Denoising and Resource Allocation in Uncoded Video Transmission

Hao Cui, Ruiqin Xiong, *Member, IEEE*, Chong Luo, *Senior Member, IEEE*, Zhihai Song, and Feng Wu, *Fellow, IEEE*

***Abstract*—The recently revitalized uncoded transmission has shown great potential in handling channel variations and user heterogeneities for wireless video communications. In this paper, we propose to adopt motion-compensated temporal filtering (MCTF) at the sender and denoising techniques at the receiver to fully exploit the temporal and spatial redundancy in video signals and therefore significantly improve the efficiency of uncoded video transmission. Although there are readily applicable MCTF and denoising techniques, integrating them into the uncoded video transmission framework is nontrivial, especially in the case of denoising. To address these challenges, we first propose spatial-domain transmission and cascaded denoising to harness the potential of denoising techniques. Second, we develop novel resource allocation algorithms based on variable-size L-shaped chunk division, which are not only compatible with spatial-domain transmission, but also achieves better energy efficiency than existing schemes based on fixed-size rectangular chunks. Our experimental results show that when transmitting 720p videos, our scheme achieves up to a 3.3 dB gain in video PSNR over the state-of-the-art uncoded transmission scheme SoftCast and up to a 5.3 dB gain over a digital scheme based on robust rate adaptation and H.264 scalable video coding (SVC).**

***Index Terms*—Image/video processing, multimedia communication, cross layer design.**

I. Introduction

TODAY'S video communication framework is designed according to Shannon's separation theorem, which suggests that source coding and channel coding can be separately designed and optimized. In the past decades, the source coding camp has made great effort to remove source redundancy through prediction, transformation, quantization and entropy coding in order to represent the video with the minimum number of bits. As a result, every bit in the encoded stream is critical for successful decoding. Therefore, during transmission, channel coding needs to be introduced to protect every single bit and to correct any possible errors. This framework has achieved great success for digital video communications.

However, with the prevalence of wireless networks and the emergence of various mobile devices, the inherent problem of the conventional framework emerges, that is the lack of flexibility in handling channel and user heterogeneity. Actually, the source coding camp realized this problem and started to investigate scalable video coding (SVC) since the 1990s, from early MPEG-4 FGS [1] and MCTF (Motion Compensated Temporal Filtering) [2], [3] to H.264 SVC [4]. Scalability in terms of quality, frame rate and resolution have all been examined. Unfortunately, SVC still has not been widely adopted in practical systems. The main obstacle is that SVC suffers from a non-negligible performance loss compared with its non-scalable counterpart.

Recently, uncoded transmission has received increasing attention. The theoretical work by Gastpar *et al.* in [5] pointed out that coding was not necessary in some cases. Subsequently, Gastpar *et al.* in [6] and Kochman *et al.* in [7] proved that uncoded transmission was optimal in a simple Gaussian sensor network and for a matched colored source/channel, respectively. The first concrete scheme for uncoded video transmission is SoftCast [8], which skips motion estimation, quantization, entropy coding and channel coding, and simply uses 3D-DCT to de-correlate the video source. The DCT coefficients are then power-scaled and directly transmitted in analog. Such uncoded transmission no longer requires channel estimation, and the quality of the received video signal naturally varies with the channel condition. Evaluations in real wireless environments have shown that SoftCast achieves significant gains over H.264 and H.264 SVC in serving heterogeneous users. The root cause of such gain is that the uncoded transmission, without compromising performance, provides finer-grained adaptation in comparison to SVC.

There remains much room for improvement in uncoded video transmission. Although it has shown outstanding performance when receivers are highly diverse and/or the channel condition varies dramatically, its performance under static settings is still inferior to conventional digital methods. We discover that SoftCast does not fully exploit the spatial and temporal correlations in video. In this paper, we improve the efficiency of uncoded video transmission through two additional modules, namely MCTF at the sender and denoising at the receiver.

Manuscript received November 01, 2013; revised April 04, 2014 and June 25, 2014; accepted July 02, 2014. Date of publication July 11, 2014; date of current version January 20, 2015. This work was carried out while H. Cui and Z. Song were interns at Microsoft Research Asia. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Béatrice Pesquet-Popescu.

H. Cui is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230026, China (e-mail: hao.cui@live.com).

R. Xiong and Z. Song are with the School of Electrical Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: xqxiong@gmail.com; zhhsng@gmail.com).

C. Luo is with Microsoft Research Asia, Beijing 100080, China (e-mail: cluo@microsoft.com).

F. Wu was with Microsoft Research Asia, Beijing 100080, China. He is now with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230026, China (e-mail: fengwu@ustc.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2014.2338279

Although we adopt existing signal processing methods, integrating them, especially the denoising module into the uncoded video transmission framework, brings non-trivial challenges. The contributions of this paper are two-fold.

- We propose spatial-domain transmission in order to leverage denoising methods. In addition, we adopt cascaded denoising to deal with different types of channel impairments.
- We propose novel resource allocation algorithms based on variable-size L-shaped chunk division which not only are compatible with spatial-domain transmission but achieve better energy efficiency than existing schemes based on fixed-size rectangular chunk division.

While the initial idea was briefly introduced in our conference paper [9], we present a complete design for denoising-aware resource allocation and perform extensive evaluations in this work. In particular, we implemented our system on a software radio platform called Sora [10] and compared it to SoftCast [8] and an adaptive digital communication system based on an H.264 SVC extension. Results show that our system achieves significant gains up to 3.3 dB over SoftCast and up to 5.3 dB over an SVC-based scheme for channel SNRs ranging from 4 dB to 20 dB.

The rest of this paper is organized as follows. In Section II, we review related theoretical work on uncoded transmission and practical uncoded video transmission systems. Section III provides an overview of the proposed system, highlighting the differences from previous systems and explaining the denoising operations in detail. Section IV presents the bandwidth and power allocation algorithms for spatial-domain transmission. Section V details the implementation of our proposed system named Cactus. Section VI presents an evaluation of our system and provides performance comparisons against reference schemes. We finally summarize our work in Section VII.

II. Related Work

Although the implementations are different, the proposed system is closely related to analog joint source-channel coding (JSCC), which has received extensive theoretical study. In this section, we will briefly review these theoretical works. In addition, we also discuss several practical video communication systems that have emerged recently.

A. Theoretical Work on Uncoded Transmission

Uncoded transmission was investigated as early as in the 1960s. Simple as it is, it has surprisingly been shown to be optimal in a few practical cases. A famous example is transmitting a uniform-distributed binary source with the Hamming distance distortion metric over a binary symmetric channel. Another one is transmitting a memoryless Gaussian source with a squared-error distortion metric over an AWGN channel.

Recently, there has been a revitalization of uncoded transmission. Gastpar *et al.* [5] point out that channel coding is not necessary in some cases, but the source and the channel have to be matched in a probabilistic sense for optimal communication. Xiao *et al.* [11] consider the transmission of a discrete memoryless Gaussian source through a discrete memoryless fading channel with AWGN. They find that linear scaling achieves the smallest mean squared error (MSE). Gastpar [6] shows that uncoded transmission is exactly optimal for a simple Gaussian "sensor" network, i.e. each sensor's channel input is merely a scaled version of its noisy observation. Recently, Kochman and Zamir [7] showed that, by combining prediction and modulo-lattice arithmetic, one can match any stationary Gaussian source to any colored-noise Gaussian channel, and hence achieve Shannon's capacity limit.

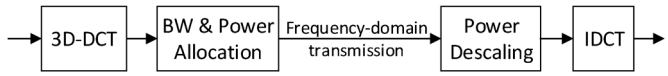


Fig. 1. Signal processing flowchart of existing uncoded video transmission.

The aforementioned review of uncoded transmission shows that there are many cases where uncoded transmission achieves optimal or near-optimal performance. Moreover, uncoded transmission is less complex and more robust than digital schemes based on separate source-channel coding and is not sensitive to exact channel knowledge at the sender.

B. Practical Schemes

SoftCast [8] is a pioneering uncoded video communication system. Fig. 1 shows its main processing modules. At the encoder, a group of pictures (GOP) are de-correlated through the 3D-DCT transform. The role of transform is thoroughly analyzed in [12]. Then the DCT coefficients are divided into equal-sized rectangular chunks, and those in the same chunk are considered as instances drawn from the same Gaussian distribution. The source dimension and channel dimension are matched through discarding chunks with the smallest variation or energy. Then, coefficients are scaled before amplitude modulation, with scaling factors set to be inversely proportional to the fourth root of the variance. At the decoder, those DCT coefficients are decoded using a linear least squares estimator (LLSE). Finally, video frames are reconstructed by an inverse DCT (IDCT).

It has been shown that SoftCast is robust to channel variations and is capable of achieving graceful degradation in a wide range of channel conditions. Its advantages over conventional digital transmission become significant in wireless multicast when different receivers have varying channel conditions. However, there still exists much room to improve the efficiency of SoftCast. First, 3D-DCT without motion alignment cannot fully exploit the temporal correlations in video. Second, spatial decorrelation based on the DCT transform does not fully utilize the local redundancy.

Following SoftCast, ParCast [13] considers uncoded transmission in MIMO-OFDM. Their contributions are on the transmission side. In particular, they propose to match the most important signal components to the best-quality sub-channels. ParCast adopts almost identical source processing as SoftCast, and therefore does not fully utilize video redundancy either. Fan *et al.* [14], [15] take motion (temporal correlation) into consideration in their design of DCast. They employ distributed source coding to take advantage of temporal correlation at the decoder. Later, Zhang *et al.* [16] improve DCast by using variable block

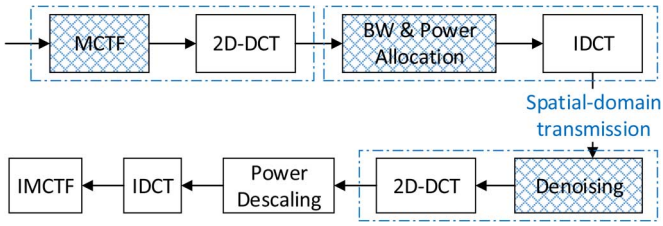


Fig. 2. Signal processing flowchart of the proposed system.

size motion estimation. However, in all the three works, spatial correlation is not fully utilized. Yu *et al.* [17] propose a hybrid digital-analog scheme which combines the low bitrate digital coding (as base layer coding) with the linear analog coding (as enhancement layer coding). Temporal correlation is only exploited in the base layer through digital coding.

Different from all the aforementioned systems, we propose to utilize temporal correlations at the sender through MCTF and spatial correlations at the receiver through denoising. Although there are existing signal processing methods that can address these needs, integrating them into the uncoded video transmission framework is non-trivial, especially in the case of denoising.

III. Overview of The Proposed System

We aim to improve the transmission efficiency of the uncoded system through exploiting the correlations in video signals. We propose to adopt MCTF at the sender to take advantage of the temporal correlations and leverage denoising techniques at the receiver to utilize the spatial correlations. Fig. 2 displays the signal processing flowchart of the proposed system. The three dashed-line boxes highlight the differences from previous work, which contribute to the significantly improved transmission efficiency.

First, we replace the 3D-DCT module in the existing uncoded framework with MCTF and 2D-DCT. The reason is that 3D-DCT without motion alignment cannot fully exploit temporal correlations. We adopt MCTF instead of the more popular motion estimation and compensation techniques in the current video coding standards because they are based on closed-loop prediction, i.e. the prediction is based on the reconstruction of previous frames. In uncoded transmission, however, the encoder is unable to obtain the exact reconstruction at the decoder. This situation is similar to that in SVC, where part of the compressed data may be dropped. Therefore, we adopt the motion-compensated temporal filtering (MCTF) [2], [18], [19] developed for SVC to remove temporal redundancy. MCTF is based on an open-loop prediction model, i.e. the prediction is based on the original pixel values instead of the reconstructed ones. It has been shown that this leads to drifting errors that are much smaller than those of its closed-loop counterpart. As MCTF is a well-developed technique in source coding and is not challenging to integrate, we do not expand on its technical details here but instead refer readers to the paper [19].

Second, after 2D-DCT, the frequency-domain coefficients are truncated in order to match the source and channel bandwidths. The remaining coefficients are then power-scaled to minimize the MSE under the given power budget. Of note in the proposed system is that we perform an inverse DCT (IDCT) over the power-scaled DCT coefficients and convert the signal back to the spatial domain for uncoded transmission. The reason is that denoising algorithms generally perform better if the signal impairments (including loss and additive noise) are presented in the spatial domain. If a pixel value is lost during transmission, the receiver can easily conceal the error through median filtering or interpolation, while in contrast, if a DCT coefficient is lost, the receiver will not have any clue about the original value and the best concealment is to set it to zero. However, we find that spatial domain transmission leads to non-trivial challenges in the resource allocation steps. In the next section, we will detail the proposed resource allocation algorithms which are well-suited for spatial-domain transmission.

Third, we propose to leverage image denoising techniques to fully exploit the spatial correlations at the receiver. We also emphasize that denoising should be immediately performed at the channel output. We will show in Section VI-C that performing denoising after power de-scaling or after the inverse MCTF is less effective than performing it at the channel output. We further propose cascaded denoising which adopts two or more denoising techniques to handle different unfavorable channel conditions. In particular, we first adopt the classic median filter [20] to handle losses. Under ideal interleaving, packet loss creates randomly dispersed pixel "holes" in the frame, as shown in Fig. 3(a). These holes are filled with the median of the surrounding eight pixel values. In the case of deep fading, the pixel values which experience deep fade can also be filled using the median filter. Fig. 3(b) shows the result after median filtering. The lost pixels become non-obvious. Then the state-of-the-art denoising algorithm BM3D [21] is adopted to reduce random noise.

The complete BM3D algorithm has two estimation steps: basic estimation and final estimation. Each estimation is also composed of two steps: block-wise estimation and aggregation. In block-wise estimation, similar blocks in a large neighborhood are found for each block and are stacked in a 3D array. Then, a 3D transformation, hard thresholding (Weiner filtering in the final estimation), and an inverse 3D transformation are consecutively performed to generate estimates for all the involved pixels. After all the blocks are processed, overlapping estimates are aggregated through a weighted sum operation. Fig. 3(c) shows the result after BM3D denoising. It can be seen that all the frames become smoother.

Note that the combination of median filtering and BM3D is just one possible choice in this step. Different denoising algorithms may be chosen for different devices according to their computational capability and end requirements. However, we discovered that using a more complex and effective algorithm than median filtering in the first denoising step does not bring much gain. This may be due to the fact that the subsequent BM3D algorithm is very powerful.

IV. Resource Allocation for Spatial-Domain Transmission

In wireless communications, bandwidth and power are the major limiting resources. In the conventional coded framework, source bandwidth and channel bandwidth are matched through quantization. Transmission power is almost evenly distributed to each information bit. In uncoded transmission, however, the bandwidth expenditure is decided by the number of coefficients

(a) Channel output (b) After median filtering (c) After BM3D (d) Recovered frame

Fig. 3. Denoising processing at the receiver in the proposed framework. (a) Channel output; (b) After median filtering; (c) After BM3D; (d) Recovered frame.

or pixels no matter whether they are quantized or not. Therefore, quantization becomes unnecessary and bandwidth matching has to be achieved through data truncation. In addition, the power used to transmit a symbol is proportional to the square of its value. When the total power is given, we shall fairly allocate the power to different variables in order to minimize the MSE.

A. Bandwidth Allocation

The source bandwidth of a video signal can be computed by $W \times H \times F$, where W and H are the width and height of a video frame and F is the frame rate. Without source compression, the source bandwidth is very large, especially for high-definition videos. The available channel bandwidth is usually smaller. Therefore, it is necessary to truncate the video data in a manner that fits the important information into the limited channel bandwidth.

It is known that data truncation should be performed in the frequency domain after the data are properly de-correlated. In a conventional digital video encoder, a video frame (either original or residual) is divided into blocks, and block-DCT is performed to transform pixel values into frequency coefficients. Then the coefficients are quantized. Using a larger quantization parameter (QP) will allow coefficients to be represented by fewer bits. In particular, small coefficients, usually appearing in the high frequency bands, may be quantized to zeros and exempted from subsequent encoding operations. In the uncoded transmission system SoftCast, frame-DCT is performed and the frequency coefficients are then divided into equal-sized chunks. The chunks with the minimal energy are discarded and the coefficients in the remaining chunks are kept completely intact.

Unfortunately, neither of the previous approaches can be adopted in our system. This is because we will transmit the video signal in the spatial domain. In order to reduce source bandwidth, we have to reduce the number of pixels. However, neither quantization nor truncation of frequency-domain coefficients will change the number of pixels after the IDCT.

We tackle this problem through a novel L-shaped data truncation. In particular, we perform 2D-DCT for each frame in a GOP after MCTF. The high frequency bands, containing more low-energy coefficients, reside on the right and the bottom of each frame. Therefore, we truncate the L-shaped coefficients as shown in Fig. 4. Let $W' \times H'$ be the resolution of the remaining coefficients. Then the IDCT could use a $W' \times H'$ transform matrix. The resulting image is actually a down-sampled image of the original one. Transmitting the down-sampled image achieves bandwidth reduction.

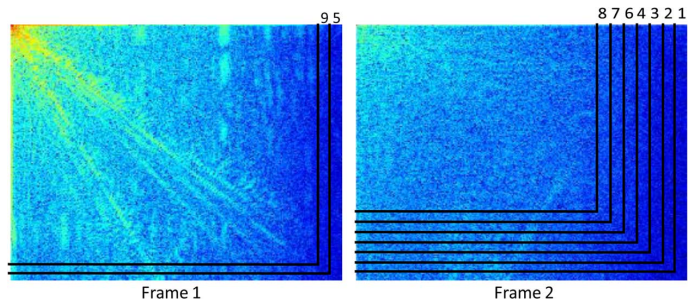


Fig. 4. L-shaped data truncation.

Algorithm 1: L-shaped data truncation

Data: $B_a, W, H, F^1, \ldots F^N, \Delta W$
Result: $W^1, \ldots W^N$

Initialization: $W^1 = W^2 = \ldots = W^N = W$, $B_s = W \times H \times N$, $\Delta H = (H/W)\Delta W$;
Define $\mathcal{L}^n = \{c_{ij}^n | W^n - \Delta W < i \leq W^n, H^n - \Delta H < j \leq H^n\}$;
for $n = 1$ to N **do**
  $\lambda^n = Var\{\mathcal{L}^n\}$;
end
while $B_s > B_a$ **do**
  $n = argmin\{\lambda^n\}$;
  $B_s - = |\mathcal{L}^n|$;
  $W^n - = \Delta W$;
  $\lambda^n = Var\{\mathcal{L}^n\}$;
end

Algorithm 1 gives the L-shaped data truncation algorithm. Since the wavelet transform is performed along the temporal axis, the low-pass and high-pass frames in a GOP have imbalanced energy. Therefore, bandwidth allocation should be performed per GOP basis. The input of the algorithm is the available bandwidth per GOP (B_a), the video resolution ($W \times H$) and the coefficients within each frame of a GOP, denoted by $F^1, \ldots F^N$, where N is the GOP size and $F^n = \{c_{ij}^n\}$ ($i = 1 \ldots W$, $j = 1 \ldots H$, $n = 1 \ldots N$). The output is the new widths of each frame, denoted by $W^1 \ldots W^N$, in the GOP. For simplicity, we fix the aspect ratio of the remaining coefficients in each frame, so it is not necessary to output the heights. The

parameters ΔW and ΔH are the horizontal and vertical truncation steps respectively.

In this algorithm, lines 3–5 compute the variance of the bottom-right L-shaped chunk for each frame. Lines 6–10 repeatedly discard L-shaped chunks that have the minimal variance (or energy) until the bandwidth requirement is met. Fig. 4 gives an example of the data truncation process when only two frames are considered. The numbers above the L-shaped chunks indicate the order that each chunk is discarded. Normally, more chunks will be discarded among high-frequency frames than from low-frequency frames.

B. Power Allocation

In wireless communications research, it has been shown that in order to optimally transmit a signal under the MSE criterion in a power-constrained system, the signal should first be de-correlated and then each coefficient should be scaled by a factor which is inversely proportional to the fourth root of its variance [22]. In our proposed video transmission system, the video signal is de-correlated by MCTF and 2D-DCT. Ideally, each transform coefficient is scaled individually according to its variance. However, as the scaling factors are required at the receiver for signal recovery, there is a tradeoff between power scaling efficiency and overhead. In our design, we adopt a compromise similar to SoftCast [8] that groups nearby coefficients into chunks and models the values in each chunk as random variables (RVs) from the same distribution. Then the coefficients in the same chunk will be scaled by the same factor and the overhead is only one scaling factor per chunk.

In contrast to SoftCast which divides coefficients into equal-sized rectangular chunks, we propose a new variable-size L-shaped chunk division method. The motivation for L-shaped chunk division is that transform coefficients decay rapidly from low-frequency to high-frequency and those belonging to a similar frequency band (constituting an L-shape) are more likely to have similar values. Grouping similar values in a chunk would allow an uncoded communication system to achieve higher power efficiency with a small overhead. An additional reason for using variable-size chunks is that the distribution of DCT coefficients differs frame by frame and video by video. While the initial idea of L-shaped chunk division has been mentioned in our earlier work [23], we present an algorithm with significantly reduced complexity in this paper.

Similar to bandwidth allocation, power allocation should also be performed on a GOP basis. Suppose that a GOP has been divided into K L-shaped chunks, then the scaling factors for each chunk are given in the following Lemma. Proof of optimality is not given here as it can be easily derived from the conclusions drawn in [22] and [8].

Lemma 1: Given K variable-size chunks, denoted by $C_1 \dots C_K$, each with size $m_k = |C_k|$, assume that the coefficients in the kth chunk are drawn from the same distribution $\mathcal{D}_k$ with zero mean and variance λ_k. Given unit average transmission power, the scaling factor for each chunk, denoted by $g_1 \dots g_K$, that minimizes MSE is:

$$g_k = \lambda_k^{-\frac{1}{4}} \sqrt{\frac{M}{\sum_k (m_k \sqrt{\lambda_k})}},\ k = 1 \dots K \tag{1}$$

where $M = \sum_k m_k$ is the total number of coefficients or equivalently the total power budget for a GOP.

Algorithm 2: Variable-size L-shaped chunk division

Data: $W^1, \dots W^N$, $F^1 \dots F^N$, K
Result: $\mathcal{B} = \{\{l_1^1 \dots l_{K_1}^1\} \dots \{l_1^N \dots l_{K_N}^N\}\}$
1 Initialization: $\mathcal{B} = \emptyset$, $l_0^n = 0$;
2 **if** $W^n > 0$ **then**
3 $W^n \rightarrow \mathcal{B}$
4 **end**
5 Greedy chunk division:
6 **for** *each chunk* k *in each frame* n $(1 \le k \le K_n)$ **do**
7 Find $l_d(n,k)$ $(l_{k-1}^n < l_d(n,k) < l_k^n)$ which minimizes $\Delta\Gamma$;
8 Record $\Delta\Gamma_{\min}(n,k)$;
9 **end**
10 **while** $|\mathcal{B}| < K$ **do**
11 Find $\Delta\Gamma_{\min}$ among all the $\Delta\Gamma_{\min}(n,k)$;
12 Add corresponding position $l_d(n,k)$ to $\mathcal{B}$;
13 Sort the boundary positions for frame n in $\mathcal{B}$;
14 For the two new chunks, find $l_d(n,k)$ and $l_d(n,k+1)$ that minimizes $\Delta\Gamma$ within the respective chunk and record $\Delta\Gamma_{\min}(n,k)$ and $\Delta\Gamma_{\min}(n,k+1)$;
15 **end**

Now the problem is how to divide a GOP into a given number (K) of chunks. For simplicity and without loss of generality, we use linear decoding and assume an AWGN channel with noise power σ^2. Then the squared error at the decoder is:

$$\varepsilon = \sum_k m_k \cdot \frac{\sigma^2}{g_k^2} = \frac{\sigma^2}{M} \left(\sum_k m_k \sqrt{\lambda_k} \right)^2. \tag{2}$$

Minimizing the squared error is equivalent to minimizing

$$\Gamma = \sum_k m_k \sqrt{\lambda_k} = \sum_k \sqrt{m_k \sum_{c_{ij}^n \in C_k} \left(c_{ij}^n\right)^2}. \tag{3}$$

Based on this analysis, we propose a greedy algorithm for chunk division. In each iteration, the algorithm will split an existing chunk into two. Let C be an existing chunk in a particular frame. For clarity, we temporarily ignore the frame and chunk indices. If chunk C is to be divided into two L-shaped chunks denoted by C^i and C^o, the change of Γ, denoted by $\Delta\Gamma$, is computed as:

$$\Delta\Gamma = \sqrt{m^i \sum_{c_{ij} \in C^i} {c_{ij}}^2} + \sqrt{m^o \sum_{c_{ij} \in C^o} {c_{ij}}^2} - \sqrt{m \sum_{c_{ij} \in C} {c_{ij}}^2},$$

where $m = |C|$, $m^i = |C^i|$ and $m^o = |C^o|$. The dividing position (l_d, t_d) should be selected among all possible positions such that $\Delta\Gamma$ is minimized.

Algorithm 2 presents the proposed variable-size L-shaped chunk division algorithm. It is performed on a per GOP basis. The inputs are the dimensions of each frame after bandwidth allocation, all the remaining coefficients and the desired number of chunks K. The outputs are the chunk boundaries in each frame. K_n indicates the number of chunks in frame n. Again,

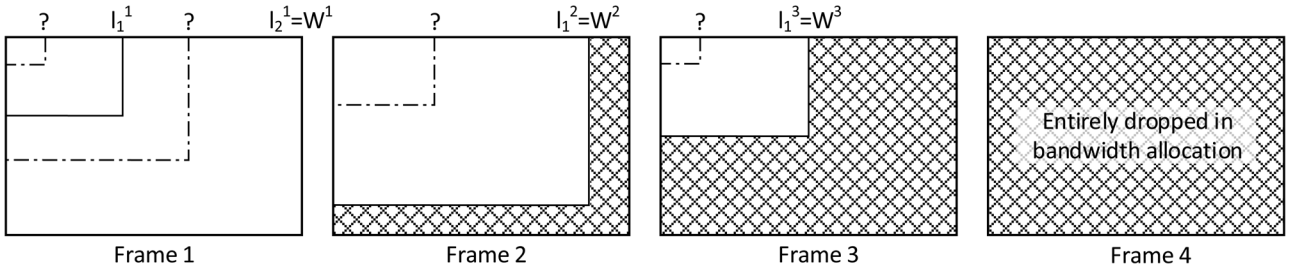


Fig. 5. Illustration of the greedy chunk division algorithm.

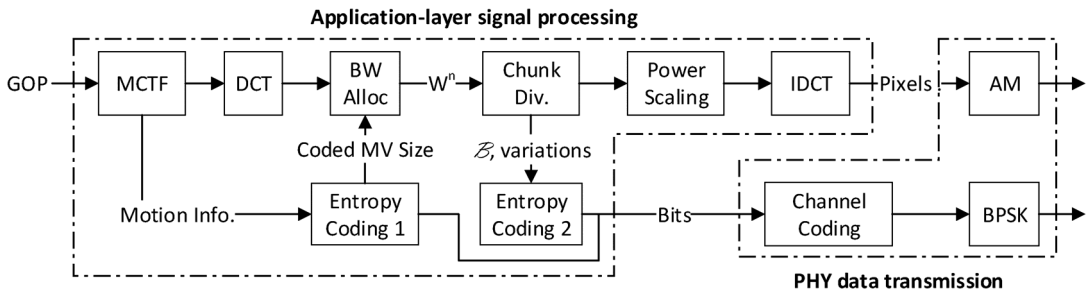


Fig. 6. Implementation of the Cactus sender is comprised of application-layer signal processing and PHY data transmission.

we fix the aspect ratio of the horizontal and vertical coordinates of each L-shaped chunk boundary. Thus it is not necessary to include the vertical coordinates in the outputs.

Fig. 5 illustrates the greedy chunk division algorithm with a simple example. The shadowed areas indicate the coefficients dropped in the bandwidth allocation process. As frame 4 is entirely dropped, the initialization step between lines 2 to 4 only adds three boundaries to $\mathcal{B}$, indicating three initial chunks. In the first round of greedy chunk division, frame 1 is divided into two chunks: the new l_1^1 is added to $\mathcal{B}$ and the previous l_1^1 becomes l_2^1 after sorting. Now, the dotted lines in Fig. 5 show four possible positions in the second round of greedy chunk division. These four positions are selected according to lines 10 to 15. The position corresponding to the minimum $\Delta\Gamma$ among the four will be added to $\mathcal{B}$ in line 12. Note that in each round of greedy selection between lines 10 to 15, we only need to find the minimum $\Delta\Gamma$ for the two new chunks. The worst case complexity is $O(KW)$.

V. Implementation

We implement the proposed uncoded video transmission system, named Cactus, through a compound approach. The application layer signal processing is implemented in Matlab Compiler Runtime (MCR) and the PHY data transmission is built on a software radio platform called SORA [10].

A. Sender

Fig. 6 depicts the Cactus sender implementation. The GOP size is set to 16. We use a reference C code to implement the barbell-lifting based MCTF [19]. The motion information, including motion vectors and modes, is entropy coded and its coded size can be calculated. The bandwidth occupied by motion information will be deducted from the overall bandwidth provisioning. The bandwidth allocation module computes the remaining frame sizes, denoted by W^n $(n = 1 \ldots 16)$. Then, we divide the GOP into 160 L-shaped chunks (10 chunks per frame on average) and compute the variation for each chunk. We use a fixed number of bits (32 bits) to record chunk boundaries and variations. These metadata are also entropy coded. The power scaling module computes the scaling factors from each chunk, denoted by g_k $(k = 1 \ldots 160)$ assuming unit power for each symbol. Then variable-size IDCT is performed on each frame to generate spatial-domain pixel values.

Two application layer signal processing steps generate metadata, which should be faithfully received by receivers. They are transmitted using a robust digital scheme. We adopt the combination of 1/2-rate channel coding and binary phase shift keying (BPSK) modulation for transmitting metadata.

The scaled pixel values are transmitted through amplitude modulation (AM). Specifically, every two pixel values are mapped to the I and Q components of one wireless symbol. Note that AM can be implemented over digital hardware using a very dense discrete modulation constellation (the precision of today's A/D converter is about 12 bits per axis). This digital implementation allows our design to be easily integrated into an existing network stack. To resist packet loss, the adjacent symbols from a frame are pseudo-randomly shuffled across different physical layer convergence procedure (PLCP) frames. We further perform inter-frame shuffling to combat fading. We limit the shuffling within a GOP to reduce the decoding delay. The shuffled symbols are sequentially placed on each orthogonal frequency division multiplexing (OFDM) symbol. Therefore, when a PLCP frame is lost, it creates randomly dispersed "holes" in the video frame, which can be easily processed by median filtering.

B. Receiver

At the receiver, the digital and pseudo-analog transmissions can be separated from the packet header. The digital BPSK symbols are demodulated and grouped for channel decoding. If all information bits are correctly decoded, they will be entropy decoded to recover the motion information, chunk division boundaries and chunk variations. Otherwise, a receiver may request for a retransmission.

The PHY at the receiver directly reads the amplitude values of pseudo-analog symbols and pieces together the 16 variable-size frames in a GOP. Each frame will be independently denoised. If packet loss is detected, we use the median function in Matlab to perform the median filter denoising. Then, we use the Matlab code published by the authors [24] to perform BM3D denoising. Then, DCT is performed and power de-scaling is applied over the frequency coefficients. The de-scaling factors can be computed from the chunk variations. For frames whose resolution is smaller than the standard resolution, we pad zeros in the high-frequency band and then perform a fixed-size IDCT. These recovered frames together with motion information can reconstruct the original frames through inverse MCTF.

VI. Evaluation

A. Methodology

Evaluations are carried out using Sora [10] (equipped with a WARP radio board) over 802.11a/g-based WLAN. The 16 bit data representation in the Sora Tool Kit is fully utilized for amplitude modulation. The carrier frequency is 2.4 GHz. The PHY is based on OFDM. Specifically, the channel is divided into 64 subcarriers and 48 of them are used to transmit modulation symbols. To reduce the overhead of the PLCP header, we use 100 OFDM symbols in each PLCP frame for data transmission. Overall, the channel bandwidth is 12 MHz and the data bandwidth is about 11.4 MHz. We have evaluated Cactus both for an exclusive stream and for two concurrent streams. The corresponding data bandwidth per stream is 11.4 MHz and 5.7 MHz, respectively. Traces are obtained at varying distances and reception power resulting in SNRs ranging from about 4 dB to about 20 dB. The SNR within each trace is fairly stable. Trace-driven evaluations ensure fairness among the comparison schemes.

We created a monochrome high-definition (HD) video sequence of resolution of 1280×720 for evaluation. It contains the first 32 frames (2 GOPs in our implementation) from 10 standard video test sequences, including *Intotree, Shields, Stockholm, City, Jets, Panslow, Parkrun, Sheriff, ShuttleStart, Spincalendar*. With a frame rate of 30 fps, the source bandwidth is 13.8 MHz (assuming 2D source samples). In order to transmit the video in a 11.4 MHz or 5.7 MHz channel, bandwidth compaction is needed and the ratio of channel bandwidth to source bandwidth is about 0.82 or 0.41.

We evaluate video delivery quality with the standard peak signal-to-noise ratio (PSNR) in dB. The PSNR is averaged across frames.

B. System Comparison

We compare our system with two reference systems, namely SoftCast and RA-SVC. The pioneering uncoded video transmission system SoftCast is implemented exactly as described in the original paper [8]. In contrast to our system Cactus, the SoftCast encoder does not generate motion information and does not need to transmit the chunk boundaries since it adopts fixed-size rectangular chunks. However, the number of chunks is 64 per video frame, which is much larger than in Cactus. The metadata of SoftCast are also transmitted with robust digital methods and

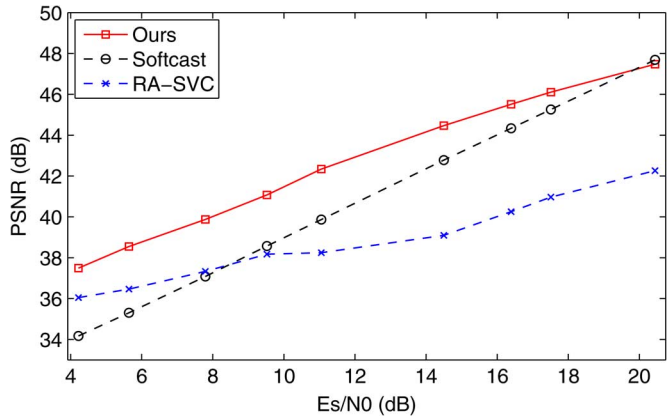


Fig. 7. Performance comparison between our system and reference systems, with a bandwidth ratio of 0.82.

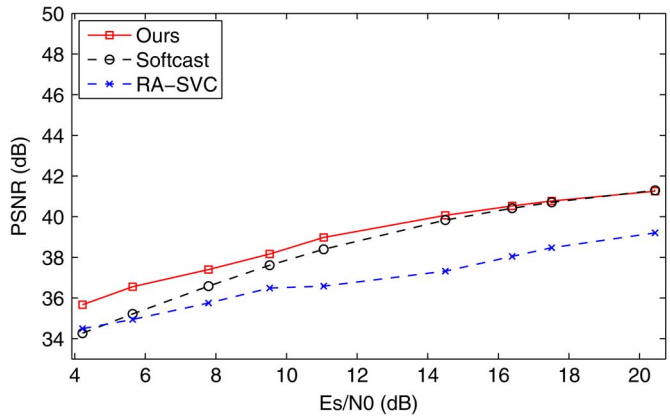


Fig. 8. Performance comparison between our system and reference systems, with a bandwidth ratio of 0.41.

can be retransmitted when there are errors. Therefore, a receiver can always assume error-free metadata.

The other reference scheme, RA-SVC, is based on the Scalable Video Coding (SVC) extension of H.264/AVC and robust rate adaptation [25]. The combined test video sequence is encoded by the H.264 reference software JSVM [26] into three quality layers. The encoding parameters are selected and tuned for each GOP to ensure the best performance. The selection criterion is that a receiver which can successfully decode (BPSK, 1/2), (QPSK, 3/4) or (16QAM, 3/4) transmissions will obtain one, two or all three quality layers. These three coding and modulation choices are as defined in 802.11a/g. We adopt the RRAA rate adaptation algorithm [25] to handle varying channel conditions. In addition, we allow instantaneous retransmission when channel decoding fails, and the base layer data are always assigned the highest priority.

Overall Performance Under Varying Channel Conditions: We evaluated the three systems over 36 traces with channel SNRs ranging from about 4 dB to about 20 dB. The SNR within each trace is fairly stable. After each transmission, the average video PSNR is computed at the receiver. We divide the receiver SNR range into 2 dB bins, and average all the (receiver SNR, PSNR) pairs whose receiver SNR falls into the same bin.

Fig. 7 compares the performance of the three systems when the bandwidth ratio is 0.82. Results show that the proposed uncoded video transmission system significantly outperforms

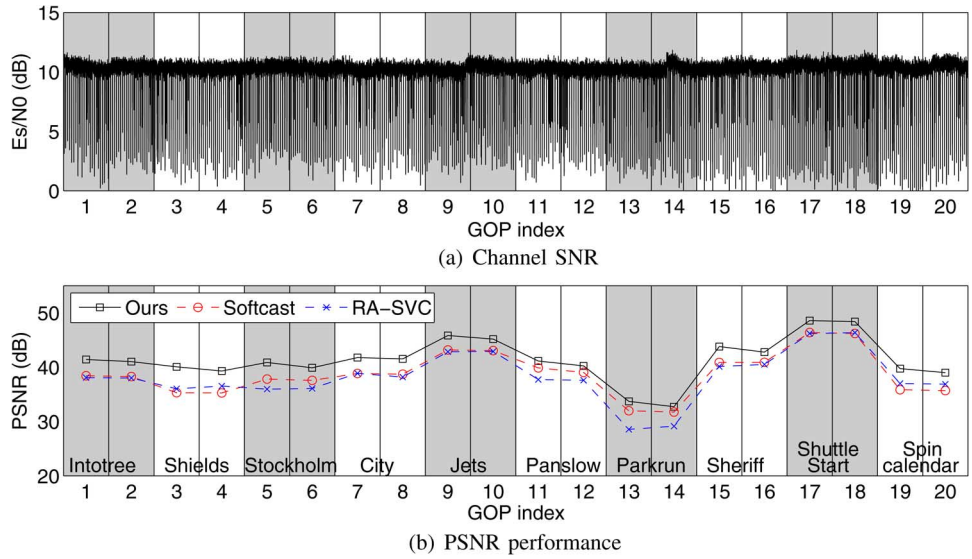


Fig. 9. Performance comparison between our system and reference systems, with a bandwidth ratio of 0.82. (a) Channel SNR; (b) PSNR performance.

SoftCast and RA-SVC. The gain over SoftCast becomes more significant as the channel condition gets poorer. When the SNR is about 4 dB, Cactus achieves a 3.3 dB gain in received video PSNR. This is due to the fact that denoising is more helpful in poor channels. When the SNR is about 20 dB, SoftCast starts to excel. This suggests that our system may turn off the denoising module when the channel SNR is above a certain threshold.

Comparing the performance of our system and RA-SVC, we find that the gain of our system increases as the channel condition improves. When the receiver SNR is between 14 dB and 20 dB, Cactus achieves about a 5 dB gain in video PSNR over RA-SVC. This is due to the fact that when more encoding layers are involved in SVC, the source coding efficiency is more heavily affected. Even when the source coding is optimal (when the channel condition is poor, only the base layer will be transmitted), the performance of RA-SVC is still inferior to our system by about 2 dB in video PSNR. The performance loss is due to the mismatched rate selection under varying channel conditions. We will give more details about this problem in the following experiment.

Fig. 8 shows the performance of the three systems when the bandwidth ratio is 0.41. This bandwidth setting is considered slightly less than adequate. In order to transmit a 120-minute 720p video with robust (BPSK, 1/2) modulation, the video has to be compressed into less than 2.57 GB. In Fig. 8, the three performance curves show trends similar to Fig. 7. Cactus achieves up to a 3.3 dB gain over SoftCast and up to a 5.3 dB gain over RA-SVC.

Performance on a Particular Trace: We next zoom into a particular trace to compare the system performance. Fig. 9 shows the per-packet channel SNRs as well as the per-GOP performance of the three systems. From Fig. 9(a), we can find that although the channel SNR is about 10 dB most of the time, there are many sudden drops to 0 to 5 dB. When the SNR drops dramatically, the selected rate (according to the previous good channel condition) in RA-SVC would be too high and the receiver may completely fail in reception. Then when the channel recovers, the selected rate could be too conservative and the channel capacity is not fully utilized. On this trace, Cactus achieves an average of 3.1 dB gain over RA-SVC.

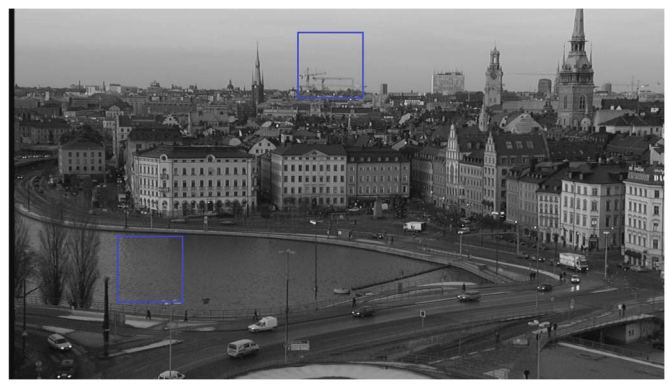


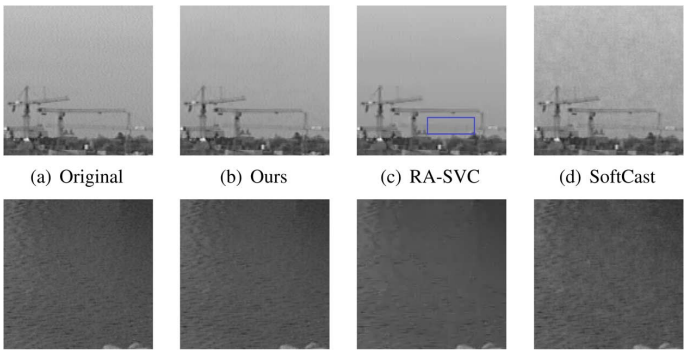
Fig. 10. Original frame #209 in the test video.

Comparing the performance of Cactus and SoftCast in Fig. 9(b), we can find that the denoising gain greatly depends on the video characteristics. It can be seen that the denoising gain over *Shields* is significant but that over *Panslow* is small. We investigate the denoising gains in more detail in Section VI-C. On average, Cactus achieves a 2.6 dB gain over SoftCast.

Fig. 11 compares the visual quality of the three schemes for frame 209, shown in Fig. 10. Frame 209 belongs to GOP 14. It was transmitted when the channel SNR was slightly above 10 dB. The PSNRs achieved by SoftCast, RA-SVC and our scheme are 30.05 dB, 31.15 dB and 35.94 dB, respectively. From the enlarged area, we can clearly see that RA-SVC tends to lose image details (see the highlighted area) and introduces some blocking effects, while the SoftCast result contains too much noise. In contrast, Cactus achieves a very clean image with details.

C. Micro-Benchmarks

In this subsection, we justify our design choices and provide insights on uncoded video transmission systems.



(a) Original (b) Ours (c) RA-SVC (d) SoftCast

Fig. 11. Comparison of image details among our method, RA-SVC and SoftCast. 11. (a) Original; (b) Ours; (c) RA-SVC; (d) SoftCast.

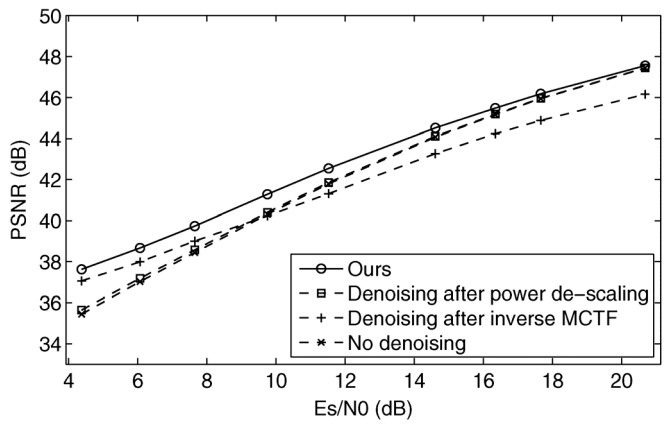


Fig. 12. Denoising gains and comparing different denoising choices.

Denoising Gains and Importance of Spatial-Domain Transmission: We have proposed to introduce denoising at the receiver to suppress channel noise. Fig. 12 shows denoising gains with respect to additive noise under varying channel conditions. This experiment is carried out without any packet loss. It can be seen that suppressing the additive noise can bring significant gains of up to 2.18 dB when the channel condition is poor. The gain decreases as the channel SNR increases and becomes negligible when the channel SNR is about 20 dB.

This figure also shows the performance of two other denoising choices, namely denoising after power de-scaling and denoising after inverse MCTF. It is clear that hardly any gain is obtained if denoising is performed after power de-scaling. This is because BM3D and most other denoising algorithms perform best for additive white Gaussian noise. After power de-scaling, the additive noise on different frequency bands will be scaled differently, and the noise distribution will be dramatically different from a Gaussian distribution, which makes denoising algorithms less effective. Performing denoising as post-processing (after inverse MCTF) also does not yield performance as good as ours. This is because the noise power, which is an important input parameter for a denoising algorithm, cannot be correctly estimated after the inverse MCTF.

In the previous subsection, we have observed that the denoising gains over different sequences could be quite different. Fig. 13 shows the denoising gain of each test sequence in a test run when the average channel SNR is 6.07 dB. While the average denoising gain is 1.64 dB, the minimal and maximal gain is 1.17 dB and 2.17 dB respectively.

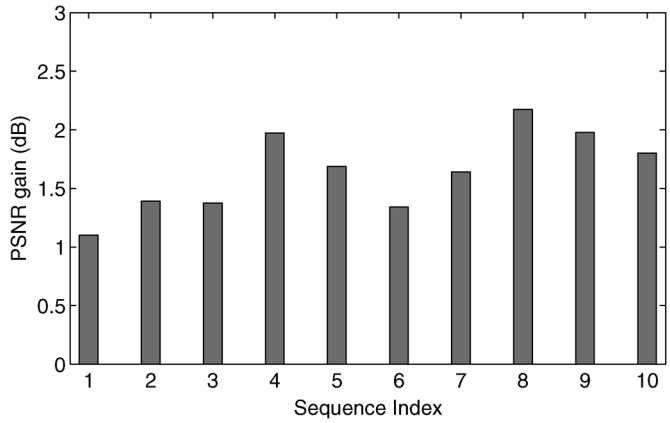


Fig. 13. Denoising gain of different sequences under a trace with an average SNR of 6.07 dB.

Denoising also plays an important role in combating losses, but only when the video signal is transmitted in the spatial domain. Fig. 14 demonstrates the robustness of our system against packet losses, although we only use a simple median filter. The results are obtained on a channel trace with 10 dB average SNR and we emulate packet losses by randomly discarding a certain percentage of the packets from the channel trace. Actually, if only additive noise is concerned, transmitting frequency-domain coefficients or spatial-domain pixel values does not make much difference, because IDCT is an orthonormal transform. However, when loss is concerned, the advantage of transmitting in the spatial domain becomes obvious. As shown in Fig. 14, our scheme achieves a 10.5 dB gain over frequency-domain transmission when the loss ratio is 10%.

SoftCast proposed using the Hadamard transform and LLSE (linear least square estimator) to combat loss. We discover that once LLSE is performed, BM3D denoising provides little improvement in the image quality, which is also due in part to the change of error distribution. From the figure, we can see that although this approach could improve the robustness against packet loss (video PSNR drops by 0.55 dB when the loss ratio increases from 0.1% to 1%), the overall performance is significantly inferior to our solution.

Benefits of L-Shaped Resource Allocation: To reduce spatial-domain bandwidth, we proposed L-shaped data truncation in the frequency domain. An experiment was conducted to show that the proposed denoising-aware L-shaped data truncation is almost as good as data truncation based on equal-chunk division, which is not well-suited for spatial-domain transmission. The evaluation metric is the total energy of the truncated data, which should be minimized. Fig. 15 shows the comparison over the first GOP of the Intotree sequence. Results on other test sequences are similar. It can be seen that the energy loss of the proposed L-shaped data truncation is almost identical to the equal-chunk data truncation when each frame is divided into 16 chunks (EC-16). It is slightly inferior to other cases when frames are divided into more chunks, but the loss is negligible.

We proposed variable-size L-shaped chunk division for power allocation. Fig. 16 shows that our solution brings significant gains over its counterpart based on fixed-size rectangular chunk division. As mentioned in Section IV, the objective of power allocation is to minimize MSE as defined in (2). As the

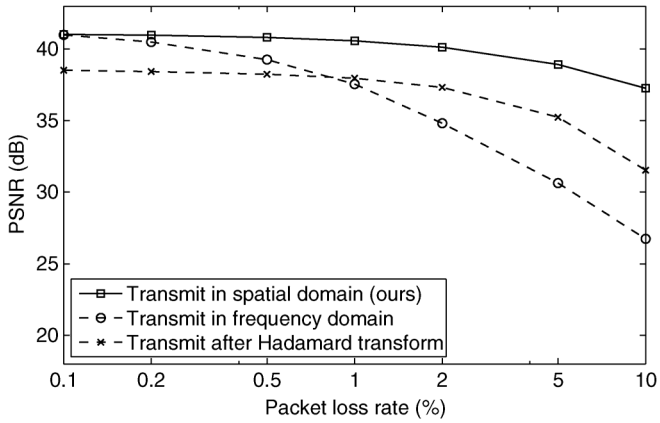


Fig. 14. Comparing different transmission strategies under packet loss.

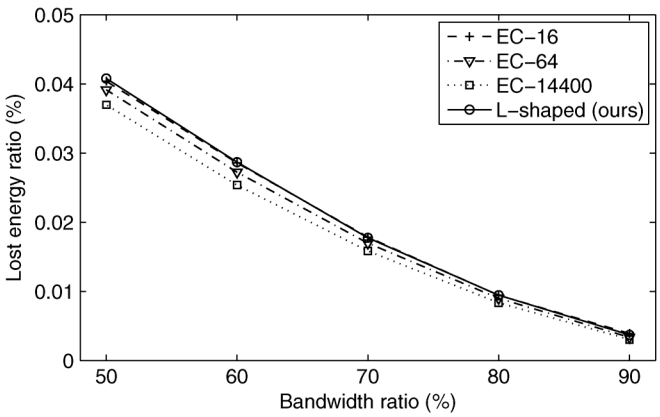


Fig. 15. Energy loss of L-shaped and equal-chunk data truncation over the first GOP of Intotree.

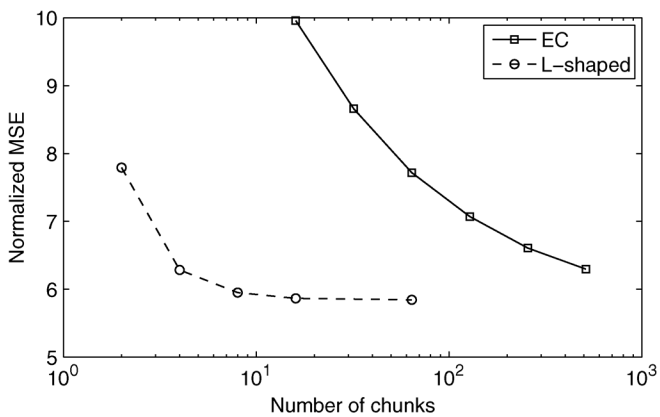


Fig. 16. Comparing power allocation performance of the proposed L-shaped chunk division and conventional equal chunk division.

absolute value of MSE is proportional to the noise power σ^2, we use the normalized MSE (ε/σ^2) as the metric for evaluating chunk division methods. From Fig. 16, we can find that the proposed L-shaped chunk division achieves much better performance than equal chunk division for the same number of chunks. When each frame is divided into 256 equal-sized chunks, the resulting normalized MSE is even larger than that with only four L-shaped chunks. We also observe that the MSE performance tends to flatten out when the number of L-shaped chunks is greater than 10 per frame.

TABLE I
BANDWIDTH PERCENTAGE OF METADATA FOR THE TEN TEST SEQUENCES

Seq. #	Motion	Total	Seq. #	Motion	Total
1	2.90%	3.06%	6	0.56%	0.73%
2	1.95%	2.12%	7	3.09%	3.26%
3	1.41%	1.58%	8	1.36%	1.52%
4	3.03%	3.20%	9	0.71%	0.88%
5	1.52%	1.69%	10	3.87%	4.04%

Metadata Overhead: In the proposed linear digital communication, there are some metadata for which there is zero tolerance for errors. We proposed to use 1/2-rate channel coding and BPSK modulation to transmit this part of the data. Table I lists the bandwidth consumption of the metadata. The total percentage includes the motion information and power scaling parameters. It can be seen that different sequences have varying amounts of motion information. On average, the metadata overhead is small, ranging from 0.73% to 4.04% of the total bandwidth.

VII. SUMMARY

We have introduced in this paper a novel uncoded video transmission system, which has the potential to provide signal processing flexibilities to wireless video communication. We show that by enabling denoising at the receiver, the efficiency of wireless video communication can be greatly improved. Extensive trace-driven experiments show that our system outperforms the conventional digital system and the state-of-the-art uncoded video transmission system SoftCast in typical channel conditions of 802.11b/g. However, when the channel SNR is extremely low (such as 0 dB), the digital scheme may outperform the proposed uncoded scheme because the former could use a high compression ratio for source coding and a very low rate channel coding for protection. It is one of our future work to explore the performance regimes of the uncoded video transmission system.

In the future, we will also explore other signal processing possibilities provided by uncoded video transmission. For instance, each receiver in a multicast session may adapt the received video resolution to the screen resolution of its mobile devices, or focus only on areas of interest area when its bandwidth is not sufficient to receive the entire video. We will also look into possibilities for information retrieval and analysis brought by this new communication framework.

REFERENCES

[1] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[2] S.-J. Choi and J. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[3] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'01)*, 2001, vol. 3, pp. 1793–1796.

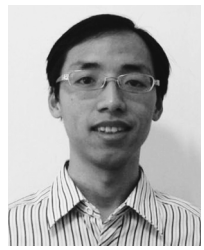
[4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[5] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[6] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "Sensor" Network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[7] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3180–3195, Jun. 2011.
[8] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. 17th Annu. ACM Int. Conf. Mobile Comput. and Netw. (MobiCom'11)*, Las Vegas, NV, USA, 2011, pp. 289–300.
[9] H. Cui, Z. Song, Z. Yang, C. Luo, R. Xiong, and F. Wu, "Cactus: A hybrid digital-analog wireless video communication system," in *Proc. 16th ACM Int. Conf. Mod., Anal. Simul. of Wireless Mobile Syst. (MSWiM'13)*, Barcelona, Spain, 2013, pp. 273–278.
[10] K. Tan, H. Liu, J. Zhang, Y. Zhang, J. Fang, and G. M. Voelker, "Sora: High-performance software radio using general-purpose multi-core processors," *ACM Commun. Mag.*, vol. 54, no. 1, pp. 99–107, Jan. 2011.
[11] J. J. Xiao, Z. Q. Luo, and N. Jindal, "CTH16-2: Linear joint source-channel coding for Gaussian sources through fading channels," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM'06)*, Nov. 2006, pp. 1–5.
[12] R. Xiong, F. Wu, J. Xu, and W. Gao, "Performance analysis of transform in uncoded wireless visual communication," in *Proc. IEEE Int. Symp. Circuits Syst., (ISCAS'13)*, May 2013, pp. 1159–1162.
[13] X. L. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "ParCast: Soft video delivery in MIMO-OFDM WLANs," in *Proc. 18th Annu. ACM Int. Conf. Mobile Comput. Netw. (MobiCom'12)*, Istanbul, Turkey, 2012, pp. 233–244.
[14] X. Fan, F. Wu, and D. Zhao, "D-cast: DSC based soft mobile video broadcast," in *Proc. 10th Int. Conf. Mobile and Ubiquitous Multimedia (MUM'11)*, Beijing, China, 2011, pp. 226–235.
[15] X. Fan, F. Wu, D. Zhao, O. C. Au, and W. Gao, "Distributed Soft Video Broadcast (DCAST) with explicit motion," in *Proc. Data Compression Conf. (DCC'12)*, Apr. 2012, pp. 199–208.
[16] A. Zhang, X. Fan, R. Xiong, and D. Zhao, "Distributed soft video broadcast with variable block size motion estimation," in *Proc. Visual Commun. Image Process. (VCIP'13)*, Nov. 2013, pp. 1–5.
[17] L. Yu, H. Li, and W. Li, "Wireless scalable video coding using a hybrid digital-analog scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 331–345, Feb. 2014.
[18] V. Bottreau, M. Benetiere, B. Felts, and B. Pesquet-Popescu, "A fully scalable 3D subband video codec," in *Proc. IEEE 2001 Int. Conf. Image Process. (ICIP'01)*, Oct. 2001, vol. 2, pp. 1017–1020.
[19] R. Xiong, J. Xu, F. Wu, and S. Li, "Barbell-lifting based 3-D wavelet coding scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1256–1269, Sep. 2007.
[20] W. K. Pratt, "Median filtering," Image Processing Institute, University of Southern California, Sep. 1975, Tech. Rep..
[21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[22] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.
[23] R. Xiong, F. Wu, X. Fan, C. Luo, S. Ma, and W. Gao, "Power-distortion optimization for wireless image/video softcast by transform coefficients energy modeling with adaptive chunk division," in *Proc. Visual Commun. Image Process. (VCIP'13)*, Nov. 2013, pp. 1–6.
[24] "BM3D algorithm and its extensions," [Online]. Available: http://www.cs.tut.fi/foi/GCFBM3D/ Tech. Rep.
[25] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust rate adaptation for 802.11 wireless networks," in *Proc. 12th Annu. ACM Int. Conf. Mobile Comput. Netw. (MobiCom06)*, 2006, pp. 146–157, Los Angeles, CA, USA: ACM.
[26] JSVM Reference Software. [Online]. Available: http://www.hhi.fraunhofer.de/

Hao Cui received his B.S. degree in electronic engineering and his Ph.D. degree in signal and information processing from University of Science and Technology of China in 2008 and 2014, respectively. His research interests include source coding, channel coding, and signal processing for wireless multimedia communication and networking.

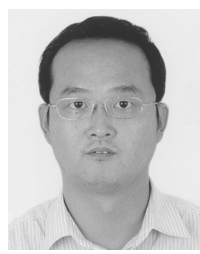


Ruiqin Xiong (M'08) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 2001, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2007.

He was with Microsoft Research Asia as a Research Intern from 2002 to 2007 and the University of New South Wales, NSW, Australia, as a Senior Research Associate, from 2007 to 2009. He joined Peking University in 2010. His current research interests include image and video processing, compression and communication.

Chong Luo (M'03–SM'14) received her B.S. degree from Fudan University in Shanghai, China, in 2000, her M.S. degree from the National University of Singapore in 2002 and her Ph.D. degree from Shanghai Jiao Tong University in 2012. She has been with Microsoft Research Asia since 2003, where she is a Lead Researcher in the Internet Media group. She is a Senior Member of the IEEE. Her research interests include wireless networking, wireless sensor networks and multimedia communications.

Zhihai Song received the B.S. degree in information security from Shanghai Jiao Tong University, Shanghai, China, in 2011, and the M.S. degree in computer science from Peking University, Beijing, China, in 2014. His research interests include image and video processing.



Feng Wu (M'99–SM'06–F'13) received the B.S. degree in electrical engineering from Xidian University in 1992. He received the M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology in 1996 and 1999, respectively. Now he is a professor in University of Science and Technology of China. Before that, he was principle researcher and research manager with Microsoft Research Asia.

His research interests include image and video compression, media communication, and media analysis and synthesis. He has authored or co-authored over 200 high quality papers (including several dozens of IEEE TRANSACTION PAPERS) and top conference papers on MOBICOM, SIGIR, CVPR and ACM MM. He has 77 granted U.S. patents. His 15 techniques have been adopted into international video coding standards. As a co-author, he got the best paper award in IEEE T-CSVT 2009, PCM 2008 and SPIE VCIP 2007. Wu has been a Fellow of IEEE. He serves as an associate editor in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA and several other International journals. He got IEEE Circuits and Systems Society 2012 Best Associate Editor Award. He also serves as TPC chair in MMSP 2011, VCIP 2010 and PCM 2009, and Special sessions chair in ICME 2010 and ISCAS 2013.