COMPRESSIVE IMAGE SAMPLING WITH SIDE INFORMATION

Vladimir Stanković, Lina Stanković

Dept Electronic and Electrical Engineering
University of Strathclyde, Glasgow, UK
{vladimir, lina}.stankovic@eee.strath.ac.uk

Samuel Cheng

Dept Electrical and Computer Engineering
University of Oklahoma, Tulsa, OK
samuel.cheng@ou.edu

ABSTRACT

Compressive sampling is a novel framework that exploits sparsity of a signal in a transform domain to perform sampling below the Nyquist rate. In this paper, we apply compressive sampling to reduce the sampling rate of images/video. The key idea is to exploit the intra- and inter-frame correlation to improve signal recovery algorithms. The image is split into non-overlapping blocks of fixed size, which are independently compressively sampled exploiting sparsity of natural scenes in the Discrete Cosine Transform (DCT) domain. At the decoder, each block is recovered using useful information extracted from the recovery of a neighboring block. In the case of video, a previous frame is used to help recovery of consecutive frames. The iterative algorithm for signal recovery with side information that extends the standard orthogonal matching pursuit (OMP) algorithm is employed. Simulation results are given for Magnetic Resonance Imaging (MRI) and video sequences to illustrate advantages of the proposed solution compared to the case when side information is not used.

Index Terms— Image processing, Image reconstruction

1. INTRODUCTION

Compressive sampling (CS) replaces conventional sampling and reconstruction with a more general linear measurement scheme and an optimization procedure to acquire a subset of signals within a source at a rate that is below Nyquist. A number of theoretical contributions have appeared on CS (see [1]) over the past few years.

In this paper we investigate the use of CS for acquisition of image/video. Indeed, using the fact that natural images are sparse in a transform domain, one can reduce the required sampling rate without sacrificing perceptual quality. Conventionally, after acquisition of a gray-scale image, Discrete Cosine Transform (DCT) is performed on the image using values assigned to each pixel. After DCT, many coefficients will be zero or will carry negligible energy; these redundant coefficients are discarded before quantization or/and entropy coding. Hence, though the image is acquired fully, much of the acquired information is discarded after DCT.

Thus CS seems a natural tool to reduce the number of acquired samples - this is envisioned to dramatically reduce imaging cost in the spectrum where silicon is blind [2]. Silicon is improving in capacity for visible light image capture so this technique will not benefit the commonplace visible-light digital camera. However, for cameras capturing scenes at infra-red or other parts of the spectrum, like terahertz imaging, CS has huge potential because silicon-based sensors are not available and thus the cost of sensing is expensive. Although the key concepts have already been laid out, many issues still remain to be resolved before practical *compressive imaging* is deployed widely. One of the main challenge is designing efficient and fast signal recovery algorithms, that are able to reconstruct an N-dimensional signal using $M < N$ measurements by exploiting sparsity of the signal in a domain of choice. Since optimal recovery is an NP-hard problem, several sub-optimal solutions have been reported (see [1, 3]). One of them is the orthogonal matching pursuit (OMP) algorithm [4], which is very popular due to its relatively lower complexity compared to other proposed reconstruction methods. However, the execution time of OMP is still high for many practical applications where sparsity of the signal and its dimension are high. One such example is image/video acquisition [2], where frame sizes are usually too large for OMP recovery.

Benefits of compressive imaging have been shown by several research groups (see [1] and references therein). For example, [5] investigates the possibility of applying compressive sampling to Magnetic Resonance Imaging (MRI) acquisition. Several optical architectures are compared in [6] for compressive imaging applications. In [7], we proposed a scheme for compressive video sampling that exploits local sparsity within a frame, followed by OMP recovery. In [2], two methods are proposed to build a single-pixel video camera. The first method is frame-by-frame reconstruction with CS in a 2D wavelet domain, and the second one exploits a 3D measurement matrix and a 3D wavelet transform to achieve joint reconstruction of all the frames. Whereas it is expected that the latter approach will show better performance, its complexity is higher. The first, lower complexity approach, reconstructs each frame without using useful information about previously reconstructed frames.

In this paper we propose a scheme for compressive imaging, which follows the first approach of [2]. The improved performance at reduced complexity is achieved by exploiting intra- and inter-frame correlation via a signal recovery algorithm with side information. The algorithm extends OMP to the case when *a priori* information about the source is present at the decoder in the form of estimated positions of significant signal's elements. If the side information is correct, the algorithm finds the solution with fewer number of iterations than OMP. In addition, the quality of the solution is generally better, which indicates that fewer measurements are needed. If the side information is noisy, i.e., some of the assumed positions at the decoder are not correct, the algorithm has a mechanism to correct them and converge to the correct solution.

The algorithm is suitable for scenarios where the decoder has access to a correlated source, like in image/video acquisition where a previously reconstructed block can play the role of side information. We suppose that the first frame is available at the decoder, and each subsequent frame is compressively sampled and reconstructed at the decoder using a previous frame. For single-image sampling, the image is split into non-overlapping blocks, and at the decoder, each block is recovered using useful information extracted from the recovery of a neighboring block. Our scheme can be used with the single-pixel camera architecture [2] or MRI [5]. In our experiments, we use high resolution images and simulate how well we could have recovered the image/video after CS. Our simulation results show that it is possible to reduce the number of acquired samples using CS without sacrificing the reconstruction performance.

2. COMPRESSIVE SAMPLING

Compressive sampling (CS) is a novel framework that enables sampling below the Nyquist rate, without (or with a small) sacrifice in reconstruction quality. It is based on exploiting sparsity of the signal in some domain. In this section we briefly review CS.

Let $\mathbf{x}$ be a vector of N samples of a real-valued, discrete-time random process. Let

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

where $\mathbf{s} = [s_1, \ldots, s_N]$ is an N-vector of weighted coefficients $s_i = \langle \mathbf{x}, \psi_i \rangle$, and $\mathbf{\Psi} = [\psi_1|\psi_2|\cdots|\psi_N]$ is an $N \times N$ orthonormal basis matrix with ψ_i being the i-th basis column vector. That is, $\mathbf{s}$ is the representation of $\mathbf{x}$ in domain $\mathbf{\Psi}$. For example, $\mathbf{s}_i$'s can be discrete cosine or Fourier coefficients.

Vector $\mathbf{x}$ is considered K-sparse in the domain $\mathbf{\Psi}$, for $K \ll N$, if only K out of N elements of $\mathbf{s}$ are non-zero. Many natural signals can be approximated as sparse since they have many non-significant (close to zero) coefficients after transform. Sparsity of a signal is used for compression with conventional transform coding, where the whole signal is first acquired (all N samples), then the N transform coefficients $\mathbf{s}$ are obtained via $\mathbf{s} = \mathbf{\Psi}^{-1}\mathbf{x}$, and finally $N-K$ non-significant coefficients of $\mathbf{s}$ are discarded and the remaining are encoded. The resulting acquisition redundancy is due to large amounts of data being discarded because they carry negligible or no energy.

The main idea of CS is to remove this "sampling redundancy" by requiring only M samples of the signal, where $K < M \ll N$. Let $\mathbf{y}$ be an M-length measurement vector given by:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x},$$

where $\mathbf{\Phi}$ is an $M \times N$ measurement matrix. Then, the above expression can be written in terms of $\mathbf{s}$ as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s} = \mathbf{\Phi}'\mathbf{s}. \quad (2)$$

Sampling is performed in $\mathbf{\Psi}$ domain by collecting coefficients $\mathbf{s}$, which will work only if signal $\mathbf{x}$ is sparse in domain $\mathbf{\Psi}$, i.e., $N-K$ elements of $\mathbf{s}$ are negligible.

Note that (2) is a dimensionality reduction thus leading to a loss in information in general. That is, there are infinitely many $\mathbf{x}'$ that when multiplied by $\mathbf{\Phi}$ give $\mathbf{y}$. However, signal $\mathbf{x}$ can be recovered losslessly from $M \approx K$ or slightly more measurements if the measurement matrix $\mathbf{\Phi}$ is properly designed, so that $\mathbf{\Phi}\mathbf{\Psi}$ satisfies the so-called restricted isometry property (RIP). This will always be true if $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are incoherent, that is, the vectors of $\mathbf{\Phi}$ cannot sparsely represent basis vectors of $\mathbf{\Psi}$ and vice versa.

It was further shown that an independent identically distributed (i.i.d.) zero-mean Gaussian matrix satisfies the above property for any orthonormal $\mathbf{\Psi}$ with high probability. Some other choices of $\mathbf{\Phi}$ that satisfy RIP are random matrix with +1/-1 entries drawn from uniform Bernoulli distribution, and randomly permuted vectors from standard orthonormal bases, such as Fourier and Walsh-Hadamard. Also, it has been shown that it is enough for a signal $\mathbf{x}$ to be r-compressible (the sorted coefficients decay under a power law with scaling exponent $-r$), instead of K-sparse (see [1] and references therein).

Unfortunately, reconstruction of $\mathbf{x}$ (or equivalently, $\mathbf{s}$) from vector $\mathbf{y}$ of M samples is not trivial. The exact solution is NP hard and consists of finding the minimum l_0 norm (the number of non-zero elements). However, an excellent approximation can be obtained via the l_1 norm minimization given by:

$$\hat{\mathbf{s}} = \arg\min ||s'||_1, \quad \text{such that} \quad \mathbf{\Phi}\mathbf{\Psi}\mathbf{s}' = \mathbf{y}. \quad (3)$$

A K-sparse signal can be recovered with high probability using (3) if $M \geq cK\log(N/K)$ for some small constant c. Thus, one can recover N measurements of $\mathbf{x}$ with high probability from only $M \approx cK\log(N/K) < N$ random measurements $\mathbf{y}$ under the assumption that $\mathbf{x}$ is K-sparse in domain $\mathbf{\Psi}$.

This convex optimization problem, namely, basis pursuit, can be solved using a linear program algorithm of $O(N^3)$ complexity. Due to complexity and low speed of linear programming algorithms, faster solutions were proposed at the expense of slightly more measurements, such as matching pursuit [3], tree matching pursuit [3], OMP [4], and group testing.

3. PROPOSED SYSTEM

In this section, we describe our scheme for CS of image/video. We start by describing a signal recovery algorithm building on OMP but with side information, and then continue with the acquisition scheme description.

3.1. Signal Recovery with Side Information

Suppose that $\mathbf{\Phi}$ is the $M \times N$ random measurement matrix, and $\mathbf{x}$ is the N-length K-sparse signal. The decoder has access to the M-length measurement vector $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$. In addition, the decoder has *a priori* knowledge about the signal, side information, in the form of estimated positions of some significant elements in $\mathbf{x}$ (which might not be correct). The problem is to reconstruct signal, $\hat{\mathbf{x}}$, based on $\mathbf{y}$, $\mathbf{\Phi}$, and side information. For a set Θ, $|\Theta|$ is its cardinality, and $\{\}$ denotes an empty set. For a matrix $\mathbf{\Omega}$, ω_j denotes its j-th column; furthermore, $\mathbf{\Omega}_\Theta$ is a matrix of $|\Theta|$ columns of $\mathbf{\Omega}$ with indices from set Θ.

The main idea of the algorithm is to start with the estimated positions of significant elements of $\mathbf{x}$, and then in each iteration find the most strongly correlated column among remaining ones. This column will either be included as an additional column or replace the column in the set of estimated positions that is least correlated.

Algorithm 1 *Signal recovery with side information*

`INPUT:`

- *An $M \times N$ measurement matrix $\mathbf{\Phi}$*
- *An M-dimensional measurement vector $\mathbf{y}$*
- *The sparsity level of the signal K*
- *Maximum number of iterations $T \geq K$*
- *The side information set Λ_1 with at the most K elements*
- *Constants $\kappa_1, \kappa_2 \leq 1$.*

`OUTPUT:`

- *An N-dimensional estimate $\hat{\mathbf{x}}$ of the signal $\mathbf{x}$*
- *A set Λ_t, $t > 1$, containing K elements from $\{1, \ldots, N\}$*

PROCEDURE:

1. Set $t = 1$*. If* $\Lambda_t = \{\}$ *then* $\Lambda_t = \{\arg\max_{j=1..N} |\langle \mathbf{y}, \phi_j \rangle|\}$.

2. $\mathbf{\Omega}_t = \mathbf{\Phi}_{\Lambda_t}$, $\mathbf{x}_t = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{\Omega}_t \mathbf{x}||_2$*. Let* p *be the projected value, and* $\tilde{m}$ *the least correlated column in* $\mathbf{\Omega}_t$ *to* $\mathbf{y}$.

3. $\mathbf{r} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}_t$

4. $q = \max_{j=1..N} |\langle \mathbf{r}, \phi_j \rangle|$
$\tilde{l} = \arg\max_{j=1..N} |\langle \mathbf{r}, \phi_j \rangle|$

5. If $(p < q\kappa_1)$ *then* $\Lambda_t = \Lambda_t / \{\tilde{m}\}$
else if $(p\kappa_2 < q)$ $\Lambda_{t+1} = \Lambda_t$ *and goto 8*

6. $\Lambda_{t+1} = \Lambda_t \bigcup \{\tilde{l}\}$*. If* $|\Lambda_{t+1}| \geq K$ *then*
$\{\mathbf{\Omega}_t = \mathbf{\Phi}_{\Lambda_{t+1}}, \mathbf{x}_t = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{\Omega}_t \mathbf{x}||_2$ *and goto 8* $\}$

7. If $t < T$ *increment* t *and goto 2.*

8. Stop: The estimate $\hat{\mathbf{x}}$ *has nonzero indices at the components listed in* Λ_{t+1}*. The value of* $\hat{\mathbf{x}}$ *in component* λ_j *equals the* j*-th component of* $\mathbf{x}_t$.

The algorithm allows possibility for a wrong guess, that is, the estimated position(s) of the significant elements in $\mathbf{x}$ are wrong. This is why the number of iterations T is allowed to be higher than K. In the case when probability of a wrong guess is zero, T is equal K.

Λ_1 is the set of the known/estimated positions of the significant elements of $\mathbf{x}$ at the decoder. If $\Lambda_1 = \{\}$, the algorithm boils down to the OMP algorithm, thus in Step 1, as in OMP, Λ_1 is set to the most strongly correlated column in $\mathbf{\Phi}$ to $\mathbf{y}$.

In Step 2, $\mathbf{\Omega}_t$ is a matrix of columns of $\mathbf{\Phi}$ that correspond to significant elements of $\mathbf{x}$ based on side information. The decoder forms a projection, $\mathbf{x}_t$, of $\mathbf{y}$ onto $\mathbf{\Omega}_t$. Then, the decoder calculates the column in $\mathbf{\Omega}_t$ that is least correlated to $\mathbf{y}$. This will be a candidate for removal since it might be a wrong guess.

In Steps 3 and 4, a residual is computed based on the current estimate $\mathbf{x}_t$, and as in OMP, the most strongly correlated column in $\mathbf{\Phi}$ from the remaining columns is computed and its index is set to $\tilde{l}$. If column $\tilde{l}$ is more correlated than column $\tilde{m}$, $\tilde{m}$ is removed. Otherwise, if $\tilde{m}$ is much more correlated (regulated by constant κ_2), the decoder concludes that there cannot be further improvement of the estimate, and exits. If this is not the case, the decoder adds column $\tilde{l}$ to $\mathbf{\Omega}_t$, increments t and goes to the next iteration.

If all initial guesses are correct, the algorithm boils down to OMP with $K - |\Lambda_1|$ iterations, resulting in reduced complexity and execution time. In more realistic situations when side information is not perfect, e.g., reconstruction of correlated sources, the algorithm effectively corrects wrong guesses after more than $K - |\Lambda_1|$ iterations.

3.2. Compressive Image Acquisition in the DCT Domain

In [2] a single-pixel camera architecture was developed based on the programmable digital micro-mirror device (DMD), which consists of a 2D array of micro-mirrors, where each micro-mirror corresponds to one pixel on the image. The micro-mirrors's orientation (+12/-12 degrees) is controlled by a pseudo-random measurement matrix $\mathbf{\Phi}$ with +1/-1 entries drawn from uniform Bernoulli distribution. For each of M different DMD configurations (selected by $\mathbf{\Phi}$), one measurement is obtained. A single photo-detector is used to generate M voltage readings corresponding to (2). Matching pursuit [3] is then employed for reconstruction.

In the case of video acquisition, two methods are proposed in [2]. In the first method, each frame is sampled in 2D domain and recovered independently from the previous frames. In the second method, a 3D measurement matrix and a 3D wavelet transform are exploited to achieve joint reconstruction of group of frames. Whereas it is expected that the latter approach will show better performance, its complexity is higher. The frame-by-frame sampling is also more flexible in terms of easy adjustment of the frame rate, it can potentially track fast motion, and is in the spirit of conventional cameras.

We build on the single-pixel camera architecture [2] with independent sampling of each frame. To reduce complexity for large image sizes, we split each frame into non-overlapping blocks. Each block is independently sampled in the DCT domain to reduce complexity. The first frame in the sequence is compressively sampled [2] block-by-block and recovered at the decoder using OMP. Each next frame is compressively sampled and recovered using Alg. 1 where Λ_1 contains $S \leq K$ positions of the most significant elements in the spatially corresponding block of a previous frame. Thus a previous frame plays the role of side information.

Since the subsequent frames are correlated, reduced complexity, in terms of lower number of iterations and/or better reconstruction quality is expected. Note that the performance improvement will depend on the motion in the scene. For scenes with slow motion, several frames can use the same side information frame.

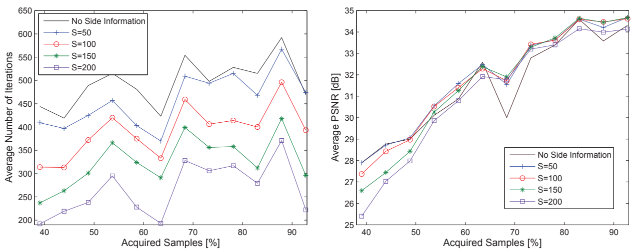
We follow the same principle in the case of the single-image acquisition for single-pixel imaging applications. We exploit spatial correlation within the image and use previously recovered blocks as side information. We split the image into non-overlapping blocks, where the first block is compressively sampled [2] and recovered at the decoder using OMP. Spatial correlation within an image is exploited by recovering other blocks using an already reconstructed neighboring block as side information.

The scheme straightforwardly applies to stereo image/video acquisition, where one view can play the role of side information. We note that it is possible to combine several neighboring blocks or frames to generate better side information.

4. RESULTS

In our simulations, we set $\kappa_1 = 0.25$ and $\kappa_2 = 0.00001$ in Algorithm 1, which empirically led to the best results. As usually done in compressive imaging literature, we use images/video already acquired at high resolution to simulate the effects if the data were acquired compressively. This serves as a good indicator of how the proposed method would perform in a real-world setup such as that of [2, 5].

First, we use a 288×288 pixel image, which was originally obtained using MRI, and then converted to the bmp file format (see Fig.2). We set the block size to $N = 32 \times 32 = 1024$ pixels. This block size was observed to provide a good tradeoff between CS efficiency, reconstruction complexity, and decoding time. We use an $M \times N$ measurement matrix with random Bernoulli +1/-1 entries, which is more realistic in this scenario [2]. The first block of the image was compressively sampled with $M < 1024$ random measurements and recovered using OMP (without side information). Each next block was sampled with M random measurements and recovered using the previous left neighbor as side information. The results as peak signal-to-noise ratio (PSNR) of the average mean square error (MSE) and the average number of iterations (over all blocks) vs. M/N are shown in Fig. 1. The curve labeled No Side Information denotes results obtained with independent recovery of each block with OMP. We vary M in the range [400, 950] with a step of 50 to obtain different sampling rates. We did not use all significant positions in the side information block, but instead used only the first S significant positions to better tradeoff complexity and quality. As expected, as S increases, the number of iterations decreases, since we rely more on side information. It can be seen that $S = 50$ and $S = 100$ perform the best. At some sampling points (e.g., $M = 700$) there is a significant quality improvement compared to the case without side information. When $S = 200$ the recovery is the fastest, but the obtained reconstruction is usually worse than without side information. The reason why smaller S's perform better in this case is, because for high S many wrong positions are accepted at the start, and the algorithm cannot correct them all. If the correlation among the blocks is higher, S can be higher (as in the video example below).



(a) (b)

Fig. 1. (a) PSNR, (b) the average number of iterations, vs. sampling rate for four different values of S and the image shown in Fig. 2.

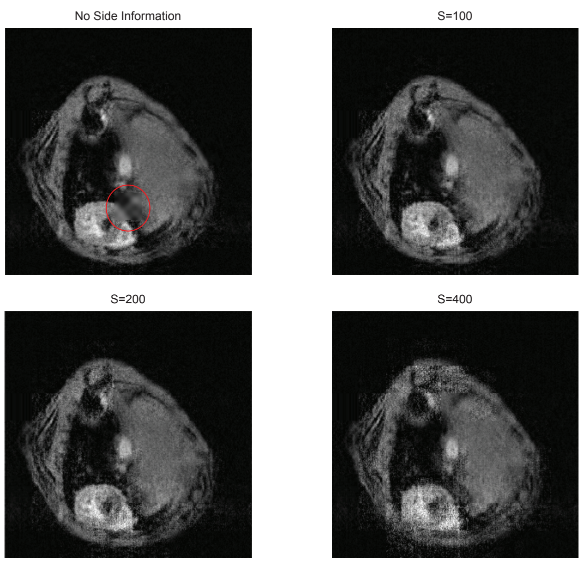
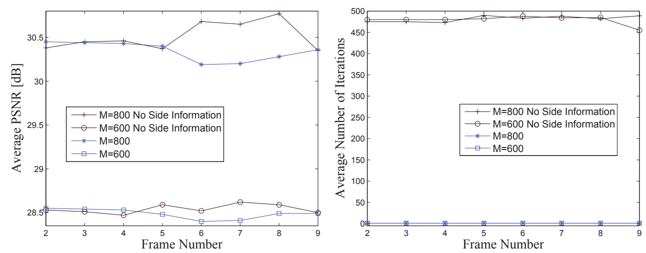


Fig. 2. MRI image recovered without side information (top left), and with side information, M =400 measurements and three different values of S.

One reconstruction example is shown in Fig. 2. We set the number of measurements per block to $M = 400$, which corresponds to 39% of acquired samples. The image without side information (OMP recovery) was reconstructed after 425 iterations, whereas the images with side information were reconstructed after 341, 192, and 40 iterations, for S =100, 200, and 400, respectively. It can be seen that in the result with the scheme without side information one block is seriously damaged, which is not the case with the proposed scheme though fewer number of iterations have been used.

Next, we simulate CS on the Y-component of the first nine frames of QCIF "Akiyo" sequence. The first frame was compressively sampled and recovered using OMP (without side information). It is then used for recovery of all subsequent 8 frames. The results as PSNR of the average MSE and the average number of iterations vs. the frame number are shown in Fig. 3. Curves labeled No Side Information denote results obtained with independent recovery of each frame with OMP. Results are shown for $M = 800$ and $M = 600$ measurements per block (similar results are obtained for different M), which correspond to 78% and 59% of acquired samples, respectively. We always used all significant positions in the side information frame, i.e., $S = K$. Roughly the same PSNR was obtained compared to the No Side Information case, with significantly fewer number of iterations for all frames. Note that due to slow motion of the "Akiyo", it is possible to use one frame as side information for a long sequence of subsequent frames.



(a) (b)

Fig. 3. (a) PSNR, (b) the average number of iterations, vs. the frame number for the "Akiyo" video sequence.

5. CONCLUSIONS AND FUTURE WORK

We propose the scheme for compressive imaging using the concept of side information. The proposed concept of recovery with side information has potential to significantly reduce complexity and improve performance of compressive image sampling by exploiting spatial and temporal correlation within image/video.

One future task is to develop a new compression scheme that should follow compressive image acquisition. Instead of using the spatially corresponding blocks as side information, research can go into the direction of performing different motion estimation techniques to better predict block sparsity. The main challenge that still remains is building a real-world system for compressive image/video sampling based on the proposed methods.

6. REFERENCES

[1] *IEEE Signal Processing Magazaine, Special Issue on Compressive Sampling,* March 2008.

[2] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R.G. Baraniuk, "Compressive imaging for video representation and coding," *Proc. PCS06 Picture Coding Symposium,* Beijing, China, April 2006.

[3] M.F. Duarte, M.B. Wakin, and R.G. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," *Proc. SPARS2005 Signal Processing with Adaptative Sparse Structured Representations,* Rennes, France, Nov. 2005.

[4] J. Tropp and A.C. Gilbert, "Signal Recovery from Random Measurements via Orthogonal Matching Pursuit," *IEEE Trans. Inform. Theory,* vol. 53, pp. 4655–4666, Dec. 2007.

[5] M. Lustig, D. Donoho, and J.M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine,* vol. 58, pp. 1182–1195, Dec. 2007.

[6] M.A. Neifeld and J. Ke, "Optical architectures for compressive imaging," *Applied Optics,* vol. 46, pp. 5293–5303, Aug. 2007.

[7] V. Stanković, L. Stanković, and S. Cheng, "Compressive video sampling," *Eusipco2008* , Lausanne, Switzerland, Aug. 2008.