BAYESIAN COMPRESSED SENSING IMAGING USING A GAUSSIAN SCALE MIXTURE

George Tzagkarakis and Panagiotis Tsakalides

Department of Computer Science, University of Crete & Institute of Computer Science - FORTH
e-mail: {*gtzag, tsakalid*}@*ics.forth.gr*

ABSTRACT

The ease of image storage and transmission in modern applications would be unfeasible without compression, which converts high-resolution images into a relatively small set of significant transform coefficients. Due to the specific content of many real-world images they are highly sparse in an appropriate orthonormal basis. The inherent property of compressed sensing (CS) theory working simultaneously as a sensing and compression protocol, using a small subset of random incoherent projection coefficients, enables a potentially significant reduction in the sampling and computation costs of images favoring its use in real-time applications which do not require an excellent reconstruction performance. In this paper, we develop a Bayesian CS (BCS) approach for obtaining highly sparse representations of images based on a set of noisy CS measurements, where the prior belief that the vector of projection coefficients should be sparse is enforced by fitting directly its prior probability distribution by means of a Gaussian Scale Mixture (GSM). The experimental results show that our proposed method, when compared with norm-based constrained optimization algorithms, maintains the reconstruction performance, in terms of the reconstruction error and the PSNR, while achieving an increased sparsity using much less basis functions.

1. INTRODUCTION

During the last years, image data increase at an explosive rate due to the increased capabilities of modern digital imaging devices. This huge amount of data necessitates the development of efficient compression techniques and standards [1, 2]. However, there are cases where even higher compression rates would suffice to carry out specific tasks, such as image classification and retrieval, where a high-quality reconstruction of the still images is not necessary.

Several studies [3] have shown that many natural images result in a highly sparse representation by projecting on localized orthonormal basis functions (e.g., wavelets, sinusoids). The traditional compression techniques exploit only a small subset of large-amplitude transform coefficients. However, this is an inherently wasteful process (in terms of both sampling rate and computational complexity), since one gathers and processes the entire image even though an exact representation is not required explicitly.

Compressed sensing (CS) is a framework for simultaneous sensing and compression [4, 5] enabling a potentially significant reduction in the sampling and computation costs at a sensing system with limited capabilities. In particular, a signal having a sparse representation in a transform basis can be reconstructed from a small set of projections onto a second, measurement basis that is incoherent with the first one. Thus, CS provides a simple compression scheme with low computational complexity, which is *asymmetrical*, with the compression part consisting of simple linear projections, while the main computational cost is on the decompression part where increased computational and storage resources are available.

The majority of previous CS methods for sparse representation and reconstruction of a signal in an over-complete dictionary solve constraint-based optimization problems. Several recent studies exploit the sparsity of images using CS to increase the compression rates [6, 7]. Recently, a Bayesian CS (BCS) framework was introduced [8] resulting in certain improvements when compared with norm-based CS methods, by employing a hierarchical model as a sparsity-enforcing prior distribution on the transform coefficients vector. In the present work, we replace this hierarchical prior by modeling directly the coefficients vector using a Gaussian Scale Mixture (GSM). The experimental results reveal that this approach yields a significantly sparser representation of several images, which is our primary objective, while maintaining a high reconstruction performance.

The paper is organized as follows: in Section 2, we briefly review the main concepts of BCS, extend the standard BCS approach by incorporating a GSM as the sparsity-enforcing prior model and discuss its potential advantages when compared with constraint-based optimization methods for obtaining even sparser representations of images with distinct content in an over-complete dictionary. In Section 3, we compare the performance of the proposed approach in terms of the degree of sparsity and the reconstruction quality with several other CS recovery methods, while we conclude in Section 4.

2. BAYESIAN CS RECONSTRUCTION

Let $\mathbf{\Psi}$ be a $N \times N$ matrix, whose columns correspond to the transform basis functions. Then, a given image $\vec{f} \in \mathbb{R}^N$ (reshaped as a column vector) can be represented as $\vec{f} = \mathbf{\Psi}\vec{w}$, where $\vec{w} \in \mathbb{R}^N$ is the weight vector. Obviously, $\vec{f}$ and $\vec{w}$ are equivalent representations of the image, with $\vec{f}$ being in the space domain and $\vec{w}$ in the (transform) $\mathbf{\Psi}$ domain. As mentioned above, for many natural images the majority of the components of $\vec{w}$ have negligible amplitude. In particular, $\vec{f}$ is L-sparse in basis $\mathbf{\Psi}$ if the corresponding weight vector $\vec{w}$ has L non-zero components ($L \ll N$). In a real-world scenario $\vec{f}$ is not strictly L-sparse, but it is said to be *compressible* when the re-ordered components of $\vec{w}$ decay at a power-law.

Consider also an $M \times N$ ($M < N$) measurement matrix $\mathbf{\Phi}$ (the over-complete dictionary) with its rows being incoherent with the columns of $\mathbf{\Psi}$. For instance, let $\mathbf{\Phi}$ be a Hadamard matrix or contain independent and identically distributed (i.i.d.) Gaussian entries. Such matrices are incoherent with any fixed transform matrix $\mathbf{\Psi}$ with high probability (universality property) [5].

If $\vec{f}$ is compressible in $\mathbf{\Psi}$, then it is possible to perform directly

This work was funded by the Greek General Secretariat for Research and Technology under Program ΠΕΝΕΔ-Code 03ΕΔ69 and by the Marie Curie TOK-DEV "ASPIRE" grant (MTKD-CT-2005-029791) within the 6th European Community Framework Program.

a compressed set of measurements $\vec{g}$, resulting in a simplified image acquisition system. The original image $\vec{f}$ and the CS measurements $\vec{g}$ are related through random projections, $\vec{g} = \mathbf{\Phi}\mathbf{\Psi}^T\vec{f} = \mathbf{\Phi}\vec{w}$, where $\mathbf{\Phi} = [\vec{\phi}_1, \ldots, \vec{\phi}_M]^T$ and $\vec{\phi}_m \in \mathbb{R}^N$ is a random vector with i.i.d. components. Thus, a sparse representation of $\vec{f}$ from $\vec{g}$ reduces to estimating a weight vector $\vec{w}$ with as few non-zero components as possible, which then can be used for reconstructing $\vec{f}$.

Most of the recent CS literature has concentrated on constrained optimization-based methods for sparse signal representation. For instance, in a real-world scenario with CS measurements corrupted by additive noise $\vec{\eta}$ with unknown variance σ_η^2, $\vec{g} = \mathbf{\Phi}\vec{w} + \vec{\eta}$, the ℓ_1-norm minimization approach seeks a sparse vector $\vec{w}$ by solving the following optimization problem,

$$\tilde{\vec{w}} = arg\min_{\vec{w}} \|\vec{w}\|_1 \ , \ s.t. \ \|\vec{g} - \mathbf{\Phi}\vec{w}\|_\infty \leq \epsilon \ , \tag{1}$$

where ϵ is the noise level ($\|\vec{\eta}\|_2 \leq \epsilon$). The main approaches for the solution of such optimization problems include linear programming [9] and greedy algorithms [10], resulting in a *point estimate* of the weight vector $\vec{w}$.

On the other hand, when the CS inversion is treated from a Bayesian perspective, then given a prior belief that $\vec{w}$ is sparse in basis $\mathbf{\Psi}$ and the set of CS measurements $\vec{g}$, the objective is to formulate a *posterior probability distribution* for $\vec{w}$. This improves the accuracy over a point estimate and provides confidence intervals (error bars) in the approximation of $\vec{f}$, which can be used to guide the optimal design of additional CS measurements with the goal of reducing the uncertainty in reconstructing $\vec{f}$.

Under the common assumption of a zero-mean Gaussian noise we obtain the following Gaussian likelihood model,

$$p(\vec{g}|\vec{w}, \sigma_\eta^2) = (2\pi\sigma_\eta^2)^{-M/2} \cdot \exp\left(-\frac{1}{2\sigma_\eta^2}\|\vec{g} - \mathbf{\Phi}\vec{w}\|\right) . \tag{2}$$

Assuming that $\mathbf{\Phi}$ is known, the quantities to be estimated, given $\vec{g}$, are the sparse vector $\vec{w}$ and the noise variance σ_η^2. This is equivalent to seeking a full posterior density function for $\vec{w}$ and σ_η^2. Thus, given $\vec{g}$ and assuming the Gaussian likelihood model in (2), it is straightforward to see that the solution of (1) corresponds to a maximum *a posteriori* (MAP) estimate for $\vec{w}$.

In a probabilistic framework the assumption that $\vec{w}$ is sparse is formalized by modeling its distribution using a sparsity-enforcing prior. A common choice of this prior is the Laplace density [11]. However, its usage raised the problem that the Bayesian inference may not be performed in closed form, since the Laplace prior is not conjugate[1] to the Gaussian likelihood model. The treatment of the CS measurements $\vec{g}$ from a Bayesian viewpoint, while overcoming the problem of conjugateness, was introduced in [8] by employing a hierarchical model, which had similar properties as the Laplace but allowed convenient conjugate-exponential analysis [12].

2.1. BCS reconstruction using Gaussian scale mixture priors

In the present study, the estimation of $\vec{w}$ is also performed in a Bayesian framework. However, in contrast to the previous BCS work, our proposed method consists in modeling directly the prior of $\vec{w}$ with a heavy-tailed distribution, which promotes its sparsity, since it is suitable for modeling highly impulsive signals. This is motivated by the fact that the content of many natural images is often well structured (e.g., containing edges), and thus $\vec{w}$ can be considered as highly impulsive, since it is characterized by a large number of small-amplitude components and a small number of large-amplitude components. For this purpose, we replace the approximate hierarchical process by modeling directly the prior distribution of $\vec{w}$ by means of a Gaussian Scale Mixture (GSM).

Definition 1 *A vector $\vec{w}$ is called a GSM (in $\mathbb{R}^N$) with underlying Gaussian vector $\vec{G}$ iff it can be written in the form $\vec{w} = \sqrt{A}\,\vec{G}$, where A is a positive random variable and $\vec{G} = (G_1, G_2, \ldots, G_N)$ is a zero-mean Gaussian random vector, independent of A, with covariance matrix $\mathbf{\Sigma}$.*

In the subsequent analysis we consider that the components of $\vec{G}$ are also independent, yielding a diagonal covariance matrix $\mathbf{\Sigma} = diag(\sigma_1^2, \ldots \sigma_N^2)$. From definition 1, the density of $\vec{w}$ conditioned on the variable A is a zero-mean multivariate Gaussian given by,

$$p(\vec{w}|A) = \frac{\exp(-\frac{1}{2}\vec{w}^T(A\mathbf{\Sigma})^{-1}\vec{w})}{(2\pi)^{N/2}|A\mathbf{\Sigma}|^{1/2}} \ , \tag{3}$$

where $|\cdot|$ denotes the determinant of a matrix. From (3), we obtain the following maximum likelihood (ML) estimate of the variable A,

$$\hat{A}(\vec{w}) = \left(\vec{w}^T\mathbf{\Sigma}^{-1}\vec{w}\right)/N \ . \tag{4}$$

Assuming that the noise variance σ_η^2, the value of A and the covariance matrix $\mathbf{\Sigma}$ have been estimated, given the CS measurements $\vec{g}$ and the matrix $\mathbf{\Phi}$, the posterior of $\vec{w}$ is given by the Bayes' rule, combining the likelihood and the prior density functions and exploiting the independence of A and $\vec{G}$, as follows:

$$p(\vec{w}|\vec{g}, A, \mathbf{\Sigma}, \sigma_\eta^2) = \frac{p(\vec{g}|\vec{w}, \sigma_\eta^2)p(\vec{w}|A, \mathbf{\Sigma})}{p(\vec{g}|A, \mathbf{\Sigma}, \sigma_\eta^2)} \ , \tag{5}$$

which is a multivariate Gaussian distribution with a mean vector $\vec{\mu}$ and a covariance matrix $\mathbf{P}$, given by,

$$\vec{\mu} \ = \ \sigma_\eta^{-2}\mathbf{P}\mathbf{\Phi}^T\vec{g} \ , \tag{6}$$

$$\mathbf{P} \ = \ (\sigma_\eta^{-2}\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{M})^{-1} \ , \tag{7}$$

where $\mathbf{M} = diag((A\sigma_1^2)^{-1}, \ldots, (A\sigma_N^2)^{-1})$. Working in this framework, the estimated vector $\vec{w}$ is equal to the most probable value of the above multivariate Gaussian model, that is, $\vec{w} \equiv \vec{\mu}$.

The critical advantage of a BCS approach, when compared with norm-based optimization methods in the processing of images, is that it fits better the true heavy-tailed statistics of the sparse vector and the noise component. In contrast, the norm-based optimization approaches consider only the noise level and not its actual distribution. For this purpose, a BCS algorithm for image data is expected to result in an increased sparsity, as our experimental results reveal. This performance could be enhanced by employing the proposed GSM-based BCS method, since it provides an additional degree of freedom through the scale parameter A, and thus, it ends up with a more accurate modeling of the true sparsity of the image of interest.

The problem of estimating the sparse vector $\vec{w}$ reduces to estimating the unknown model parameters $A, \mathbf{\Sigma}, \sigma_\eta^2$, by performing a *type-II* ML method. By noting that (7) can be re-written in the following equivalent form, $A^{-1}\,\mathbf{P} = \tilde{\mathbf{P}} = \left(A\sigma_\eta^{-2}\mathbf{\Phi}^T\mathbf{\Phi} + \mathbf{\Sigma}^{-1}\right)^{-1}$, we can estimate the unknown parameters σ_η^2, $\{\sigma_i^2\}_{i=1}^N$ iteratively by maximizing the following marginal log-likelihood function based on the matrix $\tilde{\mathbf{P}}$, with respect to the unknown parameters:

$$\begin{aligned}\mathcal{L}(\sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1}^N) &= \log[p(\vec{g}|A, \sigma_\eta^2, \{\sigma_i^{-2}\}_{i=1}^N)] \\ &= -\frac{1}{2}\left[M\,\log(2\pi) + \log(|\mathbf{C}|) + \vec{g}^T\mathbf{C}^{-1}\vec{g}\right] \ , \end{aligned} \tag{8}$$

where $\mathbf{C} = \frac{\sigma_\eta^2}{A}\mathbf{I} + \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^T$. By comparing (8) with the marginal likelihood model in [13] we notice that our proposed model is a

[1] In probability theory, a family of prior probability distributions $p(s)$ is said to be conjugate to a family of likelihood functions $p(x|s)$ if the resulting posterior distribution $p(s|x)$ is in the same family as $p(s)$.

scaled version of the previous hierarchical model by a factor of $1/A$. This factor plays an important role in the estimation process, since it controls the heavy-tailed behavior of the diagonal elements of $\mathbf{M}$ and consequently of the covariance matrix $\mathbf{P}$, and thus the sparsity of the estimated vector $\vec{w} \equiv \vec{\mu}$. An incremental algorithm is used for the addition and deletion of candidate basis functions (columns of $\mathbf{\Phi}$) to monotonically increase the marginal likelihood (8).

As it was mentioned before, one of the advantages of BCS is that it provides a measure of uncertainty in the estimation of the original image $\vec{f}$. This can be further employed to design a measurement matrix $\mathbf{\Phi}$ adapted to the information content of the sparse vector $\vec{w}$, by selecting the next projection $\vec{\phi}_{M+1}$ with the goal of reducing the uncertainty of $\vec{f}$. This is impossible with the previous norm-based approaches. Using the differential entropy minimization criterion, the next optimal projection is selected by performing an eigen-decomposition of $\mathbf{P}$ (cf. (7)) and selecting $\vec{\phi}_{M+1}$ to be the eigenvector corresponding to the largest eigenvalue.

3. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed sparse representation scheme (BCS-GSM) with the BCS method [8], as well as with the norm-based optimization schemes BP [9] and StOMP [14] (combined with a CFAR thresholding)[2], that achieved the best reconstruction performance on three 256×256 (noise-free) images of distinct content shown in Fig. 1, as well as on noisy versions of them by adding zero-mean Gaussian noise with $\mathrm{SNR} = 5, 10$ dB. Each image is transformed in the 2-D Discrete Wavelet Transform (DWT) domain [3], by decomposing it in 6 scales using the Daubechies' wavelet with 4 vanishing moments. The CS measurements $\vec{g}$ are acquired by applying partial Hadamard ensemble matrices $\mathbf{\Phi}$ on the wavelet coefficients vector $\vec{w}$.

Fig. 1. Test images.

In the following, we mainly focus on comparing the sparsity level achieved by each method, but also in conjunction with the corresponding reconstruction error. Fig. 2 shows the PSNRs between the reconstructed noise-free/noisy images and the original (noise-free) "Indor2" image, using the BCS-GSM, the linear reconstruction (inverse DWT) which achieves the best performance, and the selected norm-based CS methods for the two SNR values as a function of the number of measurements, where the number of CS measurements is equal to the sum of a portion α of the detail coefficients of the decomposition levels $3-6$, plus the coarse-scale approximation coefficients. The value of α varies in $\{0.4{:}0.2{:}1\}$. As it can be seen, the proposed BCS-GSM achieves similar reconstruction performance with the other four methods for the same number of measurements, with a decrease of PSNR which is at most ∼1dB in the noise-free case, while there is no difference in the two noisy cases.

[2]For the implementation of BCS, BP and StOMP we used the code included in the SparseLab package that is available online at `http://sparselab.stanford.edu/`.

Most importantly, Fig. 3 shows the CS ratio for each one of the four CS methods, which we define as the ratio of the number of measurements M over the number of non-zero components of $\vec{w}$ (sparsity) returned by each algorithm. The *higher the CS ratio the higher the sparsity* is for a fixed value of M. We can see that the proposed BCS-GSM method results in a much sparser representation of the original image as M increases, by reducing the number of significant basis functions by as much as 50%. Thus, Figs. 2-3 indicate that the proposed scheme could reduce significantly the storage requirements of a large image database, while maintaining a high reconstruction quality. This is very significant in applications such as image classification and retrieval, which is a subject of our ongoing research.

Finally, Fig. 4 shows the PSNR and CS ratio values for the BCS-GSM and BCS methods along with their adaptive versions applied on "Indor2" image with M varying in the interval $[2513, 2611]$. Specifically, both of the adaptive versions start with $M = 2513$ CS measurements and in each iteration M increases by one. Accordingly, the columns of $\mathbf{\Phi}$ are augmented with the eigenvector corresponding to the largest eigenvalue of the current matrix $\mathbf{P}$. The adaptive implementation achieves the same reconstruction performance in terms of the PSNR value for both methods, whereas the CS ratio increases, which is equivalent to a more efficient sparse representation (3.03% for the BCS-GSM method on average).

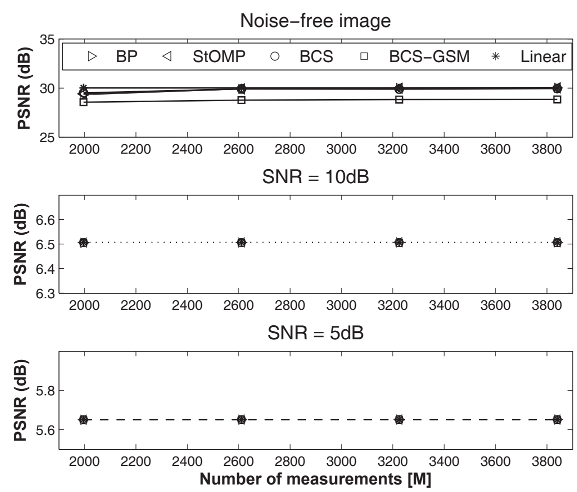

Fig. 2. PSNRs for "Indor2" image as a function of the number of CS measurements for SNR=5,10 dB.

Table 1 shows the PSNRs and CS ratios of the four CS methods for the other two test images, with $\alpha = 0.6$ resulting in $M = 2611$. As for the "Indor2" image, the proposed BCS-GSM approach achieves a significantly increased sparse representation for images with distinct contents (with a maximum increase of sparsity over 50%), while yielding the same reconstruction performance in the noise-free and the two noisy cases.

4. CONCLUSIONS AND FUTURE WORK

In this work, we described a method for CS reconstruction based on a Bayesian framework. We extended a recent work [8] by replacing the hierarchical prior model with a GSM, which models directly the weight vector with a heavy-tailed distribution that enforces

IMAGE		BP		StOMP		BCS		BCS-GSM		Linear
		PSNR	CS ratio	PSNR	CS ratio	PSNR	CS ratio	PSNR	CS ratio	PSNR
Indor4	Noise-free	29.27	0.68	29.16	1.85	29.04	3.63	28.35	5.63	29.60
	10 dB	7.08	0.68	7.09	1.87	7.09	2.05	7.08	5.49	7.09
	5 dB	6.15	0.67	6.15	2.45	6.15	4.31	6.14	5.95	6.15
Nemasup	Noise-free	25.43	0.67	25.34	2.11	25.35	2.24	25.32	4.13	25.61
	10 dB	9.95	0.68	9.96	2.92	9.96	1.90	9.95	4.30	9.96
	5 dB	5.57	0.68	5.56	2.51	5.57	2.01	5.56	4.15	5.57

Table 1. Performance comparison in terms of PSNR and CS ratio values for the reconstruction of "Indor4" and "Nemasup" images (noise-free+noisy versions) with $\alpha = 0.6$ ($M = 2611$).

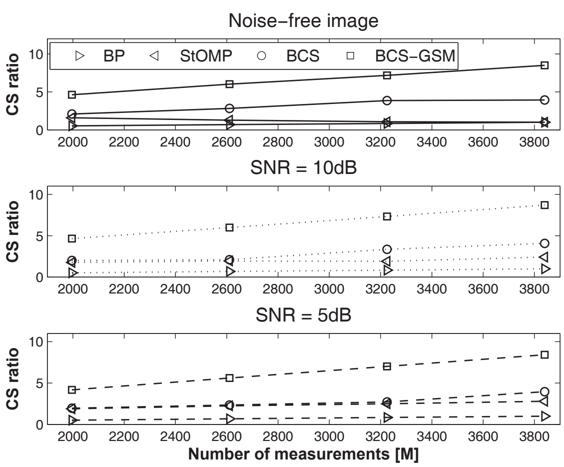

Fig. 3. CS ratios of the CS-based recovery methods for "Indor2" image as a function of the number of CS measurements for SNR=5,10 dB.

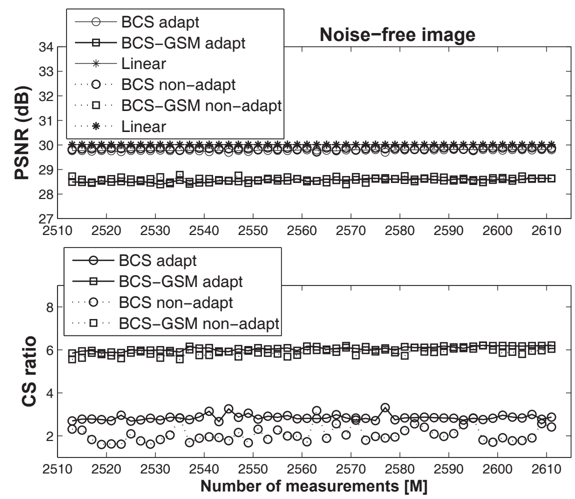

Fig. 4. PSNR and CS ratios of the BCS-GSM and BCS and their adaptive implementations for "Indor2" image.

its sparsity. The experimental results revealed a critical property of the proposed BCS-GSM approach when compared with norm-based CS reconstruction methods. In particular, we showed that the BCS-GSM implementation employs much fewer basis functions and thus, results in a more efficient sparse representation, while maintaining a comparable reconstruction performance.

In the present work, we did not make any assumption for the probability density function of the scaling factor A of the GSM. As a future work, we intend in posing a heavy-tailed distribution on A. In particular, when A follows an α-Stable distribution, then, the GSM is reduced to a sub-Gaussian model. We expect that the characteristic exponent which appears in the α-Stable distribution will provide further control on the sparsity of the weight vector.

5. REFERENCES

[1] D. Taubman and M. Marcellin, "JPEG 2000: Image Compression Fundamentals, Standards and Practice", (Int. Series in Eng. and Comp. Sci.), Norwell, MA: Kluwer, 2002.

[2] S. Grgic, K. Kers and M. Grgic, "Image compression using wavelets", *Proc. of IEEE Int. Symp. on Industr. Elec.*, Vol. 1, pp.99–104, 1999.

[3] S. Mallat, "A Wavelet Tour of Signal Processing", 2nd ed., New York: Academic Press, 1998.

[4] E. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. Inform. Th.*, Vol. 52, No. 2, pp. 489–509, Feb. 2006.

[5] D. Donoho, "Compressed Sensing", *IEEE Trans. Inform. Th.*, Vol. 52, No. 4, pp. 1289–1306, Apr. 2006.

[6] J. Haupt and R. Nowak, "Compressive sampling vs. conventional imaging", *Int. Conf. on Image Proc. (ICIP)*, Atlanta, Oct. 2006

[7] M. Duarte *et al.*, "Single-pixel imaging via compressive sampling", *IEEE Signal Proc. Mag.*, Vol. 25, No. 2, pp. 83–91, Mar. 2008.

[8] S. Ji, Y. Xue and L. Carin, "Bayesian Compressive Sensing", *IEEE Trans. on Signal Proc.*, Vol. 56, No. 6, pp. 2346–2356, June 2008.

[9] S. Chen, D. Donoho and M. Saunders, "Atomic decomposition by Basis Pursuit", *SIAM J. on Sci. Comp.*, Vol. 20, No. 1, pp. 33-61, 1999.

[10] J. Tropp and A. Gilbert, "Signal recovery from partial information via Orthogonal Matching Pursuit", Apr. 2005 [Online]. Available: `http://www.personal.umich.edu/~jtropp/papers/TG06-Signal-Recovery.pdf`

[11] M. Figueiredo, "Adaptive sparseness using Jeffreys prior", in *Advances in Neural Inf. Proc. Systems (NIPS 14)*, 2002.

[12] M. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine", *J. Mach. L. Res.*, Vol. 1, pp. 211-244, 2001.

[13] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models", *in Proc. 9th Int. Work. on Artif. Intell. and Stat.*, C. Bishop and B. Frey Eds., 2003.

[14] D. Donoho *et al.*, "Sparse solution of underdetermined linear equations by Stagewise Orthogonal Matching Pursuit", Tech. Report 06-02, Dept. of Statistics, Stanford Univ., 2006.