Mobile Netw Appl (2016) 21:1002–1012
DOI 10.1007/s11036-016-0734-4

Compressive Sensing Based Soft Video Broadcast Using Spatial and Temporal Sparsity

Wenbin Yin[1] · **Xiaopeng Fan**[1] · **Yunhui Shi**[2] · **Ruiqin Xiong**[3] · **Debin Zhao**[1]

Published online: 20 May 2016

Abstract Video broadcasting over wireless network has become a very popular application. However, the conventional digital video broadcasting framework can hardly accommodate heterogeneous users with diverse channel conditions, which is called the cliff effects. To overcome this cliff effects and provide a graceful degradation to multi-receivers, in this paper, we use the nonlocal sparsity and hierarchical GOP structure to propose a novel CS based soft video broadcast scheme. CS has properties of minimizing bandwidth consumption and generating measurements with equal importance which are exactly needed by video soft broadcast. In the proposed scheme, the measurement data are generated by block-wise compressive sensing (BCS), and then the measurement data packets are sent over a highly dense constellation though OFDM channel to achieve a simple encoder. Ideally, with the GOP structure, inter frame has lower sampling rate than intra frame to achieve better compression efficiency. At the decoder side, due to equally-important packets and property of soft broadcast, each user can receive the noise-corrupted measurements matching its channel condition and reconstruct video. The hierarchical GOP structure is presented to explode the correlation and non-local sparsity among video frames during the recover process. Additionally, using non-local sparsity, group based CS reconstruction with adaptive dictionaries is proposed to improve decoding quality. The experimental results show that the proposed scheme provides better performance compared with the traditional SoftCast with up to 8 dB coding gain for some channel conditions.

Keywords Compressive sensing · Video broadcast · SoftCast · Wireless network

1 Introduction

Wireless video broadcasting is becoming more and more popular in our daily life and its purpose is to transmit one video signal simultaneously to multiple receivers with different channel conditions. The main challenge we face is the difficulty to provide each receiver the video quality matching their channel conditions. The traditional wireless design such as digital video broadcasting (DVB) standard [1] and 802.11 standard [2] can hardly serve diverse users in broadcast due to the cliff effect: The encoder encode the video source at a fixed rate for a given channel capacity. Receivers whose channel condition cannot support the coding rate will fail in decoding the video while the receivers whose channel quality can meet the requirements will decode the content at the same suboptimal quality imposed by the encoding. The layered transmission scheme [3, 4] and scalable video coding (SVC) scheme [5, 6] have been used to overcome the cliff effect. SVC encodes the video signal into one basic layer (BL) and multiple enhancement layers (EL). In transmission, the hierarchical modulation (HM) [7] superimposes the multiple layer bits in one wireless symbol and allows the users to decode different numbers of layer according to their own channel condition. With SVC and HM, receivers with low channel quality can only decode BL while receivers with high channel quality can decode both BL and EL to reconstruct a better quality video. However, such separated source and channel

✉ Xiaopeng Fan
fxp@hit.edu.cn

[1] School of Computer Science & Technology, Harbin Institute of Technology, Harbin, China

[2] College of Metropolitan Transportation, Beijing University of Technology, Beijing, China

[3] Department of EECS, Peking University, Beijing, China

coding scheme reduces efficiency of compression and transmission. In addition, SVC only provides limited choices of BL and EL instead of a smooth performance.

Recently, a joint source channel coding and transmission scheme, named SoftCast [8, 9] is proposed for wireless video multicasting. SoftCast uses linear transform to compress the video data instead of quantization and entropy coding. It transmits the transform coefficients through a dense constellation after allocating a certain power, such that the received data is proportional to the transform coefficients of video frames. Since the channel noise is added to the transmitted signal, SoftCast can provide a graceful degradation with increasing noise. The receivers with high channel SNR will naturally get more exact data to reconstruct the video frame while receivers with low channel SNR also can decode the video with rough quality. Compared with conventional DVB framework, SoftCast achieves comparable result in wireless broadcasting application.

Several soft video broadcast frameworks have been proposed for better performance by utilizing different theory. Aditya et al. propose a unicast framework called Flexcast [10]. It does not have entropy coding, but adopts rateless channel coding to encode and transmit DCT coefficients. Fan et al. propose a soft video broadcast framework based on distributed video coding [11] called DCast [12, 13]. Later Fan et al. propose another soft video broadcast scheme call WaveCast [14] based on 3D wavelet transform [15–17]. Peng et al. propose a line-based coding and transmission system for high resolution satellite image called LineCast [18, 19] based on distributed coding. Yu et al. [20] propose a hybrid digital-analog (HAD) video broadcast framework, which contain a basic layer coded by H.264 [21] and an enhance layer coded like SoftCast. Xiong et al. propose a perception-friendly wireless soft video broadcast framework called G-Cast [22] based on gradient. Cui et al. [23] propose a method for robust uncoded video transmission over wireless fast fading channel.

Compressive sensing (CS) has drawn much attention as a methodology of sampling. The CS theory [24] demonstrates that a signal can be reconstructed with high probability from fewer measurements when it exhibits sparsity in some domain. Motivated by properties of CS, several CS based video coding methods have been proposed in recent years. There are mainly two categories of approaches for CS based video coding: independent source video coding approaches [25–30], which focus on video reconstruction algorithm without considering transmission, and joint source channel coding approaches [31–36], which concentrate on video broadcasting.

Among them, [27] and [28] utilize the correlations between spatial and temporal while other frameworks only use local information to reconstruct the video frame. Chen et al. [27] propose a framework using multi hypothesis predictions and achieves high reconstruction quality even at a low sampling rate. Mun et al. [28] propose a residual recovery method based on Motion Compensation (MC), which utilized the temporal correlation and sparsity of residual among video sequence. CS-MUVI [25] is designed for single pixel cameras based on representing a video in the Fourier domain or the wavelet domain. Wakin et al. [26] use a 3D wavelet based inversion algorithm to achieve CS based video compression. Kittle et al. [29] propose a scheme for video CS based on total variation (TV). Yang et al. [30] propose a CS based video coding method using Gaussian mixture models.

Compressive sensing can not only reduce bandwidth consumption, but also output real number with equal importance, which is significantly different from entropy coding. These properties make CS based video coding suitable for transmitting real number signal without protect coding like soft video broadcast. Liu et al. [31] propose a compressive broadcast in MIMO systems with receive antenna heterogeneity. Markus et al. [32] propose a video broadcast system. Xiang et al. [33] propose a scalable video coding method. Li et al. [34] propose a video coding method for wireless communication. Pudlewski et al. [35] propose a video streaming for wireless multimedia sensor networks. Wang et al. [36] propose a wireless soft video broadcast framework based on DCS. However, how to maintain good performance of the CS-based schemes in the wireless video multicast scenario still present us with many challenging problems.

Recently, nonlocal sparsity has shown beneficial to CS recovery by exploiting higher-order dependency among sparse coefficients [37]. In the proposed method, we use hierarchical GOP structure and nonlocal sparsity to propose a novel wireless video multicast approach based on compressive sensing. To keep the encoder simple and low complexity, we apply BCS method to each video frame to generate measurement data with equal importance. The measurement data are packetized in an interleaved fashion and then directly transmit over a very dense constellation over noisy OFDM channel like SoftCast. At decoder side, each user can receive the noise-corrupted measurement data matching its channel condition and reconstruct video frame by utilizing the nonlocal sparsity and local similarity among video frames in one hierarchical GOP structure. Finally, we utilize non-local sparsity, group based CS reconstruction with adaptive dictionaries to improve decoding quality. We compare our results with SoftCast and find out competitive results for some channel configurations.

The rest of the paper is organized as follows. Section 2 briefly reviews the related work on video broadcast and basic idea of compressive sensing. Section 3 describes the proposed scheme with details. The performance of our scheme is showed in section 4, followed by conclusion remarks in section 5.

2 Related works

2.1 SoftCast

SoftCast is a simple but comprehensive design for wireless video multicast, covering the functionality of video compression, data protection and transmission in one scheme. The SoftCast encoder consists of four steps: DCT transform, power allocation, Hadamard transform and direct dense modulation. DCT transform compresses the video frame by removes the spatial redundancy of a video frame. Power allocation reduces the total distortion by optimally scaling the DCT coefficients. Hadamard transform can make each packet with equal importance as a protect coding. The most attractive difference between SoftCast and traditional approach is that SoftCast directly map the data into wireless symbols by a very dense QAM. At decoder side, SoftCast uses Linear Least Square Estimator (LLSE) to reconstruct the video frame. Almost all the operations in SoftCast are linear and the channel noise is directly added on the transmitted signal. Therefore, SoftCast can overcome the cliff effect and achieve a smooth performance which means each user can get the visual quality matching his channel condition. However, SoftCast does not fully exploits the inter frame correlation. And SoftCast needs to transmit most of the DCT coefficients, so that it cannot achieve a satisfactory performance when the bandwidth is limited.

2.2 Compressive sensing

Compressive Sensing (CS) theory [24] shows that a signal can be decoded from many fewer measurements than that suggested by the Nyquist sampling theory, when the signal is sparse in some domain. More specifically, suppose that a real-valued signal $x \in \mathbf{R}^N$ is sparse in Ψ (i.e. x= Ψs), where only K out of N coefficients in s are nonzero, and that one has M samples (i.e. $y = \Phi x \in \mathbf{R}^M$) where $\Phi \in \mathbf{R}^{M \times N}$ is an random matrix. Then [24] proves that x can be perfectly recovered by solving the following problem

$$\hat{s} = \underset{s}{\operatorname{argmin}} \|s\|_1, \quad s.t. \quad y = \Phi\Psi s \tag{1}$$

When y is corrupted by noise, the recovery problem can be formulated as

$$\hat{s} = \operatorname{argmin} \|s\|_1, \quad s.t. \quad \|y - \Phi\Psi s\| < \varepsilon \tag{2}$$

where ε is the upper bound on the ℓ_2 norm of the noise. Once $\hat{s}$ is found, we can project it back to get $\hat{X} = \Psi\hat{s}$ in original domain.

3 Proposed method

The main problems in video broadcast have been described in above which are scalability, robustness and reconstruction performance. To rectify the above problems of traditional video broadcast, we propose a novel compressive sensing based video broadcast scheme. From many fewer acquired measurements than suggested by the Nyquist sampling theory, CS theory demonstrates that a signal can be reconstructed with high probability when it exhibits sparsity in some domain, which has greatly changed the way engineers think of data acquisition and compression. The scheme we proposed combines the advantages of compressive sensing and soft video broadcast. It can avoid the error accumulation and decrease the influence of channel noise. Figure 1 shows the structure of the proposed scheme. We divide input video sequences into group of pictures (GOP). One GOP contains intra frame and inter frame, they have different sampling rate at encoder side. In our method, the encoder first spilt the video frame to blocks and then sampling the blocks into measurements by multiply a random projection matrix. Since every measurement is a linear combination of all pixels in a block, these measurements have equal importance to one block. Then the measurement of different block will be putted into one packet and transmitted directly over OFDM channel towards multiple receivers with different channel conditions. The channel noise will add on the transmitted signal. Finally decoder uses the received measurements to reconstruct the video frame with help of group based sparse representation, adaptive dictionary learning and hierarchical GOP structure. Inter frames utilize the nonlocal similarity of current frame and correlations among video frames to improve the visual quality. The design of the proposed method fit the require of modern video transmit system, the encoder is made signal independent and computationally inexpensive at the cost of high decoder complexity, that is, simple encoder and complex decoder.

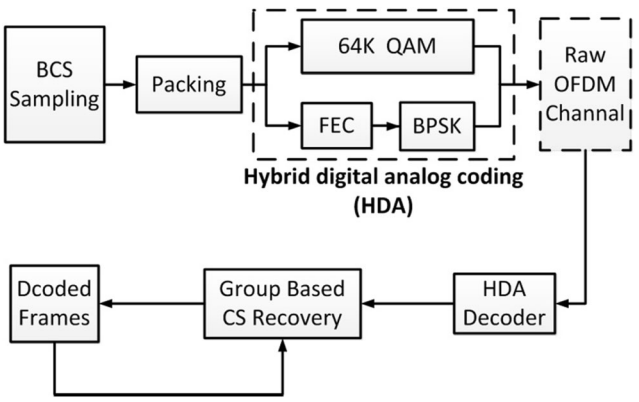


Fig. 1 Flow graph of the proposed scheme

3.1 Encoder

3.1.1 GOP structure

We use hierarchical GOP structure in our codec. The frames are classified into intra ones and inter ones. Intra frames are sampled at higher rate to achieve better recovery quality while inter frames sampled at lower rate to achieve better compress efficiency. With help of the GOP structure at decoder side, each inter frame can achieve better reconstruction quality than decoding without intra frame's help.

The GOP structure in the proposed scheme is shown in Fig. 2. One GOP consists of nine frames, the first and last frame are intra frame while others are inter frame. The encoding order equals to temporal order while the decoding order is shown on the top of each frame. Each inter frame utilizes correlate information in both forward and backward reference frames while intra frame is reconstructed respectively.

3.1.2 Encoding

At encoder side, current frame is first divided into blocks $\{u_i\}_{i=1}^{l}$ of size $\sqrt{B} \times \sqrt{B}$. Note that A is an $M \times B$ random projection matrix where M is much smaller than B. The sampling ratio equals to M/B. Each block $u_i, i = 1, 2, \ldots, l$ is represented as a column vector of length B by concatenating its columns, which is represented as $\bar{u}_i$. Intra frame is treated as key frame to help the decoding process of non-key frame. For this purpose, intra frame and inter frame use different sampling ratios. Intra frame generally has higher sampling ratio than inter frame to ensure that key frame can provide high visual quality. The measurement of each block $y_i \in \mathbb{R}^N$ is generated from

$$y_i = Au_i \tag{3}$$

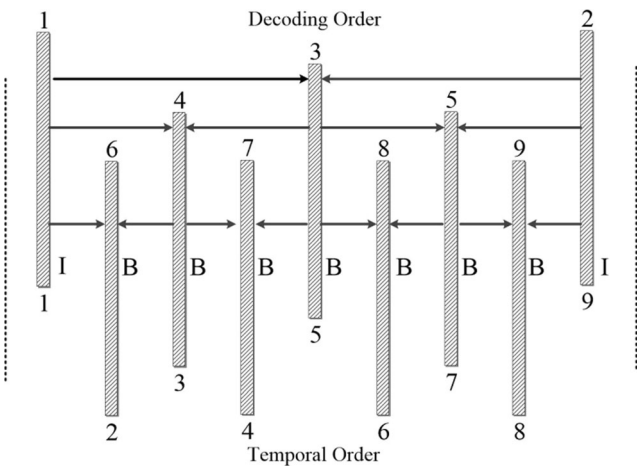


Fig. 2 GOP structure of the proposed scheme

The size of A depends on different types of frame. The sampling process is shown in Fig. 3.

The reasons why we use block based CS sampling instead of frame based CS sampling are as follow. First, the size of random projection matrix is much smaller when we use block based CS. It can save the memory and computational consumption at both encoder side and decoder side. Second, encoder can send the data packet after several blocks are sampled rather than waiting for the process of frame sampling to finish, which can save encode time.

3.1.3 Packing and transmission

The Current video multicast design is fragile to packet loss because it employs residual coding and motion compensation. These coding methods will destroy the independence between packets. The loss of one packet may cause subsequent correctly received packets to become un-decodable and bring big distortion to receiver. We need to ensure that all packets have equal importance. Hence, there are no special packets whose loss can cause loss of entire block or bring huge distortion. Since each measurement of one block contributes equally to the reconstruction of this block, we can utilize this property to generate packets with equal significance. Suppose we have several y_i in hand, put the measurement at same position in different y_i into one packet is a simple but effective way to face the packet loss. The packets create in this way have equal importance to reconstruction of these blocks, and hence offer better packet loss protection. Even one or more packet loss, the decoder still can decode these blocks without the measurements in the lost packet. The packing process is shown in Fig. 4.

In the PHY layer, the measurement data and the metadata are transmitted in different ways. The measurement data consists of real values rather than binary values. For measurement signal, it is first mapped to complex signals through a very dense 64 K-QAM constellation, i.e. every two neighboring real samples in one packet are quantized by an 8-bit quantizer and combined into one complex symbol. The complex symbols are then transmitted through raw OFDM channel directly. In contrast to traditional OFDM in 802.11 PHY layer, the raw

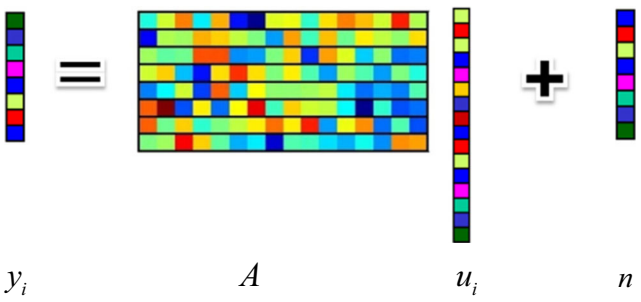


Fig. 3 Compressive sensing Sampling

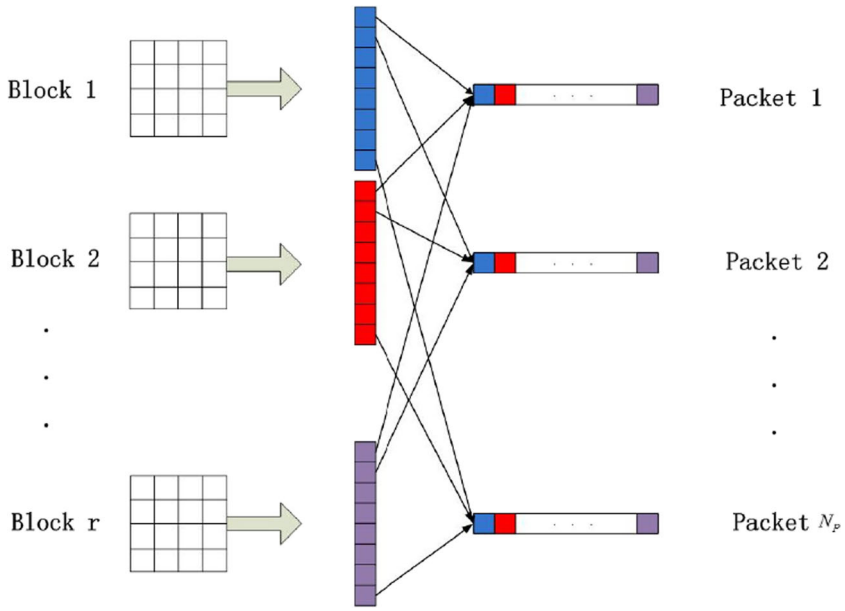


Fig. 4 The packing process in the proposed method

OFDM skips the FEC and modulation steps. An inverse FFT is computed on each packet of symbols, giving a set of complex time-domain samples. These samples are then quadrature-mixed to pass-band in the standard way. The real and imaginary components are first converted to the analogue domain using D/A converters, the analogue signals are then used to modulate cosine and sine waves at the carrier frequency respectively. And then these signals are then summed to give the transmission signal.

Unlike measurement data, the metadata is transmitted in the traditional way. Each transmitted signal is quantized by an 8-bit quantizer and encoded using variable-length-codes (VLC). The binary bits are then transmitted through OFDM in 802.11 PHY layer with FEC and modulation. To well protect the metadata part against poor channel conditions, a 1/2 convolution code is used for FEC and BPSK in used for modulation. The metadata we need to transmit is the random sampling matrix. The overhead depends on the size of the random sampling matrix. According to our experiments, the proportion of the bandwidth required by metadata is less than 3 %.

We assume the channel noise is the Additive White Gaussian Noise (AWGN). The Channel Signal to Noise Ratio (CSNR) measures the noise intensity.

3.2 Decoder initialization

Let us define n as the channel noise, the received signal is denoted by

$$\hat{y}_i = y_i + n = Au_i + n \tag{4}$$

The random projection matrix A is decoded from metadata without any loss. After we received the $\hat{y}$, a traditional CS recovery method called BCS-SPL [38] is used on it to generate an initial value for each frame.

3.3 Group based CS recovery

In this section, our goal is to reconstruct each video frame from received signal. In the following, we first discuss how to construct a group to using the nonlocal similarity. Then we will present the dictionary learning algorithm for each group. Finally, we will discuss CS recovery problem with the adaptive dictionary.

We propose a CS recovery method for video which utilizes the similarity between video frames and advantages of group based sparse representation.

3.3.1 Group construction

After we get initialization of each frame, for each frame, we first divide the initialization of current frame $X^t \in \mathbb{R}^N$ into m overlapped patches of size $\sqrt{b} \times \sqrt{b}$ and each patch is denoted by the $p_k \in \mathbb{R}^{\sqrt{b}\times\sqrt{b}}$, i.e., $k = 1, 2, \ldots, m$.

For each patch p_k, we need to search its c best matched patches in the searching window of size $L \times L$. For intra frame which has high sampling ratio at encoder side, the similar patches of current decoding patch are searching from the window only in itself. For inter frame which has low sampling ratio, the candidate list of current decoding patch is generate from 3 frames include forward reference frame, backward reference frame and current frame. Since the similar patch may be searched from any of these three frames and total number of similar patches is c, we do not fix the number of patches found in each frame.

The traditional criterion to evaluate the distance between two patches includes mean absolute difference (MAD) and Euclidean distance. These kinds of criteria lack the information of intrinsic image structures. This may cause further problem when the similar patches are only a rough estimate of current patch. To utilize the structure among patches, we use a structure-aware criterion which is proposed in [39]. It uses the gradient as structures information for distance measurement.

The criterion $Q(p, q)$ is defined as follow:

$$Q(p,q) = \|p-q\|_2^2 + \eta \times \|\nabla p - \nabla q\|_2^2 \tag{5}$$

$Q(p, q)$ is to evaluate the difference between patch q and patch q. Where ∇q represents the gradient of patch q, η is a weighting factor to adjust the contribution between pixel value and structure.

Denote all similar patches of p_k constitute S_{p_k} which is a matrix of size $\sqrt{b} \times \sqrt{b} \times c$. And then denoted by $G_k \in \mathbb{R}^{b \times c}$, which includes all patches in S_{x_k} as it columns,

i.e. $G_k = \left\{x^k_{(t-1,1)}, \ldots, x^k_{(t-1,c_1)}, x^k_{(t,1)}, \ldots, x^k_{(t,c_2)}, x^k_{(t+1,1)}, \ldots, x^k_{(t+1,c_3)}\right\}$ where $c_1 + c_2 + c_3 = c$. Especially, $x^k_{(t,1)}$ here is the current patch p_k. And then we define

$$\overline{G}_k^T = \left[x^{kT}_{(t-1,1)}, \ldots, x^{kT}_{(t-1,c_1)}, x^{kT}_{(t,1)}, \ldots, x^{kT}_{(t,c_2)}, x^{kT}_{(t+1,1)}, \ldots, x^{kT}_{(t+1,c_3)}\right] \tag{6}$$

where $\overline{G}_k^T$ is the transpose of $\overline{G}_k$. Here, we have finished the construction of group and the process is shown in Fig. 5.

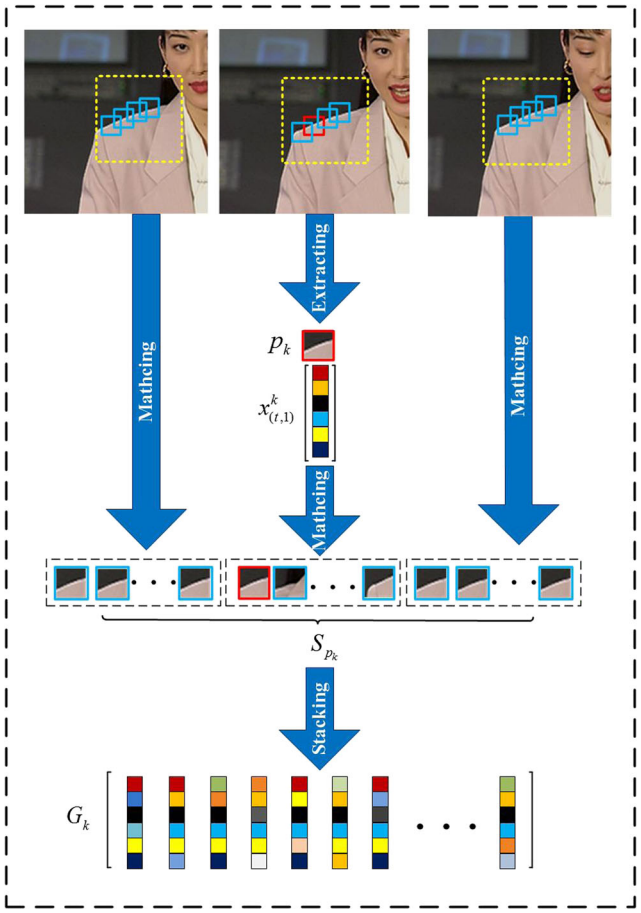


Fig. 5 The illustration for the structural group construction

3.3.2 Group sparse representation

In the traditional sparse representation model, the basic unit of sparse representation is a patch [38]. Assume X is a vector representation of original image, and then define $R_k \in R^{B \times N}$ as an operator that extracts the patch at the k-th position in the image X. Each patch X_k is denoted by $X_k = R_k X$. For the fixed dictionary D, sparse coding aims to search for sparse representations, which can be formulated as the following optimization problem

$$\hat{\beta}_k = \underset{\beta_k}{\operatorname{argmin}} \|\beta_k\|_1 + \lambda \|R_k X - D\beta_k\|_2 \tag{7}$$

where λ is a constant. Given all $\hat{\beta}_k$, the whole image X can be recovered by

$$X = \left[\sum_k R_k R_k^T\right]^{-1} \left[\sum_k R_k^T \left(D\hat{\beta}_k\right)\right] \tag{8}$$

where R_k^T is the transpose of this operation R_k which puts back a patch to the k-th position.

In the proposed method, the basic unit of sparse representation is group. Since matrix G_k represents the k-th group, the columns consist of current decoding patch p_k and all its similar patches. We assume that an adaptive group dictionary D_k is defined as $D_k = [d_k^1, d_k^2, \ldots, d_k^r]$, each d_k^i is referred to a dictionary element. We hope to represent each group G_k approximately as linear combination of D_k as

$$G_k \approx D_k \circ \alpha_k = \alpha_k^1 d_k^1 + \alpha_k^2 d_k^2 + \ldots + \alpha_k^r d_k^r \tag{9}$$

where $\alpha_k = [\alpha_k^1, \alpha_k^2, \ldots, \alpha_k^r]$. We further require that α_k is as sparse as possible. So for a known D_k, α_k can be computed by solving the following problem

$$\hat{\alpha}_k = \underset{\alpha_k}{\operatorname{argmin}} \|\alpha_k\|_1 + \lambda \|G_k - D_k \circ \alpha_k\|_2 \tag{10}$$

Note that $D_k \circ \alpha_k$ is not a strict matrix vector multiplication. At the decoder side, the patches of a GOP in current frame need to be recovered. So we define

$$R_{ki}(X^t) = \left[x^k_{(t,1)}, x^k_{(t,2)} \ldots, x^k_{(t,c_2)}\right] \tag{11}$$

where R_{ki} is actually an operator that extracts the similar patches from current frame X^t. Especially, i represent the i-th

Table. 1 A Complete Description of Decoder side

Input: the observed measurement $\hat{y}$ and the degraded operator A
Initialization: $b, c, \lambda, \alpha_k = 0, D_k = 0$
Update X^t by BCS-SPL;
Construct each structural group $\overline{G}_k$ by computing Eq. (6)
Repeat
for Each structural group G_k
Construct dictionary $\boldsymbol{D}_k$ and α_k by computing Eq. (12);
Reconstruct by $\hat{u}_i$ computing Eq. (15);
Update $\bar{u}_i$ by computing Eq. (16);
end
Until maximum iteration number is reached
Output: Final reconstructed frame.

similar patch's position of p_k. On the contrary, we denoted R_{ki}^T as its transpose, which can put back a group into the i-th position in the reconstructed image, padded with zeros elsewhere.

By averaging all the groups in current frame, the recovery of whole frame x_t from $\{G_k\}$ becomes

$$\hat{X}^t = \left[\sum_{k=1}^{m}\sum_{i=1}^{c_2} R_{ki}R_{ki}^T\right]^{-1}\left[\sum_{k=1}^{m}\sum_{i=1}^{c_2} P_{ki}\left(\overline{G}_k\right)\right] \tag{12}$$

where P_{ki} represents the operator of extracting patches of group G_k in current frame by $P_{ki}\left(\overline{G}_k\right) = x_{(t,i)}^k$. Then we will discuss the process of adaptive dictionary learning for each group G_k. Note that we hope each G_k can be faithfully represented by D_k and the sparse representation coefficients of p_k over D_k are as sparse as possible. The adaptive dictionary learning problem is defined as

$$\min_{\{D_k,\alpha_k\}} \|G_k - D_k \circ \alpha_k\|_F^2 + \mu_k\|\alpha_k\|_0 \tag{13}$$

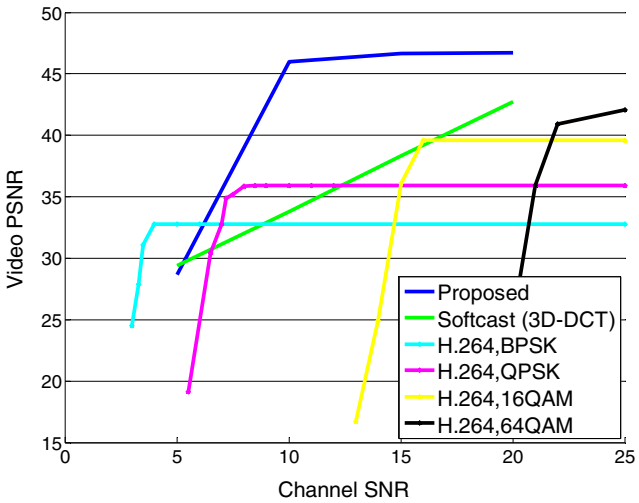
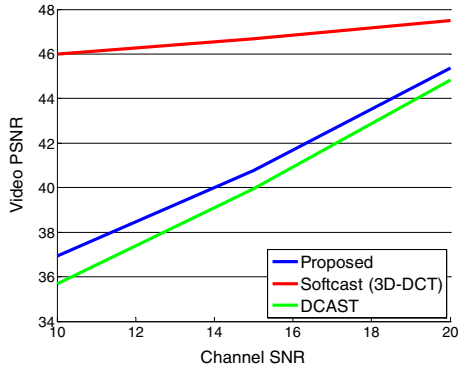
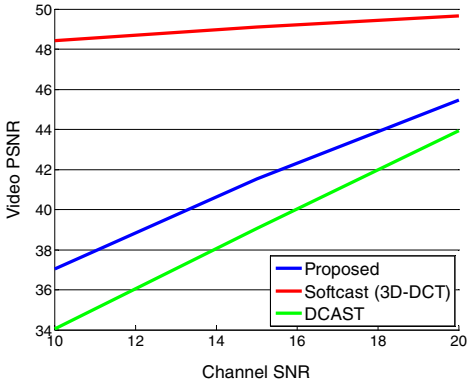


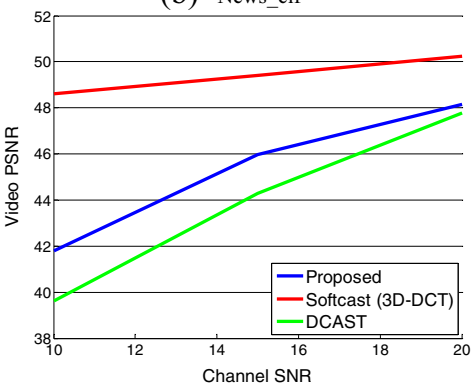
Fig. 6 Robustness between the proposed method, SoftCast and H.264



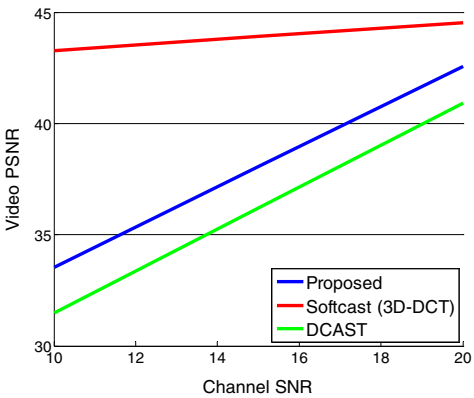
(a) Foreman_cif



(b) News_cif



(c) Mother_cif



(d) Bus_cif

Fig. 7 Multicast performance on different video sequences

Since each group consists of similar patches, G_k may be a low rank matrix. If we consider this characteristic, problem (13) can be addressed by singular value decomposition (SVD). First, we want to introduce the following hard thresh-holding operator D_τ which defined as

$$D_\tau(\Sigma_k) = diag\left(\left\{(\alpha_k^i - \tau)_+\right\}\right) \tag{14}$$

where $(\sigma_i - \tau)_+ = \max\{0, \alpha_i^k - \tau)\}$. Then each group can be represented as

$$G_k = U^k \sum^k V^{kT} = \sum_{j=1}^{r} \alpha_j^k u_j^k \left(v_j^k\right)^T = \sum_{j=1}^{r} \alpha_j^k d_j^k \tag{15}$$

where the u_j^k and v_j^k are the left and right singular vectors, and the $\sum^k = diag(\alpha_k)$ is the singular value vector of G_k. Here we denote that $d_j^k = u_j^k (v_j^k)^T$ and the matrix G_k is a linear combination of D_k. Note that each atom $d_j^k \in \mathbf{R}^{b \times c}$ is a matrix with the same size as the group G_k. And note that $diag\ (\hat{\alpha}_k) = \hat{\Sigma}_k = D_\tau(\Sigma_k)$.

3.3.3 Recovery method

According to the definition and formulation above, the group-based frame restoration scheme is formulated as

$$\begin{aligned} \left\{\hat{u}_i\right\} &= \underset{\{u_i\}}{\operatorname{argmin}} \sum_{i=1}^{m} \|Au_i - y_i\|_2^2 \\ &+ \lambda \sum_{k=1}^{m} \sum_{i=1}^{c_2} \|R_{ki} X^t - P_{ki}(D_k \circ \alpha_k)\|_2^2 \end{aligned} \tag{16}$$

We can use gradient descent algorithm to solve the above problem by $u_i = \hat{u}_i + y_i A^T A(\hat{u}_i - y_i)$ where $\hat{u}_i$ is generated by splitting the $\hat{X}^t$ into blocks with size $\sqrt{B} \times \sqrt{B}$. The $\hat{X}^t$ can be computed by formula (12).

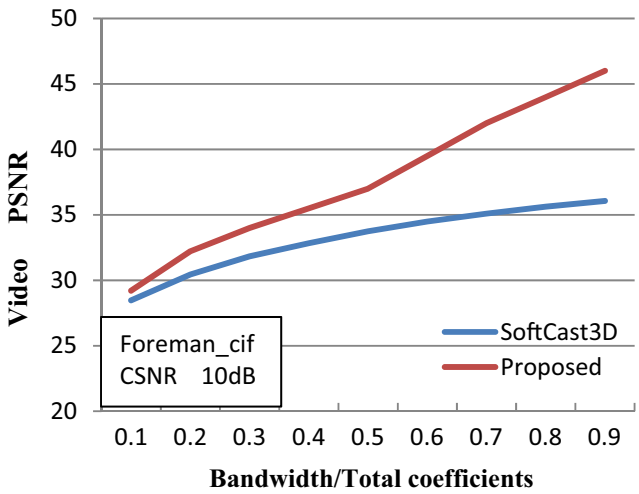


Fig. 8 Simulation of limited bandwidth network environment

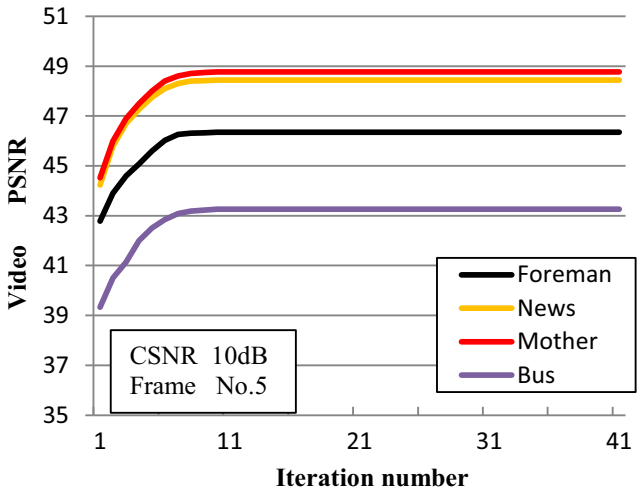


Fig. 9 Evolution of PSNR versus iteration number

So far, we have exhaustively described the decode process. Each sub-problem has been solved. The received signal is first be recovered by BCS-SPL. Then we utilize the advantage of group based spare representation and the hierarchical GOP structure to improve the reconstruction quality. In light of all derivations above, a detailed description of the process at decoder side is provided in Table. 1.

4 Experimental results

In experiments, we evaluate the performance of the proposed method in video unicast and multicast. We compare our scheme with SoftCast [9] and H.264 which use standard 802.11 PHY layer with FEC and QAM modulations. The experiment method is multicasting same video to users with different channel SNR.

The test sequences are 'foreman_cif.yuv', 'news_cif.yuv, 'mother_cif.yuv and 'bus_cif.yuv. The video frame rate is 30Hz. The block size of sampling is 32, the sampling ratio of intra frame is 0.8 and the sampling ratio of inter frame is 0.7. Notice that the average sampling ratio of our method is

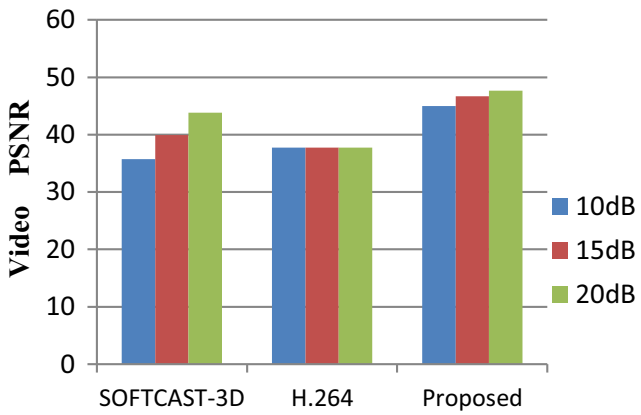


Fig. 10 Multicast to three receivers

(a) Original frame (b) SoftCast(3D)

(c) DCAST (d) Proposed

Fig. 11 Visual quality of 'Foreman_cif'

(a) Original frame (b) SoftCast(3D)

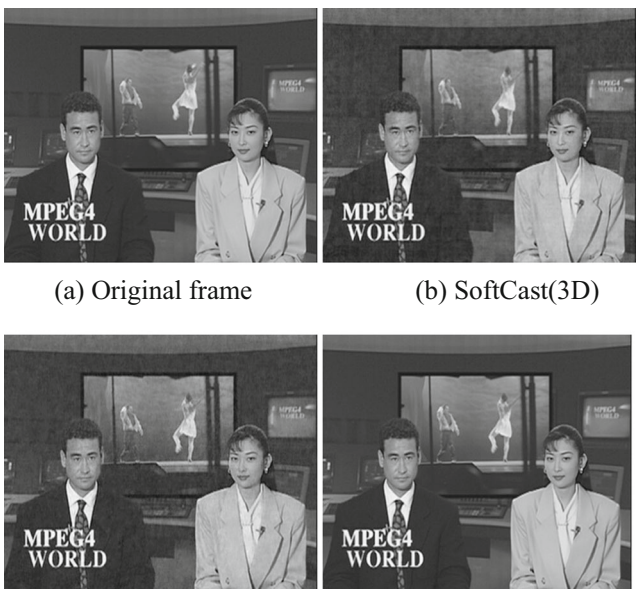
(c) DCAST (d) Proposed

Fig. 13 Visual quality of 'Mother_cif'

approximates to 0.7 and the sampling ratio of SoftCast is set at 0.8. The patch size for searching similar patch is 8. The window size for searching is 20×20. The GOP structure has shown in the above section. The video signal is transmitted over OFDM channel with AWGN.

We compare the proposed method with SoftCast and the conventional frameworks based on H.264. For conventional framework we implement 4 recommended combination of channel coding and modulation of 802.11. We calculate the corresponding bit-rate according to the bandwidth for H.264 encoder. For the proposed method, there is no bit-rate but only channel symbol rate. Note that all the frameworks consume the same bandwidth and transmission power.

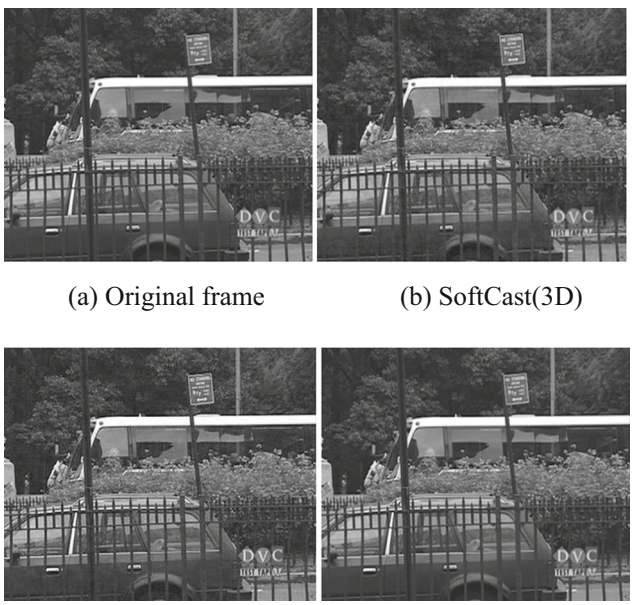
The video PSNR of each framework under different channel SNR is given in Fig. 6. In the figure, we can find out that all the four conventional transmission approaches suffer from a very serious cliff effect. For example, the approach 'H.264 with QPSK' performs well when channel SNR is between 5 dB to 8 dB, and poorly when channel SNR is out of this range. When the channel SNR is higher than 8 dB, the reconstruction quality does not increase.



(a) Original frame (b) SoftCast(3D)

(c) DCAST (d) Proposed

Fig. 12 Visual quality of 'News_cif'



(a) Original frame (b) SoftCast(3D)

(c) DCAST (d) Proposed

Fig. 14 Visual quality of 'Bus_cif'

In contrast, the SoftCast and the proposed method do not suffer from the cliff effect. When the channel SNR increases, the reconstruction quality increases accordingly. The reconstruction PSNR of the proposed method and that of SoftCast are close when channel SNR equals to 5 dB. But with the increase of channel SNR, the proposed method performs better than SoftCast. At middle channel SNR, the proposed method achieves 8 dB gains compare with SoftCast. Figure 7 gives the performance comparison on different video sequences.

Figure 8 shows the simulation result of limited bandwidth network environment. It is observed that with the decrease of bandwidth, PSNR of SoftCast and proposed method show a graceful degradation while the proposed method provides better performance than SoftCast in all limited bandwidth environments.

The Fig. 9 shows the evolutions of PSNR versus iteration numbers for test video sequences. From the figure we can observe that each sequence has different maximum iteration number but all less than 10. So we set the maximum iteration number as 8 in our experiments to balance the recovery quality and decoding time.

We then let the frameworks serve a group of three receivers with diverse channel SNR. The channel SNR for each receiver is 5 dB, 10 dB and 20 dB. In conventional frameworks based on H.264, the server transmits the video stream by using BPSK. It cannot use higher transmission rate because otherwise the dB user will not be able to decode the video. Due to this, although the other two receivers have better channel conditions, they will also only receive low speed 802.11 signal and reconstruct low quality video. In SoftCast and the proposed method, the server can accommodate all the receivers simultaneously. Using our method, the 5 dB user will get slightly lower reconstruction quality than using H.264 based conventional frameworks. However, the 10 dB and 20 dB users get 4 dB and 8 dB better reconstruction quality respectively by using our method than conventional frameworks. The test result is given in Fig. 10.

The visual quality comparison is given in Fig. 11, 12, 13, and 14. The channel SNR is set as 15 dB. The proposed method has clearly better visual quality than Softcast3D.

5 Conclusion

Our scheme achieves multicast video to receivers with different channel SNR based on compressive sensing. We compared our results with a recent method designed for the same goal and find competitive results in some channel conditions.

Our method has a simple encoder which does not require solving any resource allocation problem. The received data is reconstructed firstly at a basic quality using traditional block based CS recover scheme. And then improve the quality utilizing the advantages of GOP structure and group based sparse representation. Finally, it achieves better performance than the state-of-art multicast approach SoftCast.

Acknowledgments This work was supported in part by the National Science Foundation of China (NSFC) under grants 61472101 and 61390513, the Major State Basic Research Development Program of China (973 Program 2015CB351804), and the National High Technology Research and Development Program of China (863 Program 2015AA015903).

References

1. European Telecommunications Standards Institute (2009) Digital Video Broadcasting (DVB). ETSI Publishing ETSI official website. http://www.etsi.org/deliver/etsien/300700300799/300744/01.06.01_60/en300744v010601p.pdf. Accessed 2009
2. IEEE 802.11 Working Group et al. (2007) IEEE 802.11-2007: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE 802.11 LAN Standards-2007
3. Shacham N (1992) Multipoint communication by hierarchically encoded data In: INFOCOM '92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, pp 2107–2114 vol.3, Florence
4. McCanne S Jacobson V, Vetterli M (1996) Receiver-driven layered multicast In: Conference proceedings on Applications, technologies, architectures, and protocols for computer communications, pp 117–130, New York
5. Wu F, Li S, Zhang YQ (2001) A framework for efficient progressive fine granularity scalable video coding. IEEE Trans Circ Syst Video Technol 11(3):332–344
6. Schwarz H, Marpe D, Wiegand T (2007) Overview of the scalable video coding extension of the h.264/avc standard. IEEE Trans Circ Syst Video Technol 17(9):1103–1120
7. Ramchandran K, Ortega A, Uz K, Vetterli M (1992) Multire solution broadcast for digital hdtv using joint source-channel coding In: IEEE international conference on Communications, pp 556–560. vol.1, Chicago
8. Jakubczak S, Katabi D (2010) SoftCast: one-size-fits-all wireless video. ACM Sigcomm Comput Commun Rev 40(4):449–450
9. Jakubczak S, Katabi D (2011) A cross-layer design for scalable mobile video In: Proceedings of the 17th annual international conference on Mobile computing and networking, pp 289–300, New York
10. Aditya S Katti S (2011) FlexCast: graceful wireless video streaming In: Proceedings of the 17th annual international conference on Mobile computing and networking, pp 277–288, New York
11. Girod B, Aaron AM, Rane S, Rebollo Monedero D (2005) Distributed video coding. Proc IEEE 93(1):71–83
12. Fan X, Wu F, Zhao D, Au OC, Gao W (2012) Distributed soft video broadcast (DCAST) with explicit motion In: Data Compression Conference, pp199-208, Snowbird
13. Fan X, Wu F, Zhao D, Au OC (2013) Distributed wireless visual communication with power distortion optimization. IEEE Trans Circ Syst Video Technol 23(6):1040–1053
14. Fan X, Xiong R, Wu F, Zhao D (2012) Wavecast: Wavelet based wireless video broadcast using lossy transmission In: IEEE international conference on visual communications and image processing, pp 1–6, San Diego
15. Taubman D, Zakhor A (1994) Multirate 3-d subband coding of video. IEEE Trans Image Process 3(5):572–588
16. Secker A, Taubman D (2003) Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression. IEEE Trans Image Process 12(12):1530–1542

17. Xiong R, Xu J, Wu F, Li S (2007) Barbell-lifting based 3-D wavelet coding scheme. IEEE Trans Circ Syst Video Technol 17(9):1256–1269
18. Peng X, Xu J, Wu F (2012) Line-cast: Line-based semi-analog broadcasting of satellite images. International Conference on Image Processing, pp2929–2932
19. Wu F, Peng X, Xu J (2014) LineCast: line-based distributed coding and transmission for broadcasting satellite images. IEEE Trans Image Process 23(3):1015–1027
20. Yu L, Li H, Li W (2013) Wireless scalable video coding using a hybrid digital-analog scheme. IEEE Trans Circ Syst Video Technol 24(2):331–345
21. Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A (2003) Overview of the H. 264/AVC video coding standard. IEEE Trans Circ Syst Video Technol 13(7):560–576
22. Xiong R, Liu H, Ma S, Fan X, Wu F, Gao W (2014) G-CAST: gradient based image SoftCast for perception-friendly wireless visual communication in: data Compression Conference, pp 133–142, Snowbird
23. Cui H, Luo C, Chen C, Wu F (2014) Robust uncoded video transmission over wireless fast fading channel In: Proceedings of IEEE INFOCOM, pp 73–81, Toronto
24. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52(4):1289–1306
25. Sankaranarayanan AC, Studer C, Baraniuk RG (2012) CS-MUVI: Video compressive sensing for spatial multiplexing cameras In: IEEE International Conference on Computational Photography, pp 1–10, Seattle
26. Wakin MB, Laska JN, Duarte MF, Baron D, Sarvotham S, Takhar D, Kelly KF, Baraniuk RG (2006) Compressive imaging for video representation and coding In: Proceedings of Picture Coding Symposium, pp 1–6
27. Chen C, Tramel EW, Fowler JE (2011) Compressed sensing recovery of images and video using multi hypothesis predictions In: Conference on signals, systems and computers, pp 1193–1198, Pacific Grove
28. Mun S Fowler JE (2011) Residual reconstruction for block-based compressed sensing of video In: Proceedings of the IEEE Data Compression Conference, pp 183–192, Snowbird
29. Kittle D, Choi K, Wagadarikar A, Brady DJ (2010) Multiframe image estimation for coded aperture snap shot spectral imagers. Appl Opt 49(36):6824–6833
30. Yang J, Yuan X, Liao X, Llull P, Brady DJ, Sapiro G, Carin L (2014) Video compressive sensing using gaussian mixture models. IEEE Trans Image Process 23(11):4863–4878
31. Liu X, Luo C, Hu W, Wu F (2012) Compressive broadcast in mimo systems with receive antenna heterogeneity In: Proceedings of INFOCOM, pp 3011–3015, Orlando
32. Schenkel MB, Luo C, Wu F, Frossard P (2012) Compressed sensing based video multicast In: Proceedings of SPIE, pp 77441H-1–77441H-9, vol.7744, Huangshan
33. Xiang S, Cai L (2011) Scalable video coding with compressive sensing for wireless video cast In: Proceedings of the IEEE International Conference on Communications, pp 1–5, Kyoto
34. Li C, Jiang H, Wilford P, Zhang Y (2011) Video coding using compressive sensing for wireless communications In: Proceedings of the Wireless Communications and Networking Conference, pp 2077–2082, Cancun
35. Pudlewski S, Prasanna A, Melodia T (2012) Compressed-sensing-enabled video streaming for wireless multimedia sensor networks. IEEE Trans Mob Comput 11(6):1060–1072
36. Wang A, Zeng B, Chen H (2014) Wireless multicasting of video signals based on distributed compressed sensing. J Signal Process: Image Commun 29(5):599–606
37. Dong W, Shi G, Li X, Ma Y, Huang F (2014) Compressive sensing via nonlocal Low-rank regularization. IEEE Trans Image Process 23(8):3618–3632
38. Mun S, Fowler JE (2009) Block compressed Sensing of Image Using Directional Transforms In: Proceeding of the International Conference on Image Processing, pp 3021–3024, Cairo
39. Yue H, Sun X, Yang J (2013) Landmark image super-resolution by retrieving web images. IEEE Trans Image Process 22(12):4865–4878