

RESEARCH

Open Access

A wireless video multicasting scheme based on multi-scale compressed sensing



Anhong Wang^{1*}, Qingdian Wu¹, Xiaoli Ma² and Bing Zeng^{3,4}

Abstract

Video multicast is becoming more and more popular in wireless multimedia applications, in which one major challenge is to offer heterogeneous users with a graceful degradation against varying packet loss ratios and channel noise. In this paper, we propose a multi-scale compressed sensing-based wireless video multicast scheme, abbreviated as MCS-cast. The encoder of MCS-cast decomposes each video frame through a discrete wavelet transform (DWT) and explores an optimized compressed sensing (CS) rate to sample/measure each DWT level. The CS measurements are then packed in such a way that all packets are made as equally important as possible, while each packet includes different percentages of different DWT levels. Finally, the packets are transmitted via an analog-like modulator with mapping of the measurements into a very dense constellation. We demonstrate that because of larger percentages of more important DWT levels in each packet, packet loss leads to a much reduced influence on the reconstruction quality. Experimental results show that our MCS-cast preserves the property of graceful degradation for heterogeneous users and can outperform the state-of-the-art SoftCast by up to 3 dB in PSNR at high packet loss ratios (over the same noisy channel).

Keywords: Multi-scale; Compressed sensing; Video multicast; Discrete wavelet transform

1 Introduction

Multicasting of video signal has recently become a popular application in wireless networks, such as mobile TV, media sharing, live broadcasting of sport events, and lecturing. Because of channel heterogeneity among multiple users (e.g., the channel bandwidth they are connected and the channel error they are suffering), one big challenge imposed to video multicast is to simultaneously guarantee the best possible video quality for different users according to their individual channel characteristics.

In conventional wireless video multicast, a video bitstream coded at a specific bit rate is transmitted over the wireless network. For example, the digital video broadcasting (DVB) scheme [1] transmits the bitstream compressed by the traditional video coding standard over a wireless channel. However, this fixed bit rate usually incurs “unfairness” to users in a multicast group: if the rate accords with a low-quality receiver, users with better channel characteristics only obtain a low-quality video;

alternatively, if the rate is selected for a high-quality user, the low-quality users cannot decode the bitstream at all. On the other hand, the conventional wireless video transmission schemes have another disadvantage: they are not resilient to channel errors or noise, thus leading to a sharp decrease of reconstructed video quality (the so-called cliff effect) when the channel is corrupted by noise (it is very common in practice), and the reconstructed video quality will not improve even when the signal-to-noise ratio (SNR) becomes larger [2]. The emergence of the layered digital scheme alleviates the cliff effect through the combination of a layered video coding and a layered video transmission. Typically, scalable video coding (SVC) [3] is utilized as the layered video coding technique, and hierarchical modulation (h-mod) [4] is adopted as the layered transmission scheme. However, the layered digital scheme does not solve the cliff effect completely due to the limited number of layers, which actually generates the staircase effect.

Recently, SoftCast [2] has been proposed as a new framework to deal with the aforementioned problems, particularly the cliff effect. SoftCast sends real numbers

*Correspondence: wah_ty@163.com

¹Institute of Digital Media and Communication, Taiyuan University of Science and Technology, No. 66, Waliu Road, Taiyuan, Shanxi, China
Full list of author information is available at the end of the article

instead of a digital bitstream by using a very dense constellation, and thus it is called an analog transmission. Specifically, by discarding entropy coding and channel coding, SoftCast consists of three steps: block-wise discrete cosine transform (DCT), power allocation, and whitening. Firstly, DCT removes the spatial correlation within each video frame; secondly, power allocation generates a compact and resilient representation of DCT coefficients; and finally, whitening generates the equal-importance packets that will be transmitted directly over a wireless channel. In SoftCast, reconstruction quality at each user only depends on its channel characteristic (including packet loss rate and channel SNR) so that receivers with a good channel condition can obtain better video quality, while users with a bad channel condition can still watch a lower quality. Experimental results showed that SoftCast is more robust to channel noise and achieves a smooth degradation of quality.

Several approaches have been proposed to improve the performance of SoftCast in the past few years, such as D-cast [5] and Wave-cast [6], by making use of the inter-frame correlation of video signals to remove the temporal correlation. Furthermore, Hybrid-Cast [7] utilizes a hybrid transmission that combines the digital transmission of important information (such as the motion vector and scalar factor) and SoftCast's analog transmission of the other information (including the DCT coefficients). However, since the packets in Hybrid-Cast are not equally important, the performance may not be as good as expected when the important packets are lost. Meanwhile, there are several works appearing in the area of soft video transmission. For instance, the bandwidth expansion problem of soft video coding is solved by layered coset coding [8]; a gradient-based framework is proposed in [9] for wireless soft video broadcast; the compressive sensing (CS) is integrated into multiple-input multiple-output (MIMO) transmission to make sure that the reconstructed image/video quality is commensurate with the channel SNR and the MIMO dimension [10, 11]; the multipath case of the SoftCast is investigated in [12] in which high-energy DCT coefficients are assigned to a "good" subcarrier; and a real soft video broadcast system has been implemented and some PHY layer issues have been solved in [13].

The CS technique seems very suitable for wireless transmission (with random packet loss) due to its random measurement. One simple wireless video multicast framework based on CS has been presented in [14], which explores the random measurement to generate equal-importance packets, consequently eliminating the cliff effect and obtaining a graceful degradation. As for the CS technique, a few algorithms have been widely applied for image and video coding, such as the structurally random matrices (SRMs) and block-based compressed sensing

(BCS) [15, 16]. BCS is specially tailored to maintain a low computational burden. However, BCS is not efficient in compression because it wrecks the global random measurement. Then, MS-BCS-SPL [17] (multi-scale BCS with smooth projection Landweber) explores BCS for different levels of a wavelet-decomposed image and consequently improves the performance of BCS greatly, while retaining the low computational burden.

In this paper, we extend the state-of-the-art MS-BCS-SPL to the case of multiple users and propose a multi-scale CS-based wireless video multicast scheme: MCS-cast. Specifically, each video frame is decomposed by a multi-scale discrete wavelet transform (DWT). Then, the optimally determined CS rates are allocated to different DWT levels with an attempt of sampling the more important DWT levels with higher measurement rates. All achieved measurements are packed in such a way that all packets are equally important and each packet includes different percentages of different DWT levels. Finally, the packets are transmitted through a physical technique like SoftCast, mapping the measurements into a very dense constellation. We will demonstrate that (1) the cliff effect is avoided naturally because of application of the equal-importance packing strategy and (2) packet loss results in a (much) reduced influence on the reconstruction quality because of larger percentages of more important levels in each packet. At the same time, the linear least square estimator (LLSE) is incorporated to eliminate channel noise, thus leading to a further improved performance in reconstruction quality.

2 Related works

Our work is strongly related to the theory of CS. Therefore, we first review it briefly and then discuss two extended versions of it: BCS-SPL and MS-BCS-SPL.

2.1 Conventional CS

Suppose that a signal x with length N can be represented via a known basis $\Psi \in R^{N \times N}$ (the inverse transform), i.e.,

$$x = \sum_{i=1}^N \psi_i \vartheta_i = \Psi \Theta \quad (1)$$

where $\Theta \in R^N$ is the coefficient vector with K significant values in the transform domain, $K \ll N$. Then, we project x onto an M -dimensional space using a measurement matrix $\Phi \in R^{M \times N}$:

$$y = \Phi x = \Phi \Psi \Theta \quad (2)$$

where y is called measurements, and the CS sampling rate is M/N , $M = O(K \log(N/K))$. Since $M \ll N$, recovering $x \in R^N$ from $y \in R^M$ is an ill-posed problem. Nevertheless, provided that x is sparse enough, the

CS theory can solve it by an l_0 -constrained optimization problem:

$$\hat{\theta} = \arg \min_{\theta} \|\theta\|_{l_0} \quad \text{s.t.} \quad y = \Phi x \quad (3)$$

In practice, such a l_0 -constrained optimization is suffering from its computational infeasibility. Then, CS turns to solve an l_1 -based convex optimization, and the recovery is implemented directly by exploiting linear programming. Once $\hat{\Theta}$ is obtained, we can obtain x as $\hat{x} = \Psi \hat{\Theta}$.

2.2 BCS-SPL

In BCS-SPL, an original image with N pixels and needing M measurements is first divided into blocks of size $B \times B$. Then, each block is CS-sampled using the same matrix Φ_B :

$$y_j = \Phi_B \cdot x_j = \Phi_B \cdot (\Psi_B \cdot \Theta_j) \quad (4)$$

where Θ_j and y_j are, respectively, the transform coefficients and measurement vector of the j -th block; y_j is with size $M_B \times 1$; $M_B = \lfloor \frac{M}{N} \cdot B^2 \rfloor$; $\Psi_B \in R^{B^2 \times B^2}$; and $\Phi_B \in R^{M_B \times B^2}$. Here, Φ_B is chosen to be orthonormal, i.e., $\Phi_B \cdot \Phi_B^T = 1$. The equivalent measurement matrix Φ for the entire image is a block-wise diagonal one:

$$\Phi = \begin{bmatrix} \Phi_B & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \Phi_B \end{bmatrix} \quad (5)$$

Since BCS-SPL is based on block-wise image acquisition, only the measurement matrix Φ_B needs to be stored, which greatly saves the storage space and improves the reconstruction speed. In addition, BCS-SPL combines a Wiener filter to the SPL algorithm, thus offering a much smoothed reconstruction. Afterwards, some state-of-the-art block-based CS reconstructions for images and videos, e.g. [18] and [19], were designed, and they have shown improvement over the BCS-SPL.

2.3 MS-BCS-SPL

In MS-BCS-SPL, the sampling operator Φ is divided into two parts, a multi-scale wavelet transform matrix Ψ' and a multi-scale block-wise measurement matrix Φ' such that $\Phi = \Phi' \Psi'$, and Eq. (2) now becomes

$$y = \Phi' \Psi' x \quad (6)$$

Suppose that Ψ' generates L level wavelet decomposition. Then, the measurement matrix Φ' , composed of L different block-wise sampling operators, is adopted to these decompositions with one operator for each level. Regarding Θ as the expression of x in the transform domain, i.e., $\Theta = \Psi' x$, the sub-band s at level l of Θ can be divided into $B_l \times B_l$ blocks and each level will be measured through the appropriately sized Φ_l . Assume that $y_{l,s,i}$ is the measurement of block i of sub-band s at level l of Θ ,

with $s \in \{H, V, D\}$ (standing for horizontal, vertical, and diagonal, respectively), $0 \leq l \leq L - 1$, then,

$$y_{l,s,i} = \Phi_l \Theta_{l,s,i} \quad (7)$$

Notice that we apply the same measurement matrix Φ_l to three sub-bands (H, V, and D) at the same level l .

Next, an optimized CS rate r_l is adjusted for each level l in MS-BCS. Specifically, the rate of the baseband level₀ (LL0 sub-band) is set to be full, i.e., $r_0 = 1$ (because of the highest importance), and the rate for level l ($l \neq 0$) is selected as $r_l = w_l r'$ so that the overall rate becomes

$$r = \frac{1}{4^{L-1}} r_0 + \sum_{l=1}^{L-1} \frac{3}{4^{L-1}} w_l r' \quad (8)$$

If a target rate r and the weights w_l on each level are given, r' can be determined by Eq. (8), leading to a set of level-wise CS rates r_l .

During the MS-BCS-SPL reconstruction, an iteration process called the Landweber step [20, 21] lies between the smoothing and thresholding operations in the wavelet domain. First, the constrained optimization formulation is replaced by an unconstrained optimization problem via a Lagrangian multiplier with an l_2 -distance penalty:

$$\hat{\Theta} = \arg \min_{\theta} \|\tilde{\Theta}\|_{l_1} + \lambda \|\tilde{y} - \Phi' \tilde{\Theta}\|_{l_2} \quad (9)$$

Then, $\hat{\Theta}$ is recovered by the following successive projection and thresholding operations, assuming $\Theta^{(0)}$ is the initial approximation to the wavelet coefficients Θ :

$$\hat{\Theta}^{(i)} = \Theta^{(i)} + \frac{1}{r} \Phi^T (\tilde{y} - \Phi' \Theta^{(i)}) \quad (10)$$

$$\theta^{(i+1)} = \begin{cases} \hat{\theta}^{(i)}, & |\hat{\theta}^{(i)}| \geq \mathcal{J}^{(i)} \\ 0, & \text{else} \end{cases} \quad (11)$$

where r is a scaling factor, and $\mathcal{J}^{(i)}$ is the threshold that is used at the i -th iteration.

3 The proposed scheme

In our work, we extend MS-BCS to wireless video multicast with the aim of getting some gains over the SoftCast scheme, while preserving the property of accommodating heterogeneous users and graceful degradation of quality. To this end, we should meet the following requirements as much as possible: Firstly, a video source should be divided into packets with an equal importance. With this requirement, the video quality decoded at each user becomes independent of which packets are received; rather, it depends only on how many packets are received (according to its channel condition). Secondly, more important coefficients should occupy a higher percentage in each packet so that the reconstructed quality would still be acceptable even with the low-quality users who only receive a small number of packets. Thirdly, the overall

system should consider the channel noise, and the reconstruction quality should smoothly change proportionally to the packet loss ratio and channel signal-to-noise ratio (CSNR).

Our framework is shown in Fig. 1, where we have made use of the MS-BCS-SPL scheme. Specifically, in the encoder side, each frame is first transformed by DWT and the resulting coefficients are measured by multiplying a multi-scale measurement matrix Φ' , the same as that in Eq. (6). The measurements are packed and transmitted through a raw orthogonal frequency division multiplexing (OFDM) with an analog-like modulator to the decoders to provide wireless multicasting service. At the decoder side, the received packets at each user—the number of packets varies among different users—are first de-noised by the linear least square estimation (LLSE) algorithm, and then, each user will reconstruct a frame using the de-noised measurements.

3.1 Encoder side

Preprocessing. In our scheme, a preprocessing is performed to each frame before encoding in order to keep a low energy. In the original MS-BCS-SPL scheme, the image is first subtracted by its mean value (denoted as E_0) of all pixels included. At the decoder side, the mean value will be added back to the image after recovering. However, in the environment of wireless video multicast, the mean value will probably be lost during transmission over a noisy and lossy channel, thus leading to a very in-accurate reconstruction. To solve this problem, we propose to set E_0 to 128 (for 8-bit video frames) at the preprocessing step. Clearly, this constant (mean) value can be compensated back at the decoder side (regardless of how noisy and lossy the involved channel is, because it is not needed to be transmitted at all).

DWT. In our scheme, each frame is sampled and recovered in the wavelet domain, following what has been done in MS-BCS-SPL [17], where a dual-tree DWT (DDWT) as [22] with bivariate shrinkage [23] is applied within the DDWT domain to enforce sparsity as described in [16].

To this end, we propose to decompose each frame (after preprocessing) into L DWT levels (as shown in Fig. 2 where $L = 4$). After the DWT decomposition, it is clear that the DWT coefficients have successively decreasing importance at higher decomposition levels.

Measuring with different CS rates. Since the coefficients in different DWT levels have different importance towards the reconstruction quality, we need to apply different CS rates (for measurement) that depend on the importance at each DWT level. This means that the more important the level is, the larger measurement CS rate will be allocated to it. To this end, we first divide the sub-bands of each DWT level into blocks. We choose different sizes at different levels: the size becomes increasingly larger from level₁ to level₃. Then, we will measure the sub-bands at different DWT levels with different measurement matrices corresponding to the block size in each level. For example, if four DWT levels are used (the same as in Fig. 2) and the block sizes 4×4 , 8×8 , and 16×16 are adopted, respectively, for level₁ ~ level₃, we may select the corresponding measurement matrices to be of sizes 32×16 , 96×64 , and 176×256 , respectively, i.e., $\Phi_1 \in R^{32 \times 16}$ at level₁, $\Phi_2 \in R^{96 \times 64}$ at level₂, and $\Phi_3 \in R^{176 \times 256}$ at level₃. Such arrangement means measurement rates of 200, 150, and 68.75 %, respectively. The over-measured CS data seem completely redundant at this moment (the corresponding measurement rate at level₀ will be even larger; see discussions in the next sub-section). Nevertheless, it will be pointed out later that they are necessary when packing these CS data into packets.

Packing with optimized CS rate allocation. Packing of the CS measurements (in the DWT domain) is performed in our work in such a way that all packets have an equal importance. As a result, the reconstruction quality at each user in the multicast group depends only on how many packets are received, regardless of what packets are received. On the other hand, since the measurements from different DWT levels have different importance toward the reconstruction quality, we select different percentages for them in each packet.

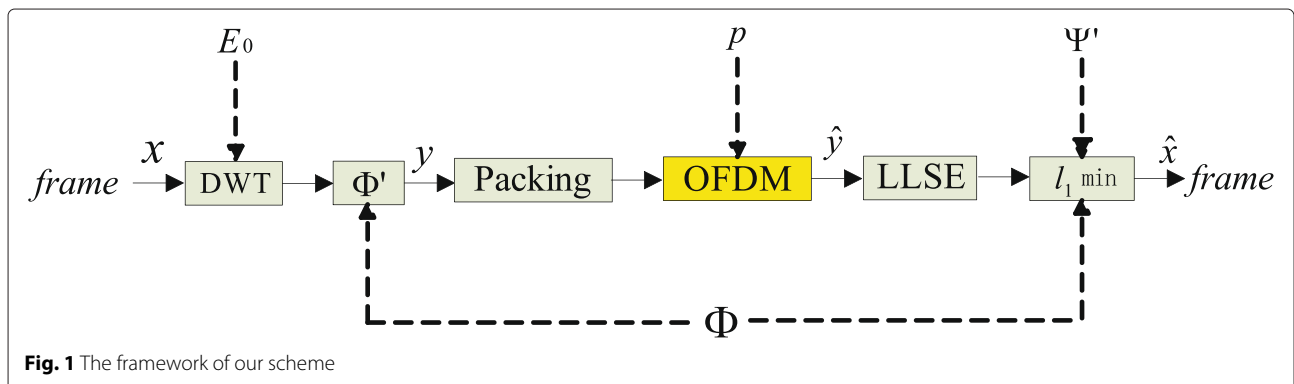


Fig. 1 The framework of our scheme

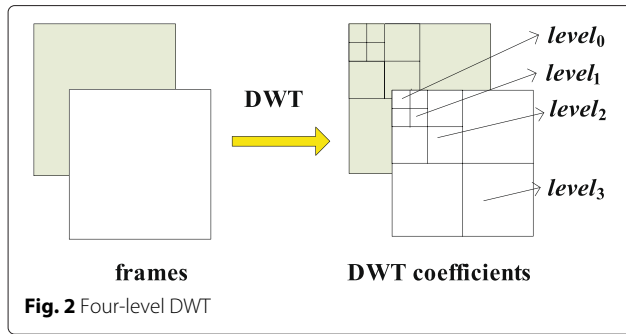


Fig. 2 Four-level DWT

Suppose that (i) the width and height of each frame are W and H , respectively; (ii) the total number of packets is k ; and (iii) the overall rate for all measurements over each frame is set at full (100%). In our scheme, we first arrange the whole baseband level₀ into each packet (as shown in Fig. 3 where $L = 4$), because it has the highest importance and losing it would lead to a sharply degraded reconstruction quality. Clearly, this repetition produces an over-sampling (by k times); nevertheless, it guarantees a minimum quality even when a user receives only one packet. On the other hand, the remaining levels need to use some well-determined CS rates so as to achieve the overall rate at 100%.

Suppose that the CS rate for level _{l} is r_l , with $l = 0, \dots, L - 1$, and $r_l \geq r_{l+1}$. To guarantee the full overall rate, we have

$$\frac{W \times H}{4^{L-1}}k + \sum_{l=1}^{L-1} 3 \frac{W \times H}{4^{L-1}}r_l = W \times H \quad (12)$$

which can be simplified as

$$\frac{1}{4^{L-1}}k + \sum_{l=1}^{L-1} 3 \frac{1}{4^{L-1}}r_l = 1 \quad (13)$$

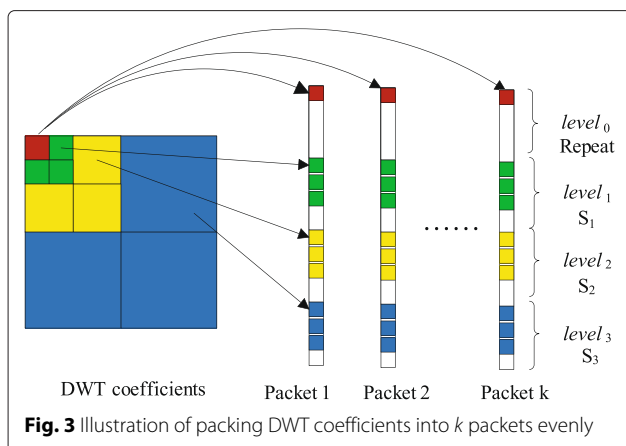


Fig. 3 Illustration of packing DWT coefficients into k packets evenly

Notice that Eq. (13) may not always be met exactly. In this case, we will choose r_l properly to make Eq. (13) hold as much as possible.

Once all r_l 's are obtained (with r_0 being always set at 1), each frame within a group of pictures (GOP) will be CS-sampled at the determined CS rates in the corresponding DWT levels. We perform the packing process on all frames within each GOP to generate a total number of k packets: the baseband level₀ is put into each packet, whereas the measurements for level₁ (the total number is $3 \frac{W \times H}{4^{L-1}}r_1$) is evenly put into k packets, i.e., $3 \frac{W \times H}{4^{L-1}}r_1/k$ CS data are put into each packet. Notice that because measurements at level _{l} are over-complete, we need to select $3 \frac{W \times H}{4^{L-1}}r_l/k$ CS data carefully so that they are as independent as possible with respect to each other.

3.2 Raw OFDM channel

Before packets are transmitted over a raw OFDM channel [24], the measurements in each packet are rounded and then directly mapped into the transmitted symbol, whereas no FEC of any kind is employed. Figure 4 shows the modulation adopted in our work: $P_s^{[k]}$ and $P_s^{[k+1]}$ are the k -th and $(k + 1)$ -th data in the s -th packet, and such pairs of data are directly mapped as the I and Q components of the transmitted symbol. Finally, the PHY layer directly transmits all symbols over OFDM channels in which we will consider different strengths of channel noise in our experimental results.

Figure 5 shows the overall OFDM channel structure in which the modulation is the same as that shown in Fig. 4. At the transmitter side, symbols obtained from modulation are inputted into some sub-bands after the serial-parallel conversion. Symbols in each sub-band go through Inverse Fast Fourier Transform (IFFT), guard interval insertion, and the parallel-serial conversion to get the OFDM signal. Then, the OFDM signal is transmitted over a wireless channel with additive white Gaussian noise (AWGN). At the receiver side, operations that are opposite to what have been done at the transmitter side

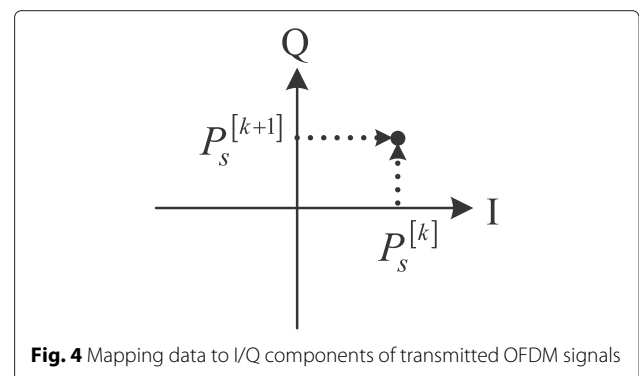
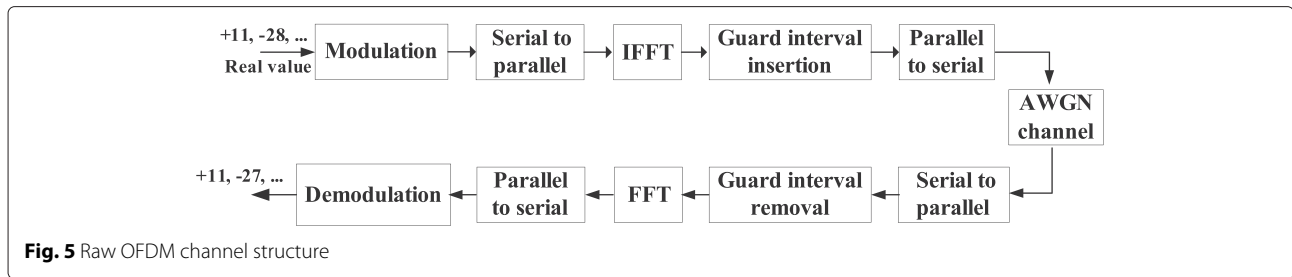


Fig. 4 Mapping data to I/Q components of transmitted OFDM signals



are carried out. This whole procedure is the same as in the SoftCast scheme.

3.3 Decoder side

LLSE. When packets are transmitted over a noisy channel, channel noise is added to packets and thus may incur a certain deviation from the original data, which may badly influence the reconstruction quality. Here, we propose to apply LLSE [25] to the received data before the MS-BCS-SPL reconstruction.

First, assuming that all packets are received (i.e., no packet loss) but with channel errors corrupting them, we can rewrite the received signal as

$$\hat{y} = y + n \quad (14)$$

where n is the additive white Gaussian noise. Then, LLSE estimates the original signal as

$$y_{\text{LLSE}} = \Lambda_y (\Lambda_y + \Sigma)^{-1} \hat{y} \quad (15)$$

where y_{LLSE} refers to the LLSE estimate of measurement y , Λ_y is the covariance matrix of y (which will be transmitted as metadata), and Σ is the covariance matrix of channel noise n . With a high channel SNR (CSNR), we obtain an approximation as

$$y_{\text{LLSE}} \approx \Lambda_y (\Lambda_y)^{-1} \hat{y} = \hat{y} \quad (16)$$

This means that the LLSE step becomes void, which is reasonable because the measurements are trustable nearly completely.

Next, when a receiver experiences certain packet loss, let us define \hat{y}_* as \hat{y} after removing all lost packets, and similarly n_* as the corresponding noise vector, and we still have

$$\hat{y}_* = y_* + n_* \quad (17)$$

Then, the LLSE decoder becomes

$$y_{\text{LLSE}} = \Lambda_{y_*} (\Lambda_{y_*} + \Sigma_*)^{-1} \hat{y}_* \quad (18)$$

Different measurement matrices. Since different users are connected with different bandwidths in the same multicast group, they receive different numbers of packets. Consequently, after LLSE, each user needs to use its own measurement matrix for each level according to the packets it receives. Suppose that the encoder uses a random

matrix $\Phi \in \mathbb{R}^{N_l}$ (which can be repeated exactly at the decoder side) to generate the measurements for level l and one user just receives M_l from N_l measurements, then the corresponding measurement matrix used at the decoder side for reconstruction can be obtained as

$$\Phi_l^T = \{(\Phi^T)_i \mid i \in \{1, \dots, M_l\}\} \quad (19)$$

where i is the row index of Φ and can be obtained from the packet index.

MS-BCS-SPL reconstruction. MS-BCS-SPL provides a multi-scale reconstruction by deploying block-based CS sampling within the wavelet domain, which applies the Landweber step to each block in each sub-band at each decomposition level independently. Hence, the reconstruction $x_{l,s,j}$ for block j of sub-band s at level l can be expressed as

$$\tilde{x}_{l,s,j} = x_{l,s,j} + \Phi_l^T (y_{l,s,j} - \Phi_l x_{l,s,j}) \quad (20)$$

where Φ_l represents the block sampling operator of level l .

3.4 Transmission of metadata

As shown in Eq. (15), the decoder requires the covariance matrix of measurement y . To this end, we transmit the standard deviations as metadata so that the covariance matrices can be calculated at the decoder from the received standard deviations. In our MS-cast, there are several standard deviations for each frame. For the video sequence of 352×288 at 30 Hz with four-level wavelet decomposition, we will place the measurements into several packets (date matrix) with size 64×1584 (this is for the convenience of subsequent 64-point IFFT). Then, for each frame, there are 1 standard deviation from the vectorized level₀ components of size 44×36 (1584) with repeated full measurement, 6 (3 × 200 %) standard deviation from three vectorized level₁ components of size 44×36 with measurement rate $r_1 = 200$ %, 18 (3 × 4 × 150 %) standard deviation from three vectorized level₂ components of size 88×72 with measurement rate $r_2 = 150$ %, and 33 (3 × 16 × 68.75 %) standard deviation from three vectorized level₃ components of size 176×144 with measurement rate $r_3 = 68.75$ %. In total, there are 58 standard deviations for 58 vectors of size 1584 that need to be transmitted as metadata for each frame of size 352×288 .

The transmission of the standard deviation is through the traditional communication scheme consisting of entropy coding, channel coding, and modulation. The standard deviations are quantized by a 32-bit scalar quantizer and compressed by entropy coding, and then further coded using the 1/2 convolutional code (with generator polynomials {133, 171}) and BPSK constellation. This forms the metadata packet. Hence, the percentage of this metadata is about $58 \times 32 \times 2/352/288 = 3.66\%$. For this extra percentage of metadata, we just cut the measurement rate of level 3 to ensure the final equivalent rate of each frame to be full (100%).

4 Simulation results

We evaluate our MCS-cast under two kinds of channel models: additive white Gaussian noise (AWGN) channel and noiseless channel. In both cases, the measurements are transmitted directly using the analog-like modulation which maps measurements into a very dense constellation. The packets are erased randomly with packet loss ratio p so that the number of measurements received by the decoder is $M = (1 - p)k$, and k is the total number of packets. The channel signal-to-noise ratio (CSNR) is defined as

$$\text{CSNR} = \|y\|_2^2 / \|n\|_2^2$$

Notice that although we did not consider a real multicast scenario that needs to define exactly how many users are included in the multicast group, it can be mimicked closely by allowing different packet loss ratios and channel noise because each ratio/noise represents a specific user.

In our experiments, the CIF (352×288 at 30 Hz) video sequences of *Football*, *Foreman*, *Coastguard*, *Hall*, and *Container* are used, and the full measurement rate is assumed. After the DWT decompositions of four levels ($L = 4$), except for the baseband, three remaining levels of each frame undergo block-based projection like that in MS-BCS. Following the example we discussed earlier in the last section, the block sizes are selected as 4×4 , 8×8 , and 16×16 , respectively, for level₁ ~ level₃. Then, we apply random measurement matrices $\Phi_1 \in R^{32 \times 16}$ to level₁, $\Phi_2 \in R^{96 \times 64}$ to level₂, and $\Phi_3 \in R^{176 \times 256}$ to level₃, corresponding to $r_1 = 200\%$, $r_2 = 150\%$, and $r_3 = 68.75\%$ CS rates, respectively. The total number of packets in the simulation is set to $k = 8$. It can be verified that the overall rate is $\frac{8}{4^3} + \sum_{l=1}^3 \frac{3}{4^{4-l}} r_l = \frac{32.5}{32}$, which exceeds 100% a little bit. This is a very minor problem, and we can fix it easily by, for instance, cutting the measurement rate at level₃ to 66.67% (corresponding to about 170 CS data, instead of 176 in the original setting). Such a full rate (over the whole frame) has been chosen because it is also used in the SoftCast scheme so that the comparison is quite fair. SoftCast and our schemes have the

same transmitting power and use the same wireless bandwidth of 1.5 MHz considering that the number of pixels of the CIF video signal is about 3 M per second. Since we transmit complex symbols, this should require a channel bandwidth of about 1.5 MHz. Sixty-four subcarriers are used for OFDM, and the prefix duration is 16. Each packet consists of a data matrix of size 64×1584 . After a 64-point IFFT, each row of data matrix is assigned to a subcarrier of OFDM randomly. Hence, the number of OFDM symbols in each subcarrier is $1584/2 = 792$. The packets are separated by the indices of packets, and the index is inserted before each packet. We did not consider the subcarrier for pilot data and the active carriers. These experimental conditions are also conducted on the SoftCast scheme.

Two groups of experiments are conducted. The first one is to show the performance of LLSE in our framework. The second compares our scheme with SoftCast under the same CSNR and packet loss ratio.

Figure 6 shows that MCS-cast with LLSE in the decoder consistently outperforms that without LLSE. This is because LLSE overcomes the negative influence of channel noise to a certain degree, especially in the case of a low packet loss ratio where the gain can be as large as 7 dB. However, as the packet loss ratio increases, the gain obtained from LLSE becomes less but still quite noticeable (especially when CSNR is low).

Figure 7 shows that our MCS-cast scheme consistently works better than the state-of-the-art SoftCast (both schemes have employed LLSE) when the packet loss ratio is more than 0.2 (for *Football*, *Foreman*, *Hall*, and *Container*) or 0.3 (for *Coastguard*). This also indicates that our scheme has better bandwidth heterogeneity. For example, the case of packet loss rate " $p = 0$ " may mimic the user's bandwidth of 1.5 MHz, " $p = 10\%$ " mimics bandwidth of $1.5 \text{ MHz} \times 90\% = 1.35 \text{ MHz}$, " $p = 20\%$ " mimics $1.5 \text{ MHz} \times 80\% = 1.2 \text{ MHz}$, and so on. It is because our scheme considers the different importance of decomposition levels and takes more measurements from these more important decomposition levels. More specifically, MCS-cast will preserve the baseband as long as one packet is received (which is nearly always true in practice), whereas other relatively more important information (e.g., at level₁ and level₃) are also likely to be received due to 200 and 150% CS rates used in our scheme. On the other hand, in the case of no packet loss, the repeated baseband into all packets brings redundancy to our scheme, thus resulting in a lower efficiency than the referenced SoftCast.

Finally, Fig. 8 shows the visual result comparison of MCS-cast with SoftCast for frame #70 of five test sequences at a CSNR of 25 dB, from which we can observe clearly that the visual results of our MS-cast scheme are better than those of SoftCast when the packet loss ratio increases above a certain level.

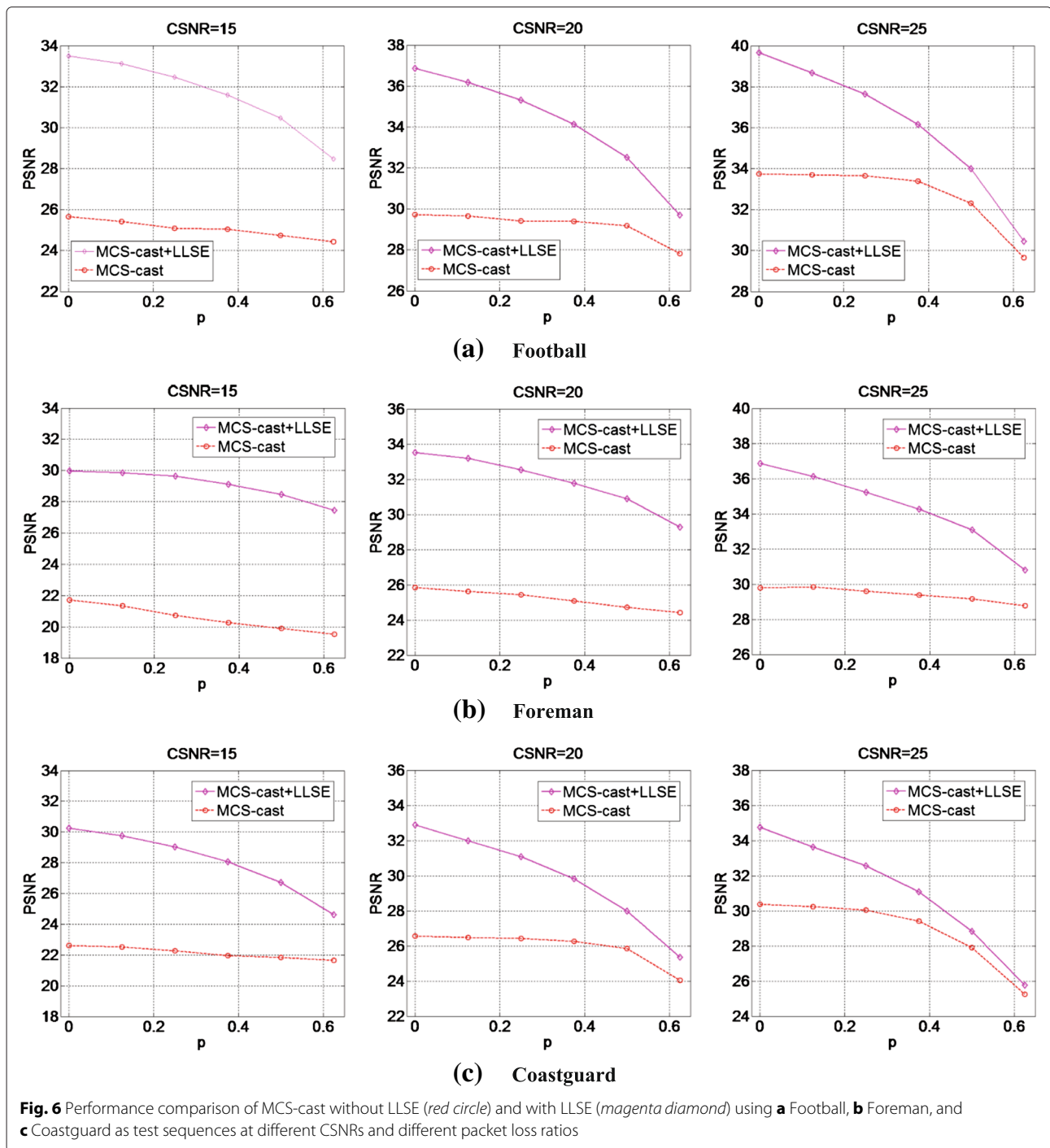


Fig. 6 Performance comparison of MCS-cast without LLSE (red circle) and with LLSE (magenta diamond) using **a** Football, **b** Foreman, and **c** Coastguard as test sequences at different CSNRs and different packet loss ratios

One remark is necessary before we conclude the paper: although we did not consider a real multicast scenario that needs to define exactly how many users are included in the multicast group, it has been mimicked closely by allowing different packet loss ratios and channel noise levels, because each ratio/noise combination truly represents a specific user. As these combinations can be many, our MCS-cast becomes fully scalable in serving an arbitrary number of users in a multicast group or even multiple

multicast groups, where each individual user receives a number of noise-corrupted packets (depending on its channel conditions) and then runs its own reconstruction independently.

5 Conclusions

In this paper, we proposed a multi-scale compressed sensing-based wireless video multicast: MCS-cast. The reconstruction quality of MCS-cast depends only on the

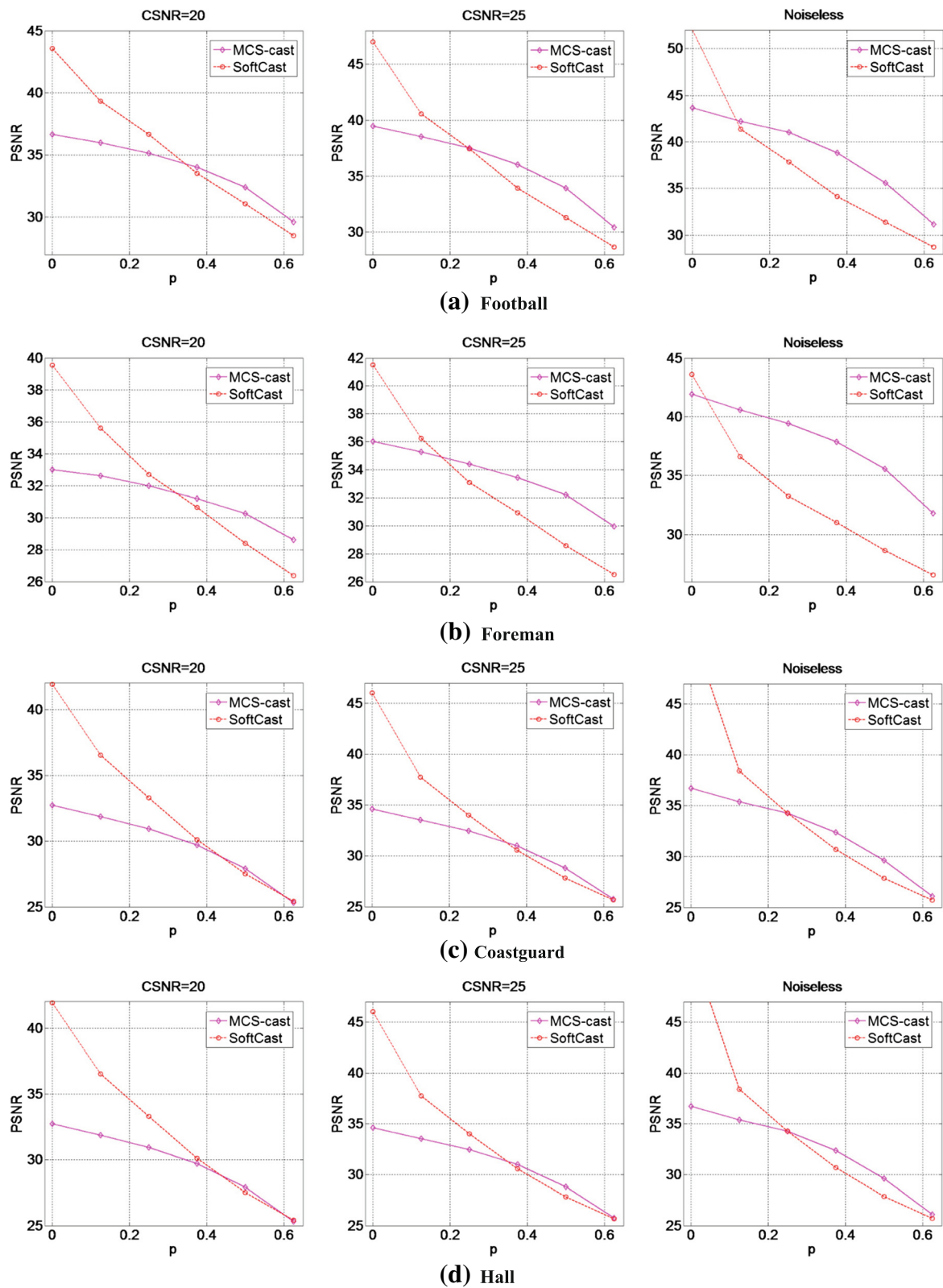
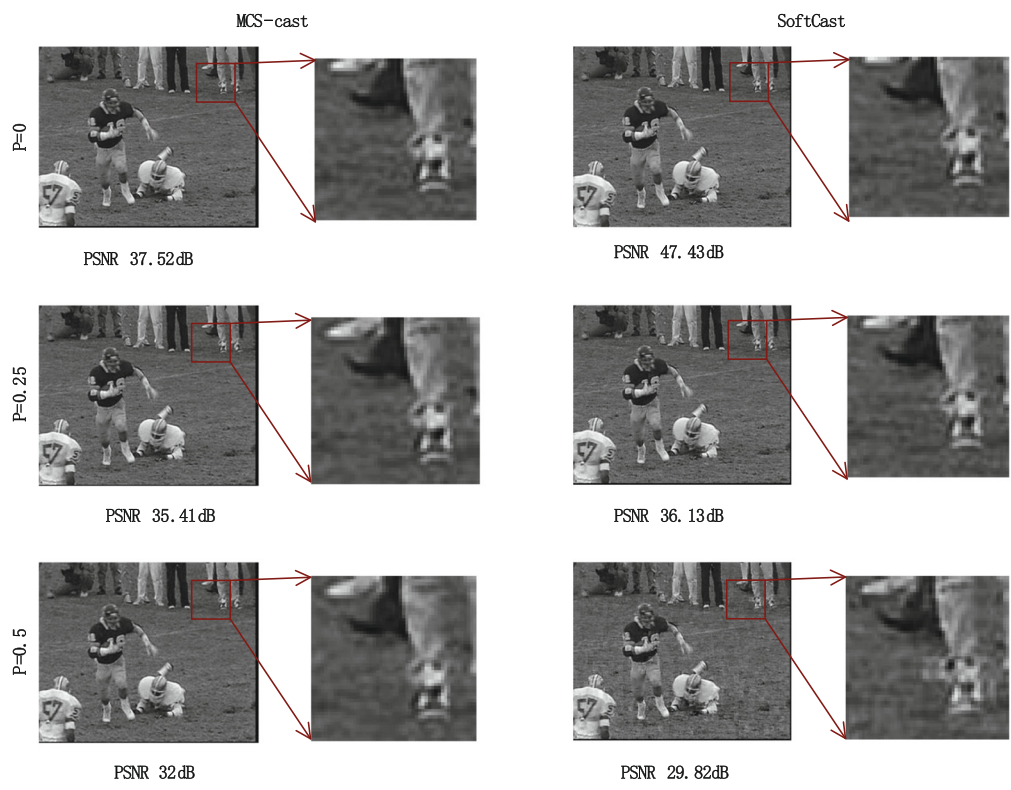
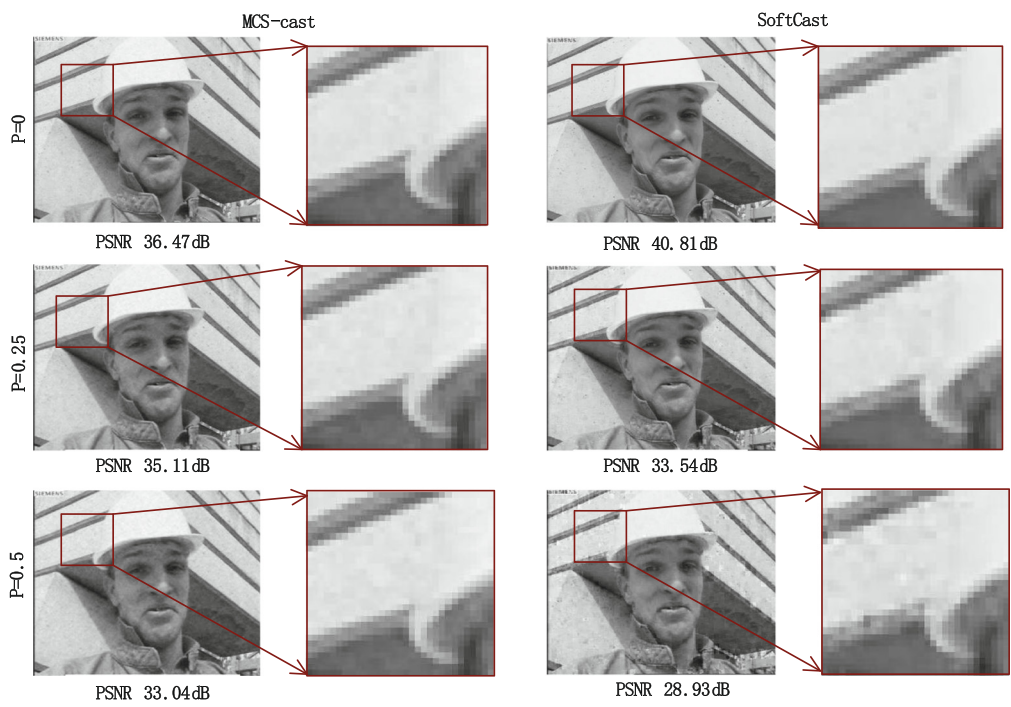


Fig. 7 Comparison of MCS-cast (magenta diamond) with SoftCast (red circle) using the three test sequences. These graphs show results of PSNR in different noise strengths and different packet loss ratios. **a** Football, **b** Foreman, **c** Coastguard, and **d** Hall



(a) Football



(b) Foreman

Fig. 8 Visual result comparison of MCS-cast with SoftCast for Frame #70 of the two test sequences at CSNR = 25 dB. a Football. b Foreman

number of packets received by each user. We further proposed a novel packing strategy such that all packets are equally important and each packet includes different percentages of measurements from different wavelet decomposition levels. Due to the equal-importance feature in various packets and the direct transmission (without entropy coding and channel coding), MCS-cast does not suffer from the cliff effect and the reconstruction quality is only degraded gracefully when channel noise and/or packet loss is considered. Meanwhile, larger CS rates used at more important DWT levels guarantee that these important coefficients are still likely to be received at a user's side even with a large packet loss ratio so that the reconstruction quality remains quite acceptable. These advantages have been clearly demonstrated in our experiments. As a future work, we will be focusing on how to utilize the correlation among adjacent frames in our MCS-cast scheme so as to make a further improvement.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work has been supported in part by the National Natural Science Foundation of China (No. 61272262 and No. 61210006), Shanxi Provincial Foundation for Leaders of Disciplines in Science (20111022), Shanxi Province Talent Introduction and Development Fund (2011), Shanxi Provincial Natural Science Foundation (2012011014-3), the program of "One Hundred Talented People" of Shanxi Province, Research Project Supported by Shanxi Scholarship Council of China (2014-056), and Program for New Century Excellent Talent in Universities (NCET-12-1037).

Author details

¹Institute of Digital Media and Communication, Taiyuan University of Science and Technology, No. 66, Waliu Road, Taiyuan, Shanxi, China. ²Georgia Institute of Technology, 225 North Avenue NW, Atlanta, GA, USA. ³Institute of Image Processing, University of Electronic Science and Technology of China, No. 2006 Xiyuan Ave., Chengdu, Sichuan, China. ⁴Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong, China.

Received: 28 November 2014 Accepted: 28 July 2015

Published online: 14 August 2015

References

- U Reimers, Digital video broadcasting. *IEEE Commun. Mag.* **36**, 104–110 (1998)
- SoftCast, One video to serve all wireless receivers. <http://dSPACE.mit.edu/handle/1721.1/44585>
- Wu YTH, YQ Zhang, Scalable video coding and transport over broadband wireless networks. *Proc. IEEE* **89**, 1–20 (2001)
- S Wang, BK Yi, in *Proceedings of IEEE Globecom*. Optimizing enhanced hierarchical modulations. New Orleans 1–4 Dec. 2008 1–5
- X Fan, F Wu, D Zhao, OC Au, Distributed wireless visual communication with power distortion optimization. *IEEE Trans. Circ. Syst. Vi. Technol.* **23**, 1040–1053 (2013)
- X Fan, R Xiong, D Zhao, F Wu, in *Proceedings of IEEE Visual Communication and Image Processing*. Wavecast: wavelet based wireless video broadcast using lossy transmission. San Diego 27–30 November 2012, 1–6
- L Yu, H Li, W Li, Wireless scalable video coding using a hybrid digital-analog scheme. *IEEE Trans. Circ. Syst. Vi. Technol.* **23**, 331–345 (2013)
- X Fan, R Xiong, D Zhao, F Wu, *Layered soft video broadcast for heterogeneous receivers*. (IEEE Transactions on Circuits and Systems for Video Technology, 2015). DOI 10.1109/TCSVT.2015.2402831, (in press)
- R Xiong, H Liu, S Ma, X Fan, F Wu, W Gao, in *Data Compression Conference (DCC)*. G-cast: gradient based image SoftCast for perception-friendly wireless visual communication. Snowbird, 26–28 March 2014, 133–142
- XL Liu, C Luo, W Hu, F Wu, in *INFOCOM, Orlando*. Compressive broadcast in MIMO systems with receive antenna heterogeneity. 25–30 March 2012, 3011–3015
- XL Liu, W Hu, C Luo, F Wu, Compressive image broadcasting in MIMO systems with receiver antenna heterogeneity. *Signal Process. Image Commun.* **29**, 361–374 (2014)
- H Cui, C Luo, CW Chen, F Wu, in *Proceedings of the IEEE INFOCOM*. Robust uncoded video transmission over wireless fast fading channel. Toronto, 1–2 May 2014, 73–81
- S Jakubczak, D Katabi, SoftCast: one-size-fits-all wireless video. *ACM SIGCOMM Comput. Commun. Rev.* **41**, 449–450 (2011)
- MB Schenkel, C Luo, F Wu, P Frossard, in *Proceedings of Visual Communication and Image Processing*. Compressed sensing based video multicast. Huangshan, 11–14 August 2010, 1–9
- L Gan, in *Proceedings of the International Conference on Digital Signal Processing*. Block compressed sensing of natural images. Cardiff, 1–4 July 2007, 403–406
- S Mun, JE Fowler, in *Proceedings of the IEEE International Conference on Image Processing*. Block compressed sensing of images using directional transforms. Cairo, 24–26 November 2009, 3021–3024
- E Fowler, S Mun, EW Tramel, in *Proceedings of the European Signal Processing Conference*. Multiscale block compressed sensing with smooth projected Landweber reconstruction. Barcelona, 29–31 August 2011, 564–568
- C Chen, EW Tramel, JE Fowler, in *Proceedings of the 5th Asilomar Conference on Signals, Systems, and Computers*. Compressed-sensing recovery of images and video using multihypothesis predictions. Pacific Grove, 8–11 November 2011, 1193–1198
- EW Tramel, JE Fowler, in *Proceedings of the IEEE Data Compression Conference*. Video compressed sensing with multihypothesis. Pacific Grove, 29–31 March 2011, 193–202
- WQ Yang, DM Spink, TA York, H McCann, An image-reconstruction algorithm based on Landweber's iteration method for electrical-capacitance tomography. *Meas. Sci. Technol.* **10**, 1065–1069 (1999)
- T Blumensath, ME Davies, Iterative thresholding for sparse approximations. *J Fourier Anal. Appl.* **14**, 629–654 (2008)
- N Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. harmonic analysis.* **10**, 234–253 (2001)
- L Sendur, IW Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Trans. Sig. Process.* **50**, 2744–2756 (2002)
- LJ Cimini, Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.* **33**, 665–675 (1985)
- CL Lawson, RJ Hanson, *Solving Least Squares Problem*. (Society for Industrial and Applied Mathematics (SIAM), 1974)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com