

# A GENERAL SPARSE IMAGE PRIOR COMBINATION IN COMPRESSED SENSING

Jorge Rubio<sup>a\*</sup>, Miguel Vega<sup>a</sup>, Rafael Molina<sup>a</sup> and Aggelos K. Katsaggelos<sup>b</sup>

a) Universidad de Granada, Granada 18071, SPAIN

Emails: georgos@correo.ugr.es, mvega@ugr.es, rms@decsai.ugr.es

b) Department of Electrical and Computer Engineering

Northwestern University, Evanston, IL 60208, USA

Email: aggk@eecs.northwestern.edu

## ABSTRACT

In this paper a general combination of sparse image priors is applied to Bayesian *Compressed Sensing* (CS) reconstruction of digital images. A simultaneous deblurring and CS reconstruction variational algorithm is derived. The application of the new algorithm, to both blurred and non-blurred images at different compression ratios, is studied. The new method is applied to *Passive Millimeter-Wave Imaging* (PMWI) CS, and its performance compared to state of the art CS reconstruction methods.

**Index Terms**— image processing, compressed sensing, millimeter wave imaging, Bayesian modeling, Bayesian inference

## 1. INTRODUCTION

It is well known that CS theory allows the recovery of sparse images from a limited number of incoherent observations (see [1]). The images are recovered by applying a CS method (see [2–4]), which typically enforces fidelity to the observations and sparsity of the unknown. In this paper we are interested in Bayesian CS (see [5, 6]). Bayesian CS enforces sparsity of the solutions through the application of sparse image priors. An image prior is considered to be sparse when it is super-Gaussian [7], i.e., compared to the Gaussian distribution, it has heavier tails, it is more peaked, and has a positive excess kurtosis (see [8]).

In this paper a general combination of sparse image priors is applied to Bayesian CS. This sparse prior combination, is based on the general formulation proposed in [8]) which provides flexibility in image modeling and algorithm design. A CS method, applicable to blurred images, will be derived, and its effectiveness both on blurred and unblurred images evaluated. The new method will be applied to PMWI.

One of the fields where CS has been successfully applied is PMWI. PMWI technology is based on the passive detection of naturally occurring millimeter-wave radiation from a scene [9]. PMWI is attracting increasing interest, because millimeter-waves are less affected by adverse conditions such as, clouds, fog, smoke, and dust, than visible or infrared light [10]. The limitations of current PMWI systems, in terms of the tradeoff between *Signal-to-Noise Ratio* (SNR) and acquisition time, and the expensiveness of radiometer devices, motivated the proposition of a novel single pixel PMWI system, based on the CS theory [11–13].

\*This work was supported in part by the Programa de Iniciación a la Investigación de la UGR, by the Spanish Ministry of Economy and Competitiveness under project TIN2010-15137 and by the European Regional Development Fund (FEDER).

The rest of this paper is organized as follows. Section 2 describes the problem formulation, and Section 3 its modeling within the Bayesian framework. The Bayesian inference and the proposed CS algorithm are described in Section 4. In Section 5 we present some experimental results. Finally, Section 6 concludes the paper.

## 2. PROBLEM FORMULATION

The compressed sensing process is usually formulated as follows,

$$\mathbf{y} = \Phi \mathbf{X} + \mathbf{n} \quad (1)$$

where  $\mathbf{y}$  is the  $M \times 1$  observations vector,  $\mathbf{X}$  the  $N \times 1$  vector representing the unknown image of size  $N = p \times q$ ,  $\Phi$  is the  $M \times N$  measurement matrix, and  $\mathbf{n}$  the  $M \times 1$  vector of observation noise, assumed to be white Gaussian with known variance  $\beta^{-1}$ .

The application of a CS method allows the recovery of the unknown  $\mathbf{X}$  image, even for a number of observations  $M \ll N$ , if the  $\mathbf{X}$  image is compressible in some basis, incoherent with the row vectors of the  $\Phi$  measurement matrix.

In this paper we assume that  $\mathbf{X}$  is a blurred version of the original  $p \times q$  image  $\mathbf{x}$ , which we want to recover,

$$\mathbf{X} = \mathbf{H} \mathbf{x}, \quad (2)$$

where  $\mathbf{H}$  is an  $N \times N$  convolution operator.

## 3. BAYESIAN MODELS

Given the degradation model of Eq. (1), the probability density for the observation  $\mathbf{y}$  given the original image  $\mathbf{x}$ , is

$$p(\mathbf{y}|\mathbf{x}) \propto \beta^{\frac{M}{2}} \exp \left\{ -\frac{\beta}{2} \|\mathbf{y} - \Phi \mathbf{H} \mathbf{x}\|^2 \right\}. \quad (3)$$

In this paper we utilize the following combination of general sparse priors

$$p(\mathbf{x}) = \prod_{j=1}^d \prod_{i=1}^N p(z_j(i)), \quad (4)$$

on the unknown filtered images set  $\{\mathbf{z}\} = \{\mathbf{z}_1, \dots, \mathbf{z}_d\}$ , where  $\mathbf{z}_j = \mathbf{F}_j \mathbf{x}$ , and  $\mathbf{F}_j$  are convolution operators. The four first order difference filters have been utilized here for the  $\mathbf{F}_j$ , but other configurations are also possible. In Eq. (4)  $z_j(i)$  denotes the  $i$ -th component of the  $\mathbf{z}_j$  vector. Notice that in Eq. (4) we are approximating the partition function as an independent product of partition functions.

The general sparse priors  $p(z_j(i))$  of Eq.(4) are defined as

$$p(z_j(i)) = \gamma \exp(-\alpha \rho(z_j(i))), \quad (5)$$

where  $\gamma$  is a normalization constant, that is  $\gamma^{-1} = \int \exp[-\alpha \rho(u)] du$ ,  $\rho(\cdot)$  is a penalty function symmetric around 0, and  $\alpha > 0$  is a

**Table 1:** Different choices for the penalty function.

Label	$\rho(\mathbf{s})$	$\rho'(\mathbf{s})/ \mathbf{s} $
$\ell_p$	$\frac{1}{p} \mathbf{s} ^p$	$ \mathbf{s} ^{p-2}$
log	$\log(\epsilon +  \mathbf{s} )$	$(\epsilon +  \mathbf{s} )^{-1} \mathbf{s} ^{-1}$
exp	$-\sigma_r \exp\left(-\frac{ \mathbf{s} ^2}{2\sigma_r}\right)$	$\exp\left(-\frac{ \mathbf{s} ^2}{2\sigma_r}\right)$

parameter regulating the intensity of the prior. Sparsity is achieved when the function  $\rho$  leads to the suppression of most coefficients  $z_j(i)$  while preserving a small number of important features. An important class of sparsity promoting priors are the so called Super Gaussian (SG) distributions.

Formally, for  $p(u) = \gamma \exp[-\alpha\rho(u)]$  to be SG, the function  $\rho(\sqrt{s})$  has to be increasing and concave for  $s \in (0, \infty)$  [7]. This condition is equivalent to  $\rho'(s)/s$  being decreasing on  $(0, \infty)$ . Table 1 shows some penalty functions, corresponding to SG distributions, proposed in [8], which will be used in this paper.

Being  $p(z_j(i))$  in Eq. (5) SG, it can be expressed as an *Scale Mixture of Gaussian* (SMG) (see [7]), i.e.

$$p(z_j(i)) = \int \mathcal{N}(z_j(i)|0, 1/\xi)p(\xi) d\xi. \quad (6)$$

Using this SMG representation, and introducing  $\{\boldsymbol{\eta}\} = \{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_d\}$ , with  $\boldsymbol{\eta}_j$  positive random variables, we transform the general sparse prior  $p(\mathbf{x})$  of Eq.(4) to the Gaussian form

$$p(\{\boldsymbol{\eta}\}, \mathbf{x}) = \prod_{j=1}^d \prod_{i=1}^N \mathcal{N}(z_j(i)|0, 1/\boldsymbol{\eta}_j(i))p(\boldsymbol{\eta}_j(i)). \quad (7)$$

Combining Eqs. (3), and (7) we obtain the following joint probability distribution

$$p(\Theta, \mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\{\boldsymbol{\eta}\}, \mathbf{x}), \quad (8)$$

where  $\Theta = \{\{\boldsymbol{\eta}\}, \mathbf{x}\}$  denote the set of all our unknowns. The estimation of the  $\alpha$  parameter of Eq. (5) is not studied in this paper and will be adjusted experimentally.

#### 4. VARIATIONAL BAYESIAN INFERENCE

The Bayesian inference will be based on the posterior distribution  $p(\Theta | \mathbf{y})$  which can be variationally approximated by the  $q(\Theta) = \prod_{\zeta \in \Theta} q(\zeta)$  distribution, with  $q(\zeta)$  given by

$$q(\zeta) \propto \exp\left(\langle \log [p(\Theta, \mathbf{y})] \rangle_{\Theta_\zeta}\right), \quad (9)$$

with  $\zeta \in \{\{\boldsymbol{\eta}\}, \mathbf{x}\}$ , and  $\Theta_\zeta$  denotes  $\Theta$  with  $\zeta$  removed, and  $E_{q(\Theta_\zeta)}[\cdot] = \langle \cdot \rangle_{\Theta_\zeta}$ . In the following, the subscript of the expected value will be removed when it is clear from the context.

From Eq.(9), we obtain for  $q(\mathbf{x})$

$$q(\mathbf{x}) \propto \exp\left\{\log(p(\mathbf{y}|\mathbf{x})) + E[\log(p(\{\boldsymbol{\eta}\}, \mathbf{x}))]_{\{\boldsymbol{\eta}\}}\right\}, \quad (10)$$

which is the multivariate Gaussian  $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}, \text{cov}_{q(\mathbf{x})})$  with

$$\text{cov}_{q(\mathbf{x})}^{-1} = \sum_{j=1}^d \mathbf{F}_j^t \text{diag}(\mathbf{w}_j) \mathbf{F}_j + \beta \mathbf{H}^t \boldsymbol{\Phi}^t \boldsymbol{\Phi} \mathbf{H}, \quad (11)$$

and

$$\hat{\mathbf{x}} = \beta \text{cov}_{q(\mathbf{x})} \mathbf{H}^t \boldsymbol{\Phi}^t \mathbf{y}. \quad (12)$$

In Eq.(11), the  $\mathbf{w}_j$  for  $j = 1, \dots, d$ , are  $N \times 1$  vectors with components  $w_j(i)$ , for  $i = 1, \dots, N$ ,

$$w_j(i) = \langle \boldsymbol{\eta}_j(i) \rangle_{\boldsymbol{\eta}_j(i)} = \int \boldsymbol{\eta}_j(i) q(\boldsymbol{\eta}_j(i)) d\boldsymbol{\eta}_j(i). \quad (13)$$

From Eq. (9), we obtain for the  $\{\boldsymbol{\eta}\}$  variables,

$$q(\boldsymbol{\eta}_j(i)) = p(\boldsymbol{\eta}_j(i)|z_j(i) = \nu_j(i)), \quad (14)$$

with

$$\nu_j(i) = \sqrt{\langle z_j^2(i) \rangle_{q(\mathbf{x})}}. \quad (15)$$

and  $\langle z_j^2(i) \rangle_{q(\mathbf{x})}$  given by

$$\langle z_j^2(i) \rangle_{q(\mathbf{x})} = \langle z_j(i) \rangle_{q(\mathbf{x})}^2 + \text{tr}(\text{cov}_{q(\mathbf{x})} \mathbf{F}_j^t \mathbf{J}^{ii} \mathbf{F}_j), \quad (16)$$

where  $\mathbf{J}^{ii}$  is an  $N \times N$  matrix with all its elements equal to 0, except the  $i$ -th entry of its diagonal, which is equal to 1. As  $\text{cov}_{q(\mathbf{x})}$  can not be exactly obtained, its Jacobi approximation has been used in this paper.

Let us finally proceed to calculate  $w_j(i)$  in Eq. (13). Notice that the whole distribution  $q(\boldsymbol{\eta}_j(i))$  is not required and only its mean is utilized in the iterative process. Using Eq. (6), for  $p(z_j(i))$ ,

$$\begin{aligned} \frac{\partial p(s)}{\partial s} \Big|_{s=\nu_j(i)} &= -\nu_j(i) \int \boldsymbol{\eta}_j(i) \mathcal{N}(\boldsymbol{\eta}_j(i)|0, 1/\boldsymbol{\eta}_j(i)) p(\boldsymbol{\eta}_j(i)) d\boldsymbol{\eta}_j(i) \\ &= -\nu_j(i) p(\nu_j(i)) \int \boldsymbol{\eta}_j(i) p(\boldsymbol{\eta}_j(i)|\nu_j(i)) d\boldsymbol{\eta}_j(i) \\ &= -\nu_j(i) p(\nu_j(i)) w_j(i). \end{aligned} \quad (17)$$

Furthermore using Eq. (5) we have  $\frac{\partial p(s)}{\partial s} = -\alpha \rho'(s) p(s)$ . Then utilizing Eq (17) we obtain

$$w_j(i) = \alpha \frac{\rho'(\nu_j(i))}{|\nu_j(i)|}. \quad (18)$$

The values of  $\rho'(s)/|s|$  for the penalty functions considered in this paper can be found in the right column of table 1.

The proposed algorithm is summarized below in Algorithm 1.

**Algorithm 1** Compressed sensing using a General Sparse Image Prior Combination

**Require:** Values for  $\alpha$  and  $\beta$ , and an initial value  $\mathbf{x}^{(0)}$  for the unknown image  $\mathbf{x}$ .

Set  $k = 1$  and  $\text{cov}_{q(\mathbf{x})}^{(0)} = \mathbf{0}$ .

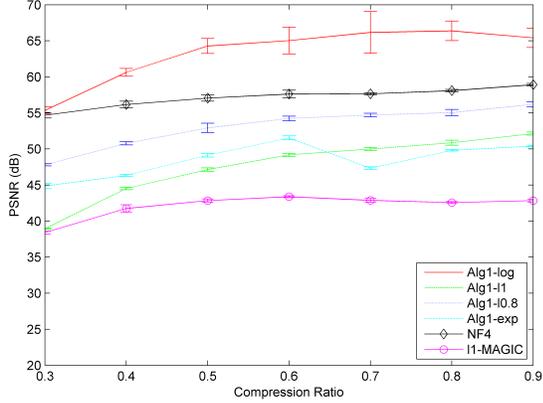
**while** convergence criterion is not met **do**

1. Compute  $\mathbf{w}_j^{(k)}$  using Eqs (15) & (18).
2. Calculate  $\left(\text{cov}_{q(\mathbf{x})}^{(k)}\right)^{-1}$  using Eq. (11).
3. Estimate the  $\hat{\mathbf{x}}^{(k)}$  image by solving Eq.(12).
4. Set  $k = k + 1$ .

For this iterative Algorithm 1, the initial value  $\mathbf{x}^{(0)} = \frac{1}{c} \mathbf{H}^t \boldsymbol{\Phi}^t \mathbf{y}$ , has been utilized, where  $c$  is the square root of the mean value of the diagonal elements of the  $\mathbf{H}^t \boldsymbol{\Phi}^t \boldsymbol{\Phi} \mathbf{H}$  matrix. In step 3 of Algorithm 1, Eq.(12) has been solved by applying a *Conjugate Gradient* (CG) algorithm. The used convergence criterion for Algorithm 1 has been  $\frac{\|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}\|^2}{\|\mathbf{x}^{(n-1)}\|^2} < 10^{-5}$  and  $n > 3$ , or  $n > 30$ .

## 5. EXPERIMENTAL RESULTS

A number of experiments have been performed using the proposed Algorithm 1, with the penalty functions shown in table 1, which will be referred henceforth as *Alg1- $\ell_{0.8}$*  for the  $\ell_p$  penalty function with  $p = 0.8$ , *Alg1- $\ell_1$*  for the  $\ell_p$  penalty function with  $p = 1$ , *Alg1-log* for the log penalty function, and *Alg1-exp* for the exp



**Fig. 1:** PSNR of the phantom image reconstructions, obtained using the different methods, as a function of the CS compression ratio.

penalty function, respectively. The results obtained have been compared with the obtained using the *NF4* combination of non-stationary edge-preserving priors in [14, 15], and also with the CS method  $\ell_1$ -MAGIC in [2], using a TV prior. In this paper the *Peak Signal to Noise Ratio* (PSNR) comparison metric has been adopted.

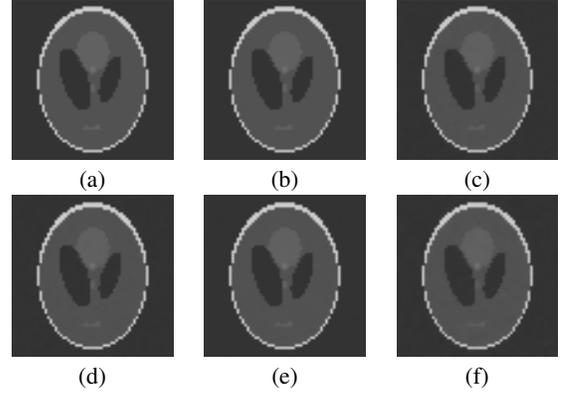
In the first experiment, the CS process described by Eq.(1) has been applied to the synthetic  $64 \times 64$  phantom image shown in figure 2(a). In this experiment the elements of the  $\Phi$  measurement matrix have been randomly drawn from a Gaussian process of mean 0 and variance  $1/N$ .  $\Phi$  matrices of different  $M \times N$  sizes, corresponding to compression ratios in the range (0.3, 0.9) have been generated, and applied to the  $\mathbf{X}$  image. In this experiment the images are not blurred, i.e.  $\mathbf{H} = \mathbf{I}$  in Eq.(2), and  $\mathbf{X} = \mathbf{x}$ . White Gaussian noise of 40 dB has been added to the obtained measures. Nine noise realizations per experiment have been generated, and the mean PSNR values, with their standard deviations, of the restorations obtained using the different methods represented in figure 1.

The reconstructions obtained in this first experiment of CS in the absence of blurring, using the considered methods, which are shown in figures 2(b-f), are visually indistinguishable from the original image in figure 2(a). The best results are the obtained using Alg1-log and NF4.

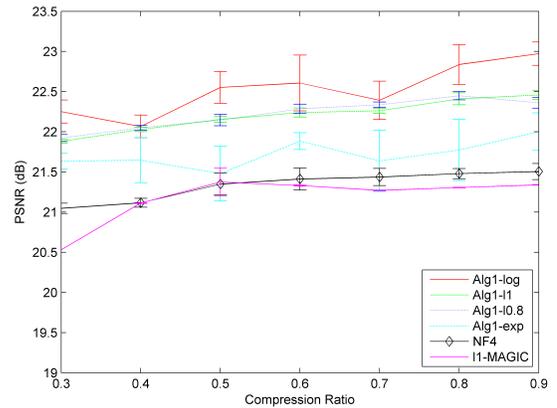
In a second experiment the same procedure described above has been applied to the blurred image shown in figure 4(a), resulting from the application of a Gaussian blur of variance 3.5 and  $9 \times 9$  support, to the phantom image of figure 2(a). The different methods have been applied to obtain reconstructions of the original unblurred image and the obtained PSNR values, as a function of the compression ratio, depicted in figure 3. It can be observed in figure 3 that the results of the application of the proposed method, specially Alg1-log, are better than the obtained using NF4 and  $\ell_1$ -MAGIC. Nevertheless the results shown in figure 3 for the blurred image, are worse in PSNR terms than the ones shown in figure 1 for the unblurred image.

Figures 4(b-f) show the reconstructions obtained using the different methods for a compression ratio of 0.5. It can be observed in those reconstructions that the thickness of the different regions, specially the skull region, are more similar to that of the blurred image of figure 4(a), than to the original image of figure 2(a). Thus, it can be asserted that the deblurring effectiveness of these CS methods has been somewhat limited.

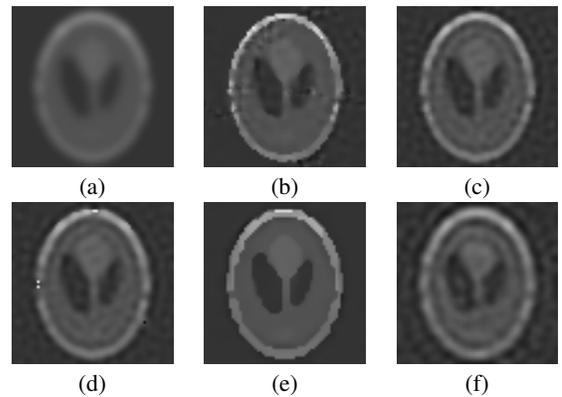
One of the reasons for the use in this section of the phantom im-



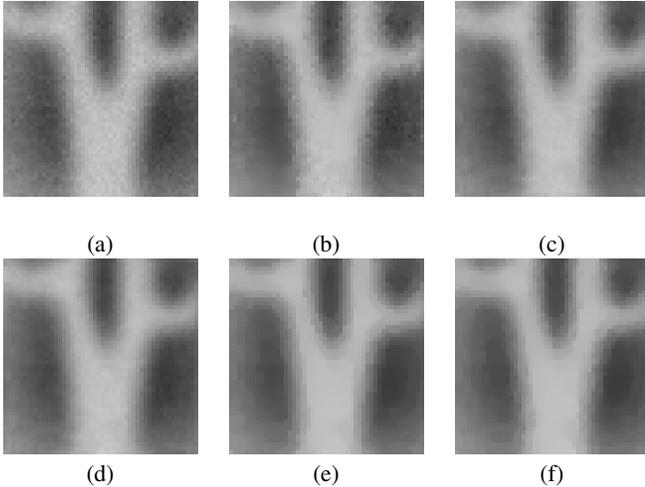
**Fig. 2:** (a)  $64 \times 64$  phantom image. Reconstructions, for a 0.5 compression ratio, obtained using: (b) Alg1-log (64 dB), (c) Alg1- $\ell_1$  (47.1 dB), (d) Alg1-exp (49.1 dB), (e) NF4 (57.1 dB) and (f)  $\ell_1$ -MAGIC (42.8 dB).



**Fig. 3:** PSNR of the phantom image reconstructions, in the presence of blurring, obtained using the different methods, as a function of the CS compression ratio.



**Fig. 4:** (a) blurred phantom image. Reconstructions, for a 0.5 compression ratio, obtained using: (b) Alg1-log (22.5 dB), (c) Alg1- $\ell_1$  (22.1 dB), (d) Alg1-exp (21.5 dB), (e) NF4 (21.3 dB) and (f)  $\ell_1$ -MAGIC (21.3 dB).



**Fig. 5:** (a) Scissors  $50 \times 50$  PMWI image from [12]. Reconstructions, for a 0.4 compression ratio, obtained using: (b) Alg1-log (33.8 dB), (c) Alg1- $\ell_1$  (35.3 dB), (d) Alg1-exp (35.2 dB), (e) NF4 (34.6 dB) and (f)  $\ell_1$ -MAGIC (34.7 dB).

age, with its flat regions, is its similarity to the typical PMW images, like the one shown in figure 5(a). In the third experiment measures from this image have been obtained using an orthogonal  $\Phi$  measurement matrix constructed from row vectors of the discrete 2D Fourier basis, to which white Gaussian noise of 40 dB has been added. Figures 5(b-f) show the reconstructions obtained using the different methods for a compression ratio of 0.4. In this case the best reconstructions in terms of PSNR have been obtained using Alg1- $\ell_1$  and Alg1-exp, although the reconstructions obtained using  $\ell_1$ -MAGIC, NF4 and Alg1-log are also good.

In the fourth experiment, we finally study a more textured image, like the  $128 \times 128$  crop of the Lena image in figure 6(a). Measures have been obtained using a Fourier basis  $\Phi$  matrix, to which white Gaussian noise of 40 dB has been added. Figures 6(b-d) show the reconstructions obtained using the different methods for a compression ratio of 0.4. The best reconstructions in terms of PSNR are the ones provided by Alg1-log and NF4. All these reconstructions present a good visual quality, though some fine textures, for example in the ribbon of the hat, have been lost.

## 6. CONCLUSIONS

In this paper the application of a general combination of sparse image priors to Bayesian CS has been studied, and a simultaneous deblurring and CS reconstruction variational algorithm has been derived. The new algorithm performs better than state of the art CS methods, both on flat by regions images, like the typical in PMWI, and on more textured images.

## 7. REFERENCES

- [1] Emmanuel J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [2] Justin Romberg Emmanuel J. Candès, "l1magic," 2006.
- [3] Chengbo Li, *An Efficient Algorithm For Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing*, Ph.D. thesis, Rice University, Houston, Texas, Sept. 2009.
- [4] Stephen Becker, Jérôme Bobin, and Emmanuel J. Candès, "NESTA: a fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sciences*, pp. 1–39, 2011.
- [5] Shihao Ji, Ya Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, June.
- [6] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *Image Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [7] Jason A. Palmer, Kenneth Kreutz-Delgado, and Scott Makeig, "Strong sub- and super-gaussianity," in *LVA/ICA*, 2010, pp. 303–310.
- [8] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Blind deconvolution with general sparse image priors," in *European Conference on Computer Vision (ECCV)*, 2012.
- [9] L. Yujiri, M. Shoucri, and P. Moffa, "Passive millimeter wave imaging," *Microwave Magazine, IEEE*, vol. 4, no. 3, pp. 39–50, 2003.
- [10] N. Kukutsu S. Oka, H. Togo and T. Nagatsuma, "Latest trends in millimeter-wave imaging technology," *Progress In Electromagnetics Research Letters*, vol. I, pp. 197–204, 2008.
- [11] S. D. Babacan, M. Luessi, L. Spinoulas, A. K. Katsaggelos, N. Gopalsami, T. Elmer, R. Ahern, S. Liao, and A. Raptis, "Compressive passive millimeter-wave imaging," in *Image Processing (ICIP), 18th IEEE International Conference on*, Sept. 2011, pp. 2705–2708.
- [12] N. Gopalsami, S. Liao, T. Elmer, A. Heifetz, and A. C. Raptis, "Compressive sampling in active and passive millimeter-wave imaging," in *Proc. 36th Int Infrared, Millimeter and Terahertz Waves (IRMMW-THz) Conf*, 2011, pp. 1–2.
- [13] Leonidas Spinoulas, Jin Qi, Aggelos K. Katsaggelos, Thomas W. Elmer, Nachappa Gopalsami, and Apostolos C. Raptis, "Optimized compressive sampling for passive millimeter-wave imaging," *Appl. Opt.*, vol. 51, no. 26, pp. 63356342, Sept. 2012.
- [14] Esteban Vera, Miguel Vega, Rafael Molina, and Aggelos K. Katsaggelos, "A novel iterative image restoration algorithm using nonstationary image priors," in *Image Processing (ICIP), 18th IEEE International Conference on*, Sept. 2011, pp. 3457–3460.
- [15] L. Spinoulas, B. Amizic, M. Vega, R. Molina, and A.K. Katsaggelos, "Simultaneous bayesian compressive sensing and blind deconvolution," in *European Signal Processing Conference*. Bucharest (Romania), August 2012, pp. 1414–1418.



**Fig. 6:** (a)  $128 \times 128$  crop of the Lena image. Reconstructions, for a 0.4 compression ratio, obtained using: (b) Alg1-log (37.8 dB), (c) NF4 (37.6 dB) and (d)  $\ell_1$ -MAGIC (36.9 dB).