

Compressed Sensing Based Video Multicast

Markus B. SCHENKEL^{a,b}, Feng WU^a, Chong LUO^a and Pascal FROSSARD^b

^a Microsoft Research Asia, Beijing, China

^b Signal Processing Laboratory (LTS4), EPFL, Lausanne, Switzerland

ABSTRACT

We propose a new scheme for wireless video multicast based on compressed sensing. It has the property of graceful degradation and, unlike systems adhering to traditional separate coding, it does not suffer from a cliff effect. Compressed sensing is applied to generate measurements of equal importance from a video such that a receiver with a better channel will naturally have more information at hands to reconstruct the content without penalizing others.

We experimentally compare different random matrices at the encoder side in terms of their performance for video transmission. We further investigate how properties of natural images can be exploited to improve the reconstruction performance by transmitting a small amount of side information. And we propose a way of exploiting inter-frame correlation by extending only the decoder. Finally we compare our results with a different scheme targeting the same problem with simulations and find competitive results for some channel configurations.

Keywords: Compressed Sensing, Video Multicast, Joint Source-Channel Coding.

1. INTRODUCTION

In a multicast scenario where a video is simultaneously transmitted to multiple receivers with different channels, it is difficult to allocate a fixed rate for the encoder to guarantee a low distortion over all receivers.

Following a traditional approach we would separate the source and channel coding. Shannons separation theorem¹⁴ largely simplifies the problem of optimal coding by splitting it into two different subproblems and still guarantees optimality to be achievable. Even though it holds in many cases, it does not apply to the case of multi-user channels as we target it. Because such schemes are designed for a given channel capacity, all receivers who can meet the requirements will be able to decode the content at the same suboptimal quality imposed by the encoding while some with less favorable channels will fail to decode it at all for most of the time. Because of the design of both variable length codes and video codecs such as MPEG, even a single error can propagate for a long time and lead to a big distortion. This behavior is commonly referred to as cliff effect and will appear in all separation based multicast schemes. Additionally, parameters of a wireless channel such as Signal to Noise Ratio (SNR), bandwidth or losses are prone to vary significantly between different receivers and over time, which further complicates the optimal resource allocation at the encoder.

This cliff effect has lead to the idea of Joint Source-Channel Coding (JSSC) where the source symbols are directly mapped to channel symbols by a single encoder. In this case a part of the distortion will be introduced by the channel rather than the encoding, leading to a graceful degradation with respect to the channel capacity. As videos are suitable for lossy compression it would be desirable to have a scheme which allows any receiver to decode the content with a quality corresponding to its channel and that at the same time remains efficient.

One way to practically implement a JSSC solution for video multicast was recently presented under the name *SoftCast*.⁸ Its main idea is inspired by Peterson¹² who showed that a group of coefficients \mathbf{s}_j with variance σ_j^2 should be scaled proportionally to $\sigma_j^{-1/2}$ (at the same time obeying some constraint on the total energy) before transmission through an Additive White Gaussian Noise (AWGN) channel in order to suffer the least distortion. *SoftCast* first decorrelates an image by applying a Discrete Cosine Transform (DCT) and then applies above scaling to groups of coefficients depending on their variances. They then compare their scheme with MPEG 4

E-mail: { markus.schenkel, pascal.frossard }@epfl.ch and { fengwu, chong.luo }@microsoft.com

and an MDC scheme at different, but fixed rates. They claim to have distortions comparable to the best out of those schemes at a given rate while having smooth degradation across rates. However, only intra-coding is used by the approach and for the comparisons.

The main contribution of this paper is a practical scheme for video multicast. It is based on the theory of compressed sensing to generate measurements of equal importance from the content. We first compare various combinations of measurement matrices in experiments for this application and evaluate the performance with respect to noise and loss levels in the channel. We further introduce two possible enhancements to the basic scheme. In the first case the encoder transmits additional side information about the distribution of the sparse coefficients which helps to improve the reconstruction quality. In the second the inter-frame correlation of videos is exploited at the decoder side without changing the encoder design. Finally we use simulations to compare the scheme with the above mentioned *SoftCast* as a main benchmark and find better results for some region of operating points.

The following section gives a brief introduction to Compressed Sensing; it is followed by a description of the proposed scheme (Sec. 3) and an analysis of the experimental results (Sec. 4).

2. COMPRESSED SENSING

2.1 Measurements

Compressed Sensing is a method to perfectly reconstruct a signal $\mathbf{x} \in \mathbb{R}^N$ from less than N non-adaptive, linear projections. It relies essentially on the possibility to represent \mathbf{x} sparsely in a known basis $\Psi \in \mathbb{R}^{N \times N}$ with only $K \ll N$ non-zero coefficients. We will denote this representation as

$$\mathbf{x} = \sum_{i=1}^N \psi_i s_i = \Psi \mathbf{s}$$

where $\mathbf{s} \in \mathbb{R}^N$ is the K -sparse coefficient vector. Instead of directly working with the signal \mathbf{x} , we project it onto a different basis $\Phi \in \mathbb{R}^{M \times N}$. In this context the components of the resulting vector

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} \in \mathbb{R}^M$$

are then called *measurements*. The compressed sensing theory now states that if Φ and Ψ are sufficiently incoherent (meaning that the columns $\{\psi_i\}$ of Ψ can not represent sparsely the rows of Φ) and if \mathbf{s} is sufficiently sparse, we can recover the signal \mathbf{x} from its measurements \mathbf{y} with high probability even though the system $\mathbf{y} = \Phi \mathbf{x}$ is highly underdetermined in terms of linear algebra.

Many pairs of bases are known to be incoherent, so for example the Fourier and Wavelet bases. In particular a basis built from i.i.d. random draws from a Gaussian or a Bernoulli distribution will be incoherent with any other fixed basis with a high probability in high dimensions. A sufficient condition linking the matrices Φ , Ψ and the sparsity K is the Restricted Isometry Property (RIP) introduced by Candès and Tao.³ Although it is a hard problem to exactly verify the RIP for a given matrix, it has been shown that certain random matrices satisfy it with high probability if

$$M \geq cK \ln \left(\frac{N}{K} \right)$$

is satisfied for a constant c .

2.2 Signal recovery

If the number of measurements M is chosen sufficiently big, then the minimization

$$(\mathbf{P}_0) : \quad \hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \hat{\mathbf{y}} = \Phi \Psi \mathbf{s} \quad (1)$$

is shown to find the unique solution $\hat{\mathbf{s}} = \mathbf{s}$. Once $\hat{\mathbf{s}}$ is found we can project it back to get $\hat{\mathbf{x}} = \Psi \hat{\mathbf{s}}$ in the original domain.

However, this minimization problem is of combinatorial nature and therefore computationally impossible for all but the smallest problem sizes. In particular, applications to images with their high number of dimensions would never be possible. But luckily above problem (\mathbf{P}_0) can be relaxed and reformulated as a convex optimization! Among all the ℓ_p norms, the one with the smallest value of $p \in \mathbb{R}$, which still satisfies the triangle inequality and hence is a norm and convex, is $\|\cdot\|_1$. Therefore it is a good candidate to replace the $\|\cdot\|_0$ pseudo-norm and to restate the problem as a constrained convex optimization problem.

It can be shown that if $\Phi\Psi$ satisfies the RIP of order $2K$ with radius $\delta_{2K} < \sqrt{(2)} - 1 \approx 0.414$, then the two minimizations are equivalent and will find the same unique $\hat{\mathbf{s}}$.

$$(\mathbf{P}_1): \quad \hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \hat{\mathbf{y}} = \Phi\Psi\mathbf{s} \quad (2)$$

Finally this problem (\mathbf{P}_1) can be solved by standard optimization approaches such as a linear program. In practice the minimal ratio M/K is sufficiently small (typically in the order of 2 to 4) for applications to be possible.

These findings do not only hold for exactly sparse signals, but can also be applied to *compressible signals* where the coefficient magnitudes decay sufficiently fast, but do not necessarily reach zero. If such a representation exists, the signal can be approximated by a best K -term approximation using a thresholded coefficient vector. In particular if the sorted magnitudes $|s|_{(i)}$ of the coefficients closely follow a power-law such that

$$|s|_{(i)} < C i^{-\alpha} \quad (3)$$

for some constants C and α we can relax above problem (\mathbf{P}_1) to solve

$$(\mathbf{P}_{QC}): \quad \hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\hat{\mathbf{y}} - \Phi\Psi\mathbf{s}\|_2 < \epsilon \quad (4)$$

instead. This formulation can also be used if the measurements are affected by noise \mathbf{e} such that $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{e}$. In which case (\mathbf{P}_{QC}) is called basis pursuit denoising (BPDN). The relaxation parameter ϵ needs to be chosen carefully. Let us assume that the noise is Gaussian with $e_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. and that we know its variance σ_n^2 . Then, even if we knew the perfect $\hat{\mathbf{x}} = \mathbf{x}$ by oracle we had an error of at least $\epsilon_0 = \|\Phi\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_2 = \|\mathbf{e}\|_2$ with an expected value of $E(\epsilon_0) = \sqrt{M}E(e_i) = \sqrt{M}\sigma$. Thus a practical ϵ needs to be chosen bigger and a value proposed by Candès² is

$$\epsilon = \sigma\sqrt{M}\sqrt{1 + 2\sqrt{2/M}} \geq \epsilon_0.$$

This is the reconstruction method we will use in the following.

Other promising applications of compressed sensing include among others Magnetic Resonance Imaging,¹⁰ imaging with single pixel cameras,⁵ spread spectrum receivers or data gathering in large wireless sensor networks.⁹

3. PROPOSED SCHEME

In our scheme (Illustrated by Fig. 1 on the following page) a single encoder transforms a video into a number of measurements through multiplication with the random measurement matrix. These measurements are then directly transmitted towards the receivers over possibly very different channels distorting the signal with losses and noise. Finally each decoder will use the measurements he receives to reconstruct the sparsest signal in some basis such that the decoded signal is compatible with the measurements. Most natural images are compressible in some bases, given by the DCT, wavelet transform or possibly others. This implies that a sparse approximation exists and makes our approach possible.

3.1 Encoding

In our application a grayscale image is split into a subblocks of size $n \times n$. A block is represented as a column vector of length $N = n^2$ by concatenating its columns. We will use the terms *image* and *signal* synonym in the following. The encoder works on each block \mathbf{x} of each frame of a video. It first subtracts the mean value from a frame and then applies the full random matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ to \mathbf{x} . Those measurements are then interleaved across

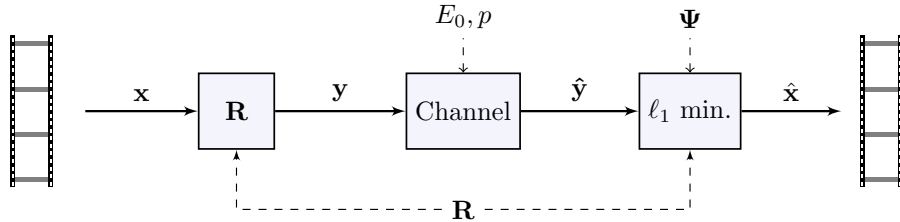


Figure 1. Flow graph of the proposed scheme (only one channel and receiver shown).

blocks and assembled into packages. This ensures that lost packages will not erase a single block completely but rather affect all of them evenly.

We have evaluated different random matrices \mathbf{R} and sparse bases Ψ ; in particular matrices drawn uniformly from a Gaussian distribution (called \mathbf{G} in the following) with $G_{ij} \sim \mathcal{N}(0, 1)/\sqrt{N}$, \mathbf{G}^\perp derived from \mathbf{G} by normalizing each column, \mathbf{G}^\perp an orthogonalized \mathbf{G} , \mathbf{B} drawn from a Bernoulli distribution with $B_{ij} \in \{-1, 1\}/\sqrt{N}$ and a Hadamard transform \mathbf{H} . In terms of sparse bases we compared the discrete cosine transform basis \mathbf{D} and a Haar Wavelet basis \mathbf{W} ; both of them can be implemented with an efficient algorithm.

For the analysis of our approach we assume two possible channel models. In both cases we use direct analog transmission of the measurement values, i.e. no quantization or other coding is applied. First, we depart from a channel with Additive White Gaussian Noise (AWGN) in which packages are erased at random with a loss rate p . After a full set of N measurements is created, these two distortions are supposed to interfere as follows: First noise is added to the full measurements such that the Channel Signal to Noise Ratio (CSNR) satisfies

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

exactly in order to make results easier to compare. Finally a fraction $M = \lfloor (1 - p)N \rfloor$ of the measurements are kept and handed over to the decoder.

Secondly we also apply a noiseless, block erasure channel, as it could arise from a best-effort network. For the noisy case described above this corresponds to the limit of an infinite CSNR.

Both channel models act directly on the baseband, no assumptions about modulation are made. It allows us to compare the trade-off between noisy and lossy channels as will be discussed in Sec. 4. Possible reasons causing package loss include collisions occurring for some receivers or congestion at the transmitter side.

3.2 Decoding

Before the measurements are decoded we need to compose the random matrix for each block according to the package loss pattern and then estimate the noise level.

If the encoder used the random matrix $\mathbf{R} \in \mathbb{R}^N$ to generate the measurements and among the N measurements of a given block the ones indexed by the set $S \subseteq \{1, \dots, N\}$ with cardinality $|S| = M$ were received, we construct the measurement matrix

$$\Phi^T = \{(\mathbf{R}^T)_i \mid i \in S\}$$

from the rows of \mathbf{R} indexed by S . The received measurements are deinterleaved and arranged accordingly into $\hat{\mathbf{y}}$. We assume that \mathbf{R} is either known to all parties of the system or communicated by the transmitter by means of a simple random seed.

A noiseless channel would lead to a zero error vector, but this will not be the case for a practical channel. Hence we will use the $(\mathbf{P}_{\mathbf{QC}})$ decoding algorithm together with the estimate for ϵ from Sec. 2.2. Then $\hat{\mathbf{s}}$ is calculated by solving $(\mathbf{P}_{\mathbf{QC}})$ using the ℓ_1 -magic² implementation for Matlab. Finally $\hat{\mathbf{s}}$ is projected back into the pixel domain and $\hat{\mathbf{x}}$ is displayed.

So far we have only assumed that an image can be sparsely approximated in some fixed bases. But we can also use previous knowledge about the coefficient distribution as well as the inter-frame correlation of videos to improve the performance of this basic scheme. These two modifications to the decoder will be presented in the following.

Reweighted decoding At first we show how previous knowledge about the distribution of the image coefficients in the transform domain can improve the performance. The intuition behind this approach is, that if we knew the perfect solution \mathbf{x} by oracle, we could define the diagonal weighting matrix $\mathbf{\Lambda}$ as

$$\Lambda_{ii} = \frac{1}{|s_i| + \gamma}$$

such that the reweighted ℓ_1 norm

$$\|\mathbf{\Lambda}\mathbf{s}\|_1 = \sum_i |\Lambda_{ii}s_i| = \sum_i \left| \frac{s_i}{|s_i| + \gamma} \right| \approx |\{s_i | s_i \neq 0\}| = \|\mathbf{s}\|_0.$$

would actually approximate the ℓ_0 norm. Above, the parameter $0 < \gamma \ll 1$ serves for regularization of s_i close or equal to zero. The reweighted ℓ_1 minimization then becomes

$$(\mathbf{P}_{\mathbf{RW}}) : \hat{\mathbf{s}} = \arg \min_{\mathbf{s}} \|\mathbf{\Lambda}\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\hat{\mathbf{y}} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}\|_2 < \epsilon \quad (5)$$

By introducing $\mathbf{q} = \mathbf{\Lambda}\mathbf{s}$ we can equivalently write

$$\hat{\mathbf{q}} = \arg \min_{\mathbf{q}} \|\mathbf{q}\|_1 \quad \text{s.t.} \quad \|\hat{\mathbf{y}} - \mathbf{\Phi}\mathbf{\Psi}\mathbf{\Lambda}^{-1}\mathbf{q}\|_2 < \epsilon$$

$$\hat{\mathbf{s}} = \mathbf{\Lambda}^{-1}\hat{\mathbf{q}}$$

and reuse existing optimization algorithms without change. Candès et al.⁴ suggest using this approach iteratively to outperform unweighted ℓ_1 minimization even in the case of not exactly sparse signals as long as the exponent α in (3) remains below 1. Unfortunately this is empirically not the case for natural images in a DCT (or any other) basis. But their low frequency components are usually still dominant over the high frequency components. The estimated coefficient magnitudes are unevenly distributed and remain consistent within a given frame. This allows us to use the variances of each coefficient in a frame as an estimate for its magnitude and plug these values into the reweighted minimization as follows:

$$W_{ii} \propto \frac{1}{\sqrt{E(s_i^2)}}.$$

This choice actually leads us to a norm similar to the Mahalanobis distance but for the ℓ_1 case. In contrast to the basic scheme, the encoder now also needs to apply the transform $\mathbf{\Psi}$ to all blocks of the image which adds to its complexity and will also fix the choice of $\mathbf{\Psi}$ for all decoders that use this side-information. We notice that the weights need to be transmitted from the encoder to all the receivers and hence we need a lossless side-channel for this scheme to work. But the rate required for this is small: If a frame is divided into B blocks then a fraction of $1/B$ of the total data will be required for this side information, remaining below 1% for practical block-sizes. The fact that these variances vary only slowly from one frame to the next can further reduce the rate. Furthermore the *SoftCast* scheme we compare to assumes the same information to be available at the decoder.

Inter-frame decoding Without increasing the encoder complexity – or even changing it at all – we can use the correlation among frames to decode two (or more) frames together. This follows the paradigm of distributed video coding (DVC) and similar ideas have been studied by Prades-Nebot,¹³ Marcia¹¹ and Chen.⁶

Although subsequent frames of a video are usually highly correlated, this correlation is subject to motion. Hence we can expect a joint sparsity model to be successful only if this motion is reflected by some sort of geometric transform between the bases of each frame. The fact that the compressed sensing model allows a decoder to independently choose any matrix $\mathbf{\Psi}$ as long as \mathbf{s} remains sparse makes this approach possible.

We will use Joint Sparsity Models (JSM) to exploit the dependency between successive frames. JSM are based on the fact that both the location and the values of the non-zero values in a sparse vector are unknowns.

Hence if we already knew the support of a sparse signal we could drastically reduce the number of measurements required to decode it because only the non-zero values would remain as unknowns. One way of reducing this number of unknowns is by assuming that two or more subsequent signals will have the same or a very similar support. We will treat two frames \mathbf{x}_1 and \mathbf{x}_2 together, but this easily extends to three or more as well. We assume a joint sparsity model of a common sparse component plus innovations, commonly referred to as *JSM-1* following the framework established by Baron et al.¹

The main idea is to first estimate the local motion between two frames using an intra-decoded frame 1 as a reference for the next frame 2. For each block j of frame 2 we are about to reconstruct, we take all possible patches in frame 1 that are supported around block j as spanning vectors for a sparse representation. We call the matrix constructed from these patches $\Psi^{(j)}$. It will not form a complete basis, but if the two frames are highly correlated by motion one of its elements will be much more consistent with the measurements of that block. The index of this best atom will give us an estimate for the motion vector \mathbf{v} between the two frames around block j .

After the motion is estimated based on two independent reconstructions a joint basis can represent two concatenated blocks (one from each frame). This joint basis consists of a common part which supports both frames using a given basis for the first and a shifted version of it for the second frame. It is augmented with two separate bases for each frame. This makes the joint basis over-complete but possibly leads to a higher sparsity of \mathbf{s} . Finally we get

$$\Psi^{(j)} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Psi & \sqrt{2}\Psi & \mathbf{0} \\ T_{\mathbf{v}^{(j)}}(\Psi) & \mathbf{0} & \sqrt{2}\Psi \end{pmatrix} \quad (6)$$

where the transform $T_{\mathbf{v}^{(j)}}(\cdot)$ denotes the shift by the motion vector $\mathbf{v}^{(j)}$. The two concatenated blocks can then be represented by a set of joint ($\mathbf{s}_{\text{joint}}$) and independent ($\mathbf{s}_1, \mathbf{s}_2$) coefficients as follows:

$$\hat{\mathbf{x}}_{\text{joint}} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \Psi \mathbf{s} = \Psi \begin{pmatrix} \mathbf{s}_{\text{joint}} \\ \mathbf{s}_1 \\ \mathbf{s}_2 \end{pmatrix}.$$

Each of the steps is performed locally for each block and the block size will also determine the search space of the motion estimation. We use a circular shift for the transform $T(\cdot)$ in the joint optimization. Whereas performance is not affected too much, this provides the two advantages that we can guarantee a high level of incoherence without the need to remove duplicate atoms and that it can be implemented using a fast transform.

4. EXPERIMENTAL RESULTS

We have implemented our scheme in Matlab and evaluated it at various operating points defined by the channel loss rate p and the CSNR. This allows us to better understand how the trade-off between noisy and lossy channels acts on CS decoding. For comparison the Peak Signal to Noise Ratio (PSNR) was used, a standard measure for image applications. It is defined as

$$\text{PSNR} = 10 \log_{10} \frac{\max_i \{x_i\}}{\frac{1}{N} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}.$$

All graphs in the following figures compare the image PSNR for different operating points in this parameter space.

1. Basic scheme For all the investigated combinations of \mathbf{R} and Ψ , experiments were run for a wide range of these two parameters in order to better understand the trade-off. They are visualized in Fig. 2.

The best performance is achieved when an orthogonal random matrix (\mathbf{G}^\perp) is used together with decoding into the DCT basis \mathbf{D} . It is slightly worse for just a random Gaussian matrix (\mathbf{G}) and for one with orthonormal columns (\mathbf{G}^1); however, the differences are quite small and graphs are omitted. The same holds if the encoder additionally performs a decorrelating transform (i.e. $\mathbf{R} = \mathbf{G}\mathbf{D}$); this is to be expected from the CS theory because the overall matrix $\Phi\Psi$ stays the same. On the other hand when we use the deterministic measurement matrix \mathbf{H} , results drop by around 1 dB for almost all operating points.

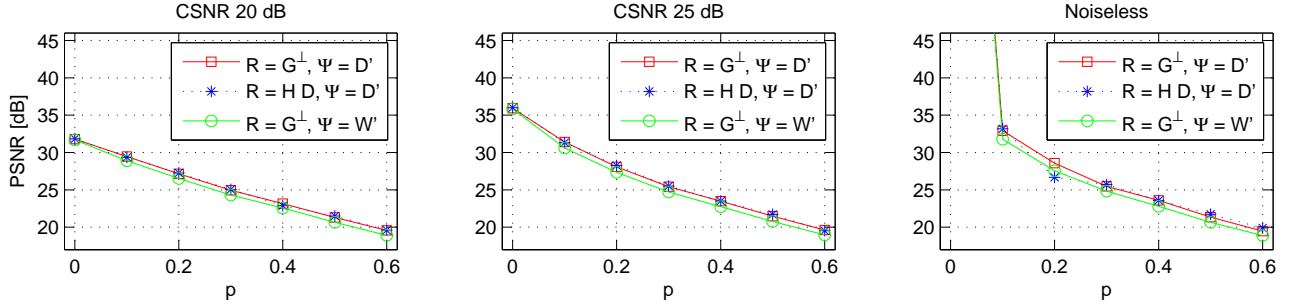


Figure 2. Comparison of different matrices Φ and Ψ as described above: (red) $\mathbf{R} = \mathbf{G}^\perp, \Psi = \mathbf{D}^T$, (blue) $\mathbf{R} = \mathbf{H}, \Psi = \mathbf{D}^T$ and (green) $\mathbf{R} = \mathbf{G}^\perp \mathbf{D}, \Psi = \mathbf{W}^T$. The graphs show results at different noise levels where each compares the image PSNR vs. the loss ratio p for the sequence *football*.

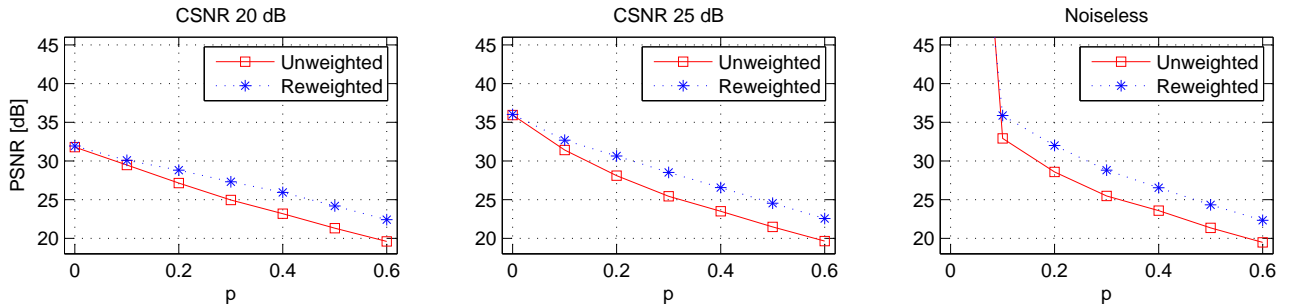


Figure 3. Comparison of (blue) default ℓ_1 minimization with (red) reweighting. The graphs show results at different noise levels and each compares the image PSNR vs. the loss ratio p for the sequence *football*.

Concerning the choice of the sparse representation we observe that a wavelet basis \mathbf{W} performs worse than the DCT. We can also see that there is no visible cliff effect and we achieve a graceful degradation with respect to both dimensions of distortion. Hence one of the principal design goals is met. The decay with respect to the loss rate p is smooth, but faster than we anticipated. In most cases we could further improve the performance slightly by operating on bigger blocks, but considering the higher complexity this is a less attractive option.

2. Reweighting Figure 3 shows the achieved improvements for various operating points. We see that at higher loss rates p , reweighting improves results by around 1–2 dB in PSNR while penalizing the results at full measurement rate slightly. But the choice of the decoding method is left to the decoder which could easily switch back to unweighted decoding at those operating points. The same holds in the rare cases where the reweighted optimization does not converge.

3. Inter-frame decoding Our implementation uses blocks of size 8×8 for both motion estimation and reconstruction. Figure 6 illustrates the reconstruction of one block in the joint reconstruction step. There we see the distribution of the coefficients among the joint and the two distinct bases. If we consider that videos have usually a high correlation between frames, then in the spirit of the Slepian-Wolf theorem it should be possible to get a considerable performance gain from inter-frame decoding alone. Nevertheless our results (shown in Fig. 4) are less favorable. Overall we achieve only little gain for low CSNR and improve by up to 1 dB for better channels even in the presence of losses. Although experiments show that an over-complete basis as given by (6) leads to slightly better results than a complete one, it also increases the complexity. We conclude that a more sophisticated joint sparsity model or a different approach should be sought after in order to achieve a more significant performance gain.

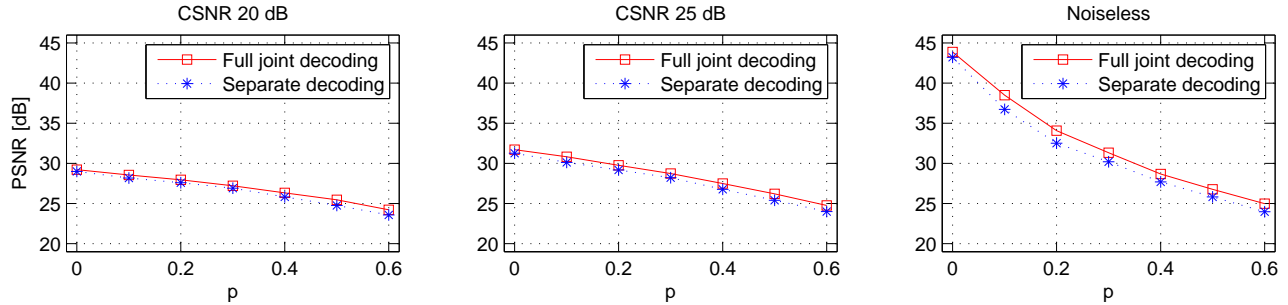


Figure 4. Comparison of (blue) default ℓ_1 minimization with (red) inter-frame decoding for two consecutive frames of the sequence *football*. The graphs show results at different noise levels and each compares the image PSNR vs. the loss ratio p .

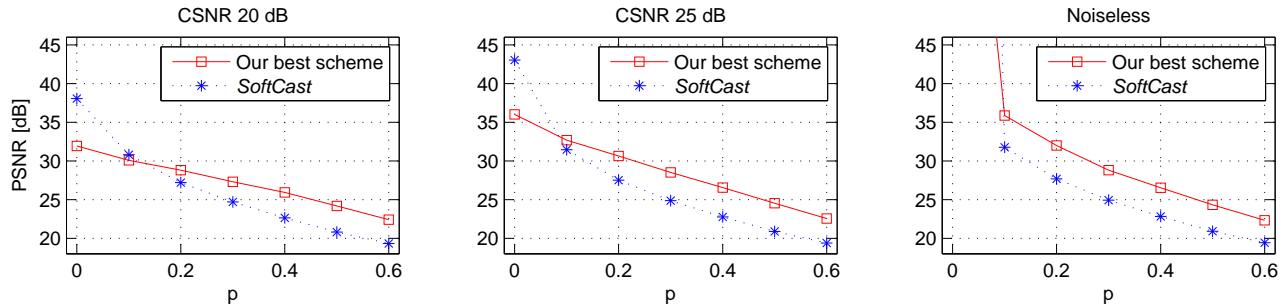


Figure 5. Comparison of (blue) our scheme with (red) *SoftCast*. The four graphs show results at different noise levels and each compares the image PSNR vs. the loss ratio p for the sequence *football*.

4. Comparison with *SoftCast* Finally we compare our best configuration (Reweighted decoding applied to $\mathbf{R} = \mathbf{G}^\perp, \Psi = \mathbf{D}$) with *SoftCast*; the results are shown in Fig. 5. Our scheme is competitive at high loss rates p and low channel noise. However, the *SoftCast* scheme is clearly superior in the contrary cases of low losses and for high noise. The fact that our scheme lies behind *SoftCast* at low losses is less surprising considering that we perform only direct random projections at the encoder where *SoftCast* is based on optimal scaling as the essential encoding step. On the other hand we achieve a performance gain in the order of 2 dB over their scheme starting from only 10–20% loss.

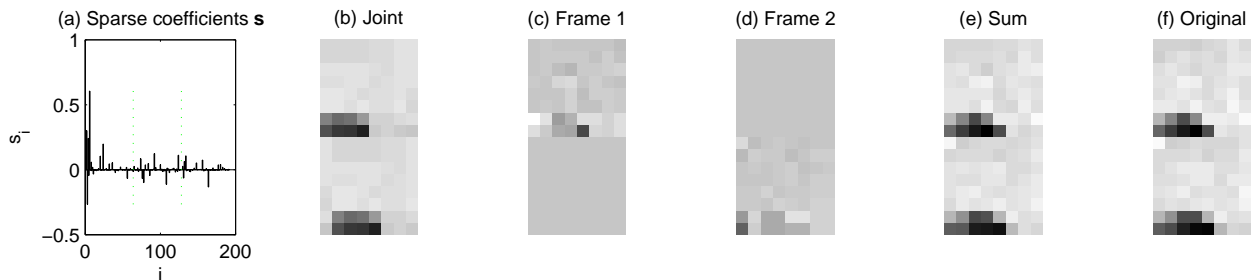


Figure 6. An arbitrary block being reconstructed into the over-complete dictionary $\Psi^{(j)}$. (a) shows the sparse coefficient vector split into the three parts of the joint basis and the two separate bases. The other subfigures show the same block for both frames on top of each other as follows: (b) – (d) the joint and distinct contributions respectively, (e) the sum of them and (f) the original.

5. CONCLUSIONS

We have proposed a scheme for wireless video multicast based on compressed sensing which does not suffer from a cliff-effect. We compare our results with a recent scheme designed for the same purpose and find competitive results in the case of high losses and low noise. Furthermore extensive experiments were deployed to compare various measurement matrices.

The advantages of our scheme are a non adaptive, simple encoder and transmission. It automatically scales with the number of receivers and does not require solving any resource allocation problem at the transmitter. Additionally this introduces the interesting property that opposed to the traditional approach where the decoder is fully specified and encoding is left for various implementations, here a decoder has the free choice of its method, in practice the choice of the sparse basis and a correlation model. This adds a certain universality and makes the setup future proof. The main drawback is obviously a high complexity at the receiver side, that is required to solve the optimization problem. This makes the implementation of a real-time application for higher resolutions challenging.

As previously investigated by Goyal⁷ and others, CS can not be expected to be optimal for compression in the information theoretic sense when applied to data that is already fully sampled. Nevertheless we conjectured a higher overall gain to be achievable in such a JSCC scenario than what we found through experiments. We will focus our future work on better exploiting the inherent correlation of videos to achieve a higher reconstruction performance.

REFERENCES

- [1] D. Baron, R.G. Baraniuk, M.B. Wakin, M.F. Duarte, and S. Sarvotham. Distributed compressed sensing. *preprint*, pages 1–50, 2005.
- [2] E.J. Candes and J. Romberg. l1-magic : Recovery of Sparse Signals via Convex Programming. *URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf*, 2005.
- [3] E.J. Candes, J. Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52:489–509, 2006.
- [4] E.J. Candes, M.B. Wakin, and Stephen P. Boyd. Enhancing Sparsity by Reweighted l1 Minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, October 2008.
- [5] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, K.F. Kelly, and R.G. Baraniuk. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Processing Magazine*, 25:83–91, March 2008.
- [6] P. Frossard and X. Chen. Joint reconstruction of compressed multi-view images. *Proc. ICASSP*, pages 1005–1008, April 2009.
- [7] V.K. Goyal, A.K. Fletcher, and S. Rangan. Compressive Sampling and Lossy Compression. *IEEE Signal Processing Magazine*, 25:48–56, March 2008.
- [8] S. Jakubczak, H. Rahul, and D. Katabi. One-Size-Fits-All Wireless Video. *HotNets*, 2009.
- [9] C. Luo, F. Wu, J. Sun, and C.W. Chen. Compressive Data Gathering for Large-Scale Wireless Sensor Networks. In *MobiCom*, pages 145–156. ACM, 2009.
- [10] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Process. Mag.*, 25:72–82, 2007.
- [11] R. Marcia and R.M. Willett. Compressive coded aperture video reconstruction. *Proc. European Signal Processing Conf.(EUSIPCO)*, 2008.
- [12] D.P. Petersen and K-H. Lee. Optimal Linear Coding for Vector Channels. *IEEE Transactions on Communications*, 24:1283–1290, December 1976.
- [13] J. Prades-Nebot, Y. Ma, and T. Huang. Distributed video coding using compressive sampling. *Proceedings of the Picture Coding Symposium*, 2009.
- [14] C.E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.