VIDEO COMPRESSIVE SENSING USING MULTIPLE MEASUREMENT VECTORS

*Michael Iliadis, Jeremy Watt, Leonidas Spinoulas, Aggelos K. Katsaggelos**

Dept. of Electrical Engineering and Comp. Sc., Northwestern University, Evanston, IL 60208, USA

ABSTRACT

Compressive Sensing (CS) suggests that, under certain conditions, a signal can be reconstructed using a small number of incoherent measurements. We propose a novel video CS framework based on Multiple Measurement Vectors (MMV) which is suitable for signals with temporal correlation such as video sequences. In addition, a CS circulant matrix is employed for fast reconstruction. Furthermore, the proposed framework allows the number of CS measurements associated with each frame to be chosen in the decoder rather than the encoder offering robustness compared to the multi-scale approaches. Experimental results on two video sequences exhibiting fast motion and occlusions, show the advantages of the proposed method over the current state-of-the-art in video CS.

Index Terms— Video compressive sensing, multiple measurement vectors, circulant matrix, fast motion.

1. INTRODUCTION

Compressive Sensing (CS) has become increasingly popular in recent years and CS theory has been incorporated in various applications [1, 2]. CS theory suggests that a signal can be perfectly reconstructed using a small number of random incoherent linear projections. In other words, one can sample well-below the Nyquist rate and still be able to reconstruct a signal. An underlying assumption is that the signal needs to be sparse in some transform domain. Many signals, such as natural images, are sparse in well-known bases (e.g., Wavelet).

Multiple algorithms have been proposed for reconstructing still images using CS. However, for time-varying scenes, reconstruction becomes far more challenging. In this case, in every few CS measurements a different scene is captured by a CS imaging system such as the Single Pixel Camera (SPC) [3]. More recently, several methods exploit the motion of the sequence in order to find sparser solutions [4,5] while other approaches use multi-scale frameworks [6,7].

This paper is organized as follows. We present the modeling of CS and Multiple Measurement Vectors (MMV) in Section 2. In Section 3, we overview the existing video CS literature. The proposed video CS algorithm is analyzed in Section 4. Finally, experimental results and discussion about the performance of our framework are presented in Section 5 and conclusions are drawn in Section 6.

2. VIDEO COMPRESSIVE SENSING AND MMV

The standard Single Measurement Vector (SMV) CS acquisition model is given by,

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the $N \times 1$ desired solution vector, $\mathbf{n}$ is the $M \times 1$ observation noise, $\mathbf{\Phi}$ is the $M \times N$ measurement matrix and $\mathbf{y}$ is the $M \times 1$ measurement vector. The desired solution $\mathbf{x}$ has dimensions of $m \times n = N$ and we assume that $M \ll N$. CS theory states that the desired solution vector $\mathbf{x}$ needs to be sparse in a basis $\mathbf{\Psi}$ such that $\mathbf{x} = \mathbf{\Psi}\theta$, where θ has a few non-zero coefficients.

Many studies, [8–13], have shown that the MMV model outperforms the SMV model when the solution vectors have the same sparsity structure (i.e., the indices of the nonzero entries are the same for all solution vectors). The advantage of the MMV approach over the SMV is the ability to attain sparser solutions. In this paper the SMV CS model in (1) is modified to the MMV model for a video setting,

$$\mathbf{Y} = \mathbf{\Phi X} + \mathbf{N}, \tag{2}$$

where $\mathbf{X} = [\mathbf{x}^1, ..., \mathbf{x}^L]$ is a $N \times L$ matrix containing the solution vector frames $\mathbf{x}^i$ of size $N \times 1$, and $\mathbf{Y} = [\mathbf{y}^1, ..., \mathbf{y}^L]$ is a $M \times L$ matrix containing the measurement vectors, where $\mathbf{y}^i$ has size $M \times 1$. L is the number of frames to be recovered, and $\mathbf{N}$ is the noise matrix with i.i.d. Gaussian entries. Since it is assumed that the solution vectors are sparse in a basis $\mathbf{\Psi}$ we can define $\mathbf{X} = \mathbf{\Psi\Theta}$, where $\mathbf{\Theta} = [\theta^1, ..., \theta^L]$ and θ^i is a $N \times 1$ sparse vector. It seems natural to assume that the MMV method could increase the recovery performance of signals exhibiting temporal correlations such as in video.

The solution vector frames can be recovered by solving the following convex-optimization problem [8],

$$\begin{aligned} &\min \sum_{k=1}^{N} ||\theta_k||_2, \\ &\text{subject to} \left\|\mathbf{Y} - \mathbf{\Phi\Psi\Theta}\right\|_F \leq \epsilon, \end{aligned} \tag{3}$$

where F is the Frobenius matrix norm, and θ_k is the k^{th} row of $\mathbf{\Theta}$. Thus, the objective term of the MMV approach is the sum of the ℓ_2-norm of the rows of $\mathbf{\Theta}$ and is used to promote zero rows in $\mathbf{\Theta}$.

3. RELATED WORK

Early attempts of CS video reconstruction use 3D sparsifying transforms such as the 3D discrete Wavelet transform, [14], where the entire video is reconstructed at once by stacking all sequential CS measurements in a vector. The main drawback of this approach is that computational time and memory requirements are intensified since all frames are recovered simultaneously. A different approach, [15], aims at the reconstruction of the inter-frame difference of two or more subsequent frames. Since consecutive frames are similar, the inter-frame difference is expected to be sparser compared to the actual frame.

In video sequences of natural scenes, where non-stationary translational object motions are present, the aforementioned techniques fail to exploit temporal redundancies [15, 16]. For such

*This work was supported by the Department of Energy under grant DE-0000457

sequences, CS video reconstruction methods that utilize motion estimation and compensation techniques have been proposed [4–7]. Accurate frame reconstruction is necessary for faithful inter-frame motion estimation, while precise motion estimates can significantly improve reconstruction quality. This is often referred to as the "chicken-and-egg" problem [6].

Several attempts towards this direction have been proposed. K-t FOCUSS algorithm, [5], is used for MRI video reconstruction under the assumption of slow motion. A multi-scale framework has been proposed in [7] by acquiring CS measurements of video in different spatial resolutions. Low resolution inter-frame motion estimates are then used to enhance reconstruction quality at higher resolutions.

In previous studies, it is assumed that the CS imager is capable of acquiring a large number of measurements per frame, sufficient for acceptable reconstruction. This cannot necessarily be considered as a real-world model since the observed scene can exhibit fast motion, exceeding the imager's capabilities in acquisition speed. That is, the imager may not be able to acquire the necessary number of measurements per frame. A study where the limitations on imager's acquisition speed were considered is the CS-MUVI [6]. The authors limit the amount of measurements per frame while exploiting motion information to improve performance. In particular, the optical-flow estimates extracted by a low-resolution version of the video are used for the final full-resolution reconstruction.

4. PROPOSED METHOD

We propose a novel framework for reconstructing video sequences from CS measurements when in every few acquisitions a slightly different scene is captured. Our framework is based on MMV. We first reconstruct the entire sequence by solving the problem in (3), then we calculate inter-frame motion-estimates and finally we refine the reconstruction of the video sequence using the motion estimates as additional constraints. This problem can be described as,

$$\begin{aligned} &\min \sum_{k=1}^{N} ||\theta_k||_2, \\ &\text{subject to} \left\|\mathbf{Y} - \mathbf{\Phi\Psi\Theta}\right\|_F \leq \epsilon_1, \\ &\left\|\mathbf{x}^i(a,b) - \mathbf{x}^j(a+u,b+v)\right\|_2 \leq \epsilon_2, \forall\, i,j, \end{aligned} \tag{4}$$

where $\mathbf{\Psi}$ is selected as the Wavelet basis and $\|\cdot\|_2 \leq \epsilon_2$ corresponds to the incorporated motion estimation constraints. In particular, $\mathbf{x}^i(a,b)$ indicates the value of frame i at position (a,b) and u, v indicate pixel translations between the i^{th} and j^{th} frames in the vertical (horizontal) directions. We also note that $\epsilon_1 \geq 0$ and $\epsilon_2 \geq 0$ can be adjusted in order to improve recovery performance.

A related method to our video recovery framework is described in [17]. In this study, motion estimates extracted from an initial reconstruction of the sequence are used as additional constraints in the final reconstruction. However, a key difference is that the CS acquisitions are obtained from a temporal rather than spatial multiplexing camera. In this study we focus on reconstructing video sequences that have been acquired by a SPC configuration.

The method presented in (4) is solved in three separate steps,

1. We reconstruct the frames $\mathbf{x}^i$, $\forall\, i = 1,\ldots,L$ solving the MMV convex optimization problem in (3). Note that, in this step, we select $W \geq M$ measurements such that $\mathbf{Y} = [\mathbf{y}^1,...,\mathbf{y}^L]$ is a $W \times L$ matrix which contains the measurement vectors $\mathbf{y}^i$ of size $W \times 1$. Thus, the measurement matrix $\mathbf{\Phi}$ has dimensions $W \times N$. See section 4.1.2 for more details on the selection of W.
2. We calculate the optical-flow between the recovered frames from step (1.). Specifically, we use the same optical-flow algorithm, as in [6], since we want to compare our methods. However, we note that one can use any existing motion estimation algorithm for this step. The optical-flow algorithm we use, [18], is invariant to occlusions and sub-pixel translations are approximated by the four closest neighboring pixels. See [6] for more details on the optical-flow setup.
3. Finally, we solve the proposed constrained optimization problem in (4) to acquire the desired reconstructed frames.

4.1. Implementation Details

4.1.1. Hadamard Circulant Matrix

One important practical aspect of the video CS reconstruction is to keep the computational time and memory requirements as low as possible. For example, it seems very impractical to store the entire measurement matrix especially for sequences with high-dimensional frames. Our framework enables the usage of circulant matrices as proposed in [19]. The structure of a circulant measurement matrix $\mathbf{\Phi}'$ is defined as,

$$\mathbf{\Phi}' = \begin{bmatrix} a_N & a_{N-1} & \ldots & a_1 \\ a_1 & a_N & \ldots & a_2 \\ \ddots & \ddots & \ddots & \\ a_{N-1} & a_{N-1} & \ldots & a_N \end{bmatrix}. \tag{5}$$

In our framework, the first row of $\mathbf{\Phi}'$ is generated as,

$$\mathbf{a} = (U\mathbf{h})^T, \tag{6}$$

where U is a $N \times N_d$ up-sampling operator and $\mathbf{h}$ is a $N_d \times 1$ vectorized version of a $N_d = n_d \times n_d$ Hadamard matrix. Note that $N_d < N$.

Let $\mathbf{b} = \mathbf{F}^*\mathbf{a}$, where $\mathbf{F}$ is the $N \times N$ discrete Fourier matrix, and $\mathbf{C_b}$ be a diagonal matrix with the entries of $\mathbf{b}$ as its diagonal components. Therefore, $\mathbf{\Phi}'$ can be written as,

$$\mathbf{\Phi}' = \mathbf{F}^{-1}\mathbf{C_b}\mathbf{F}. \tag{7}$$

One of the key advantages of this sampling scheme is that one only needs to store the first row of $\mathbf{\Phi}'$. In addition, the product $\mathbf{\Phi}'\mathbf{x}$ can be efficiently performed through the Fast Fourier Transform (FFT). The final CS matrix $\mathbf{\Phi}$ of size $M \times N$ is constructed as $\mathbf{\Phi} = P\mathbf{\Phi}'$ where P represents a $M \times N$ random-subsampling operator for better incoherence. Furthermore, it has been proven that, in terms of incoherence, CS circulant matrices are as effective as i.i.d. Gaussian random matrices [19]. Also, circulant matrices are known to be incoherent with any orthonormal basis $\mathbf{\Psi}$ with high probability [20] and can be easily implemented through an optical imaging system [19].

4.1.2. Frame Rate Parameters

In video CS, the input to the decoder is a sequence of CS measurements acquired during a certain period of time. A parameter to be chosen is the number of frames to be recovered (L). This can be determined by the number of CS measurements associated with each frame. Let the encoder generate a certain number of CS measurements per second, M be the number of measurements per frame and T be the total number of acquired CS measurements. Thus, the total number of frames would be $L = T/M$. We desire to recover

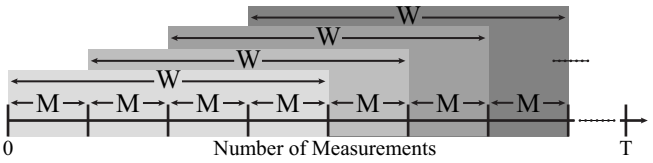

Fig. 1: Graphical representation of the relation between the selected parameters W and M of our proposed method.

frames with small or no motion-blur, i.e., solely through measurements acquired while the scene was static. Hence, if we assume that the scene exhibits fast motion, M should take small values. However, if M is too small, reconstruction performance could deteriorate dramatically. On the contrary, increasing M would increase motion-blur effects in each reconstructed video frame.

Another parameter to be selected in our framework is the number of CS measurements per frame for the initial recovery of the sequence (W). As mentioned before, $W \geq M$. The main objective of the initial frame recovery is the extraction of motion estimates. Therefore, it is desirable that the initial recovery preserves inter-frame motion information. Choosing $M = W$ might not always be the best option since M can be very small, as discussed above. On the other hand, we could allow a little motion-blur on the initial recovery by choosing $W > M$. In this case, additional CS measurements would be used for the initial frame reconstruction and thus motion information would be better preserved providing increased accuracy in the optical-flow estimates. It is difficult to determine the exact W that would lead to the best motion estimates but as a general rule we choose a large enough W such that the best motion information is obtained after the initial CS recovery.

Figure 1 represents graphically the relation between the selected parameters W and M in our framework. W and M correspond to the number of CS measurements used per frame, as acquired over time, for the initial and final reconstruction steps, respectively (steps (1.) and (3.) of our proposed method). Note that W is selected with significant overlap over subsequent frames to increase reconstruction quality and better preserve motion information, essential for the optical-flow calculation step (step (2.) of our proposed method).

Choosing appropriate values for M and W is critical for the video sequence reconstruction performance since motion speed can vary in different sequences. Therefore, as opposed to [6, 7], it is advantageous to choose these parameters, according to the speed of the sequence, at the decoder rather than at the encoder.

5. EXPERIMENTAL RESULTS

The performance of the proposed method is evaluated on two video sequences captured by a high-speed video camera that can obtain 250 frames per second, as reported in [6]. The total number of acquired frames is 2048. The video sequences exhibit fast motion and occlusions and each frame has size of 256×256 pixels. We simulated CS by obtaining 16 measurements per frame acquiring in total $T = 16 \times 2048 = 32768$ measurements from each video sequence. To the best of our knowledge, this is the first study that attempts to reconstruct a video sequence by acquiring so few CS measurements per frame. Therefore, we assume that the CS imaging system acquires $16 \times 250 = 4000$ CS measurements per second. In order to construct the first row of $\mathbf{\Phi}$, as described in 4.1.1, we use nearest-neighbor up-sampling and $N_d = 4096$. As described in section 4.1.2, M and W can be chosen at the decoder. For M it is desirable to be as small as possible, so that little motion would be incorporated in the final recovered frames. We choose $M = 512$ for both sequences so that the number of frames to be recovered is $L = 64$. The reconstruction performance in the final recovery step is primarily determined by the accuracy of the motion estimates. As stated in section 4.1.2, we choose a large enough W such that motion information is preserved after the initial CS reconstruction. Since both sequences contain fast motion we choose $W = 2048$. For the CS-MUVI framework, [6], we perform the experiments with $W = 4096$. In this framework the W parameter is fixed and coupled to the encoder, as opposed to ours, and hence it cannot be tuned at the decoder according to the speed of the sequence.

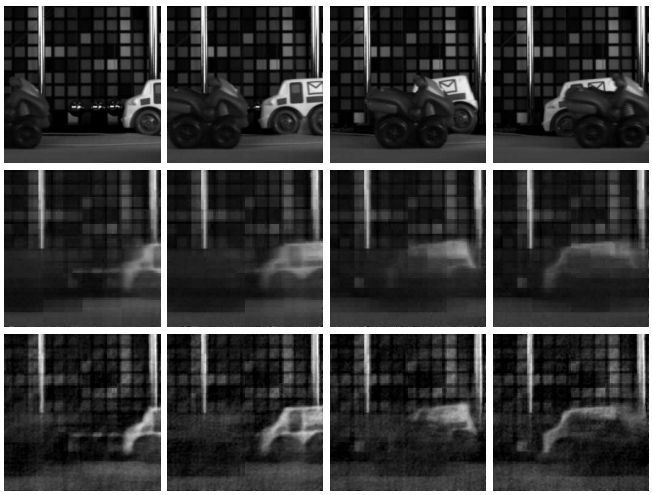

Fig. 2: Example reconstructions from the "Two Cars" sequence: 1^{st} row represents the original frames; 2^{nd} row represents the reconstruction through CS-MUVI [6] and 3^{rd} row corresponds to our proposed method.

The extraction of the optical-flow estimates is performed using [18] and the MMV optimization problems (3), (4) are solved using SPGL1 [21]. Furthermore, we add noise to the CS measurements such that the signal-to-noise ratio (SNR) equals 60dB (same noise level as in [6]).

As a performance metric, we adopt the peak signal to noise ratio (PSNR). For video sequences, PSNR is defined as,

$$\mathrm{PSNR(i)} = 10 log_{10} \frac{NR^2}{\|\mathbf{b}^i - \mathbf{x}^i\|^2}, \tag{8}$$

where $\mathbf{x}^i$ denotes the i^{th} estimated frame and R denotes the maximum possible intensity value in $\mathbf{x}^i$. Then, the overall PSNR value for the video can be calculated as,

$$\mathrm{PSNR} = \frac{1}{L} \sum_{i=1}^{L} \mathrm{PSNR(i)}. \tag{9}$$

In order to calculate PSNR for video sequences we need the number of original frames to be equal with the number of recovered frames. Since in this paper these two numbers differ, we averaged the 2048 original frames as,

$$\mathbf{b}^i = \frac{1}{c} \sum_{n=(i-1)c+1}^{ic} f_n, \tag{10}$$

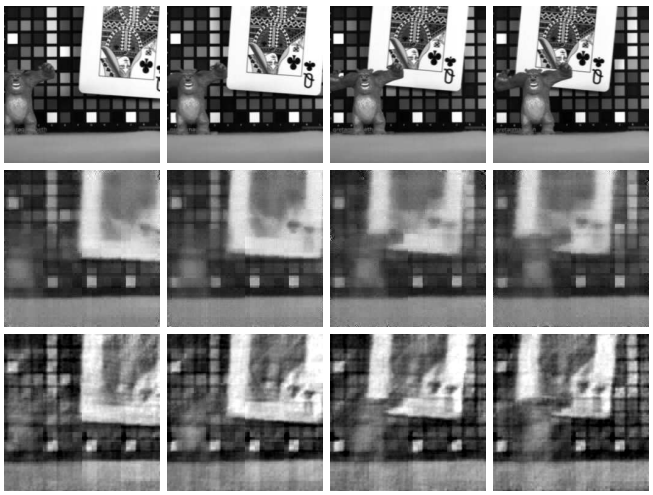

Fig. 3: Example reconstructions from the "Card and Monster" sequence: 1^{st} row represents the original frames; 2^{nd} row represents the reconstruction through CS-MUVI [6] and 3^{rd} row corresponds to our proposed method.

Table 1: PSNR values and computational time between CS-MUVI, [6], and our proposed method for the two video sequences with 16 CS measurements per frame, $M = 512$, $N = 65536$.

Video Sequences	PSNR (dB)		Time (hours)	
	CS-MUVI	Our Method	CS-MUVI	Our Method
Two Cars	20.61	21.81	3.0	1.9
Card and Monster	16.08	17.32	3.0	1.8

where $n \in \{1, \ldots, 2048\}$, f_n denotes the original frame, $c = 2048/L$ and $\mathbf{b}^i$ denotes the mean original frame and $i \in \{1, \ldots, L\}$.

Example reconstructions are shown in Figure 2 and Figure 3. Also, in Figure 4 we present the initial reconstruction results when the first row of $\mathbf{\Phi}$, is constructed as described in section 4.1.1, takes i.i.d. Gaussian random values, or when the reconstruction is performed using the SMV instead of the MMV approach. For the SMV approach the frames are recovered solving the convex-optimization problem, as defined in [6],

$$\begin{aligned} &\min \sum_{i=1}^{L} \left\| \mathbf{\Psi^T} \mathbf{x}^i \right\|_1, \\ &\text{subject to} \left\| \mathbf{y}^i - \mathbf{\Phi}^i \mathbf{x}^i \right\|_2 \leq \epsilon_1, \forall\, i, \end{aligned} \tag{11}$$

where $i \in \{1, \ldots, L\}$, $\mathbf{x}^i$ is a solution vector of the i^{th} frame of size $N \times 1$, $\mathbf{\Phi}^i$ is the $W \times N$ measurement matrix and $\mathbf{y}^i$ is the observation vector of size $W \times 1$. Theory suggests that MMV is beneficial over SMV when the solution vectors have the same sparsity structure, i.e., the indices of the nonzero entries are the same for all solution vectors [8–13]. In this paper the solution frames do not have exactly the same but very similar sparsity structure since every solution vector corresponds to a slightly different scene. Nevertheless, it is evident in Figure 4 that the MMV approach outperforms the SMV. Also, it is evident that the up-sampling Hadamard first row of $\mathbf{\Phi}$ performs better than the i.i.d. Gaussian case.

Fig. 4: Initial reconstruction of our framework ("Two Cars" sequence): 1^{st} row shows results (PSNR = 20.73dB) when the first row of $\mathbf{\Phi}$ is constructed as defined in 4.1.1; 2^{nd} row shows results (PSNR = 19.27dB) based on SMV instead of MMV when the first row of $\mathbf{\Phi}$ is constructed as defined in 4.1.1; 3^{rd} row shows results (PSNR = 19.10dB) when the first row of $\mathbf{\Phi}$ is constructed using i.i.d. Gaussian random values and the MMV approach is used.

Finally, as dictated by Table 1 and Figures 2 and 3, the proposed method outperforms the method in [6] while reducing computational time. Moreover, the reconstructed frames through our method contain less motion blur than the CS-MUVI results and finer details can be observed in the scene. All experiments were performed on a quad-core computer with 8GB RAM.

6. CONCLUSIONS

In this paper we presented a novel video compressive sensing (CS) framework based on the Single Pixel Camera (SPC). The proposed approach takes advantage of Multiple Measurement Vectors (MMV), seeking for significantly sparser solutions, assuming that the solution vectors have similar sparsity structure. This assumption is beneficial in video sequences exhibiting temporal correlations between frames where the MMV approach was proven superior to the Single Measurement Vector (SMV) one. Our framework performs fast video reconstruction using circulant matrices. In addition, the number of CS measurements associated with each frame can be chosen in the decoder rather than the encoder which is an advantage over proposed multi-scale approaches. We performed experiments using two video sequences with fast motion and occlusions and showed that our method outperforms the state-of-the-art algorithm not only in terms of PSNR and computational time but also in terms of the visual quality of the recovered frames.

7. ACKNOWLEDGMENT

We wish to thank Dr. Aswin Sankaranarayanan for providing the video sequences used in the experimental section as well as the source code of CS-MUVI [6].

8. REFERENCES

[1] Y. Tsaig and D. L. Donoho, "Compressed sensing," *IEEE Transactions on information Theory*, vol. 52, pp. 1289–1306, 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489– 509, 2006.

[3] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, 2008.

[4] M. Sungkwang and J. E. Fowler, "Residual reconstruction for block-based compressed sensing of video," in *Data Compression Conference (DCC)*, 2011, pp. 183 –192.

[5] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009.

[6] A. C. Sankaranarayanan, C. Studer, and R. G. Baraniuk, "CS-MUVI: Video compressive sensing for spatial-multiplexing cameras," in *IEEE International Conference on Computational Photography (ICCP)*, 2012, pp. 1 –10.

[7] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Proceedings of Picture Coding Symposium (PCS)*, 2009, pp. 1 –4.

[8] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2477–2488, 2005.

[9] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[10] D. P Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[11] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[12] J. Yuzhe and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2008, pp. 3921 –3924.

[13] Z. Zhang and B.D. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2010, pp. 3986 –3989.

[14] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Proceedings of Picture Coding Symposium (PCS)*, 2006.

[15] R. F. Marcia and R. M. Willett, "Compressive coded aperture video reconstruction," in *European Signal Processing Conference*, 2008.

[16] J. E. Fowler, S. Mun, and E. W. Tramel, "Block-based compressed sensing of images and video," *Found. Trends Signal Processing*, vol. 4, no. 4, pp. 297–416, 2012.

[17] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 329 –336.

[18] C. Liu, *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*, Ph.D. thesis, Massachusetts Institute of Technology, 2009.

[19] W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with toeplitz and circulant matrices," *Proceedings of SPIE*, vol. 7744, pp. 77440K–77440K–10, 2010.

[20] J. Romberg, "Compressive sensing by random convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.

[21] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.