Compressive imaging using approximate message passing and a Cauchy prior in the wavelet domain

Paul Hill, Jong-Hoon Kim, Adrian Basarab, Denis Kouamé, David Bull, Alin Marian Achim

COMPRESSIVE IMAGING USING APPROXIMATE MESSAGE PASSING AND A CAUCHY PRIOR IN THE WAVELET DOMAIN

P.R. Hill[⋆]*, J-H. Kim*[⋆]*, A. Basarab*[†]*, D. Kouamé*[†]*, D.R. Bull*[⋆] *and A. Achim*[⋆]

[⋆]Department of Electrical and Electronic Engineering,
The University of Bristol, Bristol,BS8 1UB, UK e-mail: (paul.hill @bristol.ac.uk)
[†]IRIT, University of Toulouse, CNRS, INPT, UPS, UT1C, UT2J, France

ABSTRACT

Approximate Message Passing (AMP) is an iterative reconstruction algorithm that performs signal denoising within a compressive sensing framework. We propose the use of heavy tailed distribution based image denoising, specifically using a Cauchy prior based Maximum A-Posteriori (MAP) estimate within a wavelet based AMP compressive sensing structure. The use of this MAP denoising algorithm provides extremely fast convergence for image based compressive sensing. The proposed method converges approximately twice as fast as the compared AMP methods whilst providing superior final MSE results over a range of measurement rates.

1. INTRODUCTION

Compressed sensing is a recent breakthrough in signal processing whereby a sparse signal can be effectively captured within the compressed domain by making a reduced number of measurements compared to the Nyquist limit [1, 2]. Conventional compression algorithms fully sample a signal and transform this signal into a more sparse domain before compression. Conversely, compressive sensing aims to directly sample within such a domain therefore reducing the number of input measurements and the need to store and process a fully sampled signal before compression.

The wavelet transform of a natural image is typically sparse. It thus provides a suitable space within which to perform compressive sampling. Sampling the image in the transform eliminates the need for full sampling and transformation, significantly reducing computational and memory requirements. The reconstruction of the image in such a framework is conventionally achieved using Linear Programming (LP) [3] but has been recently more efficiently achieved using AMP based methods [3–5]. A key element of image based AMP reconstruction is the use of denoising methods. A summary of the most effective image denoising methods (within the structure of AMP) is given by Metzler [6].

1.1. Contributions

This work extends the two dimensional wavelet AMP system proposed by Som and Schniter [5] and Tan et al. [4] through the integration of a novel denoising algorithm. This denoising method utilises a heavy tailed Cauchy prior within a MAP denoising structure leading to faster convergence and improved quantitative results. This is enabled through the solution of the analytic differential equations within the MAP and AMP frameworks (equations (17) and (19)). Additionally, an improved AMP structure is proposed using more effective wavelet transforms (**sym4** filters replace **Haar** filters as previously used in [4]) in conjunction with analytic solutions to the derivatives of the previously implemented Soft Thresholding (ST) and Amplitude-Scale-Invariant Bayes Estimator (ABE) methods (summarised in table 1).

A review of related work and the general background of AMP and compressive sensing / sampling is first presented in section 2. Subsequently, section 3 describes the Cauchy-MAP image denoising method and how it is applied to an AMP framework. Results and a discussion are presented in section 4 followed by a conclusion in section 5.

The authors would like to thank and acknowledge the EPRSC for funding under the Vision for the Future Platform Grant: EP/M000885/1

2. APPROXIMATE MESSAGE PASSING

The Approximate Message Passing (AMP) reconstruction algorithm defined by Donoho et al. [3] is a matrix based iterative reconstruction algorithm for single dimensional signals inspired by belief propagation techniques common in graphical networks.

AMP has been extended by Tan et al. [4] and Som and Schniter [5] for image reconstruction within a compressive image sampling framework. These authors have used various image denoising algorithms to implement the AMP iterative structure. The D-AMP method has been recently developed in order to integrate non-thresholding image denoising algorithms [6]. However, these denoising algorithms (such as BM3D [7]) do not have an analytical expression, the AMP structure is approximated.

2.1. Compressive Sensing Framework for Images

Assuming an image with dimensions $W \times H$, the image is first formed into a column vector of length N (= $W \times H$). Assuming a noiseless compressive sensing model, the observed measurements (comprising a vector $\mathbf{y}$ of length M: $\mathbf{y} \in \mathbb{R}^M$) are defined as the multiplication of the measurement matrix ($\mathbf{\Phi} \in \mathbb{R}^{M \times N}$) with the image column vector $\mathbf{x}$:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}. \tag{1}$$

The measurement vector $\mathbf{y}$ is observed and the original image signal $\mathbf{x}$ is reconstructed using one of a choice of reconstruction algorithms. Approximate Message Passing is one such reconstruction algorithm.

2.2. Algorithmic Framework

A surprising result of compressive sensing is that, although the measurement matrix $\mathbf{\Phi}$ can defined in many different forms, it it is often

most effective with random entries. Specifically, within the scheme of Donoho et al. [3] the matrix entries of $\mathbf{\Phi}$ are independent and i.i.d. $\mathcal{N}(0, 1/M)$ distributed.

Given an initial guess of $\mathbf{x_0} = 0$, the first order AMP algorithm iterates to convergence using the following alternating expressions [3]:

$$\mathbf{x}^{t+1} = \eta_t\left(\mathbf{\Phi}^T\mathbf{z^t} + \mathbf{x^t}\right), \quad (2)$$

$$\mathbf{z^t} = \mathbf{y} - \mathbf{\Phi x^t} + \frac{1}{\delta}\mathbf{z^{t-1}}\langle\eta'_{t-1}\left(\mathbf{\Phi}^T\mathbf{z^{t-1}} + \mathbf{x^{t-1}}\right)\rangle, \quad (3)$$

where the under-sampling fraction (measurement rate) is $\delta = M/N$ and $\mathbf{\Phi}^T$ is transpose the measurement matrix $\mathbf{\Phi}$. Furthermore, the functions $\eta(\cdot)$ and $\eta'(\cdot)$ are the threshold function and its first derivative respectively (known as the Onsager term [4]).

2.3. Wavelet Based AMP

A discrete wavelet transform provides a sparse representation domain for natural images. In order to make the compressive sensing framework effective for such images, the compressive sensing AMP framework encapsulated by (2) and (3) can be integrated with the wavelet transform. The signal thresholding function $\eta(\cdot)$ transforms into a wavelet based image denoising function [4]. Denoting the wavelet transform and its inverse as $\mathcal{W}$ and $\mathcal{W}^{-1}$, the wavelet coefficients $\theta_{\mathbf{x}}$ are related to the (vectorised) image $\mathbf{x}$ as $\theta_{\mathbf{x}} = \mathcal{W}\mathbf{x}$ and $\mathbf{x} = \mathcal{W}^{-1}\theta_{\mathbf{x}}$.

In this scenario the iterative forms (2) and (3) become:

$$\theta_{\mathbf{x}}^{t+1} = \eta_t\left(\left(\mathbf{\Phi}\mathcal{W}^{-1}\right)^T\mathbf{z^t} + \theta_{\mathbf{x}}^t\right), \quad (4)$$

$$\mathbf{z^t} = \mathbf{y} - \left(\mathbf{\Phi}\mathcal{W}^{-1}\right)\theta_{\mathbf{x}}^t + \frac{1}{\delta}\mathbf{z^{t-1}}\langle\eta'_{t-1}\left(\left(\mathbf{\Phi}\mathcal{W}^{-1}\right)^T\mathbf{z^{t-1}} + \theta_{\mathbf{x}}^{t-1}\right)\rangle,$$

$$\mathbf{z^t} = \mathbf{y} - \mathbf{\Phi x}^t + \frac{1}{\delta}\mathbf{z^{t-1}}\langle\eta'_{t-1}\left(\left(\mathbf{\Phi}\mathcal{W}^{-1}\right)^T\mathbf{z^{t-1}} + \theta_{\mathbf{x}}^{t-1}\right)\rangle. \quad (5)$$

As we have chosen an orthonormal wavelet transform $\mathcal{W}\mathcal{W}^T = \mathcal{W}^T\mathcal{W} = I$, the input into the denoising function $\eta(\cdot)$ (4) simplifies to

$$\left(\mathbf{\Phi}\mathcal{W}^{-1}\right)^T\mathbf{z^t} + \theta_{\mathbf{x}}^t = \mathcal{W}\Phi^T\mathbf{z}^t + \mathcal{W}\mathbf{x}^t = \mathcal{W}\mathbf{q}^t = \theta_{\mathbf{q}}^t, \quad (6)$$

where $\mathbf{q^t}$ is the noisy measured image vector defined as:

$$\mathbf{q^t} = \mathbf{\Phi}^T\mathbf{z}^t + \mathbf{x}^t. \quad (7)$$

3. MAP BASED IMAGE DENOISING USING A HEAVY TAILED CAUCHY PRIOR FOR AMP

It has been identified that wavelet coefficients are accurately modelled by heavy tailed representations such as α-stable distributions [8]. However, for the general case, α-stable distributions have no closed form expressions for analytic manipulation and therefore have no effective use within subsequently defined MAP denoising algorithms [9, 10]. However, the Cauchy distribution (a special case of the α-stable family) not only has a heavy tailed form but has a compact analytical PDF expression of the form:

$$P(w) = \frac{\gamma}{(w^2+\gamma^2)}, \quad (8)$$

where w is the value of a wavelet coefficient and γ is the parameter controlling the spread of the distribution known as the dispersion.

Firstly, the original image is assumed to be contaminated with additive Gaussian noise. This leads to an equivalent additive noise model for the wavelet coefficients in the transform domain [10]:

$$v = w + n, \quad (9)$$

where v is the measured wavelet coefficient, w is the clean coefficient and n is the Gaussian distributed error signal (n and w are assumed to be statistically independent).

In order to denoise the observed coefficients, a MAP estimator of the clean coefficient value w given the observed coefficient value v can be defined as

$$\hat{w}(v) = \arg\max_w P_{w|v}(w|v). \quad (10)$$

An expression for the a-posteriori distribution $P_{w|v}(w|v)$ can be given using the Bayes' theorem:

$$P_{w|v}(w|v) = \frac{P_{v|w}(v|w)P_w(w)}{P_v(v)}. \quad (11)$$

Assuming a normal noise distribution ($P_n(v-w) = P_{v|w}(v|w)$), disregarding the constant evidence term $P_v(v)$ and taking the natural logarithm, (11) can be expressed as:

$$\hat{w}(v) = \arg\max_w\left[-\frac{(v-w)^2}{2\sigma^2} + \log\left(P_w(w)\right)\right],$$

$$= \arg\max_w\left[-\frac{(v-w)^2}{2\sigma^2} + \log\left(\frac{\gamma}{(w^2+\gamma^2)}\right)\right]. \quad (12)$$

The solution to (12) can be found by taking the first derivative of the argmax argument (with respect to w) and setting to zero:

$$\frac{v-w}{\sigma^2} - \frac{2w}{w^2+\gamma^2} = 0. \quad (13)$$

This leads to the cubic:

$$\hat{w}^3 - v\hat{w}^2 + (\gamma^2 + 2\sigma^2)\hat{w} - \gamma^2 v = 0. \quad (14)$$

The solution to (14) is found (using Cardano's method) to be:

$$\hat{w} = \eta(v) = \frac{v}{3} + s + t, \quad (15)$$

where s and t are defined as:

$$s = \sqrt[3]{\frac{q}{2} + dd}, \quad t = \sqrt[3]{\frac{q}{2} - dd}, \quad (16)$$

$$dd = \sqrt{p^3/27 + q^2/4},$$

$$p = \gamma^2 + 2\sigma^2 - v^2/3,$$

$$q = v\gamma^2 + 2v^3/27 - (\gamma^2 + 2\sigma^2)v/3.$$

Using some simple manipulations, the first derivative of the denoising function defined in (15) is given by:

$$\hat{w}' = \eta'(v) = 1/3 + s' + t'. \quad (17)$$

Where s' and t' are found as follows:

$$s' = \frac{q'/2+dd'}{3(q/2+dd)^{(2/3)}}, \quad t' = \frac{q'/2-dd'}{3(q/2-dd)^{(2/3)}},$$
$$dd' = \frac{p'p^2/9+q'q/2}{2dd}, \tag{18}$$
$$p' = -2v/3,$$
$$q' = -2\sigma^2/3 + 2\gamma^2/3 + 2v^2/9.$$

The noise variance σ^2 is estimated as the variance of the elements of vector $\mathbf{z}$ defined in (4) and (5) [4]. The dispersion parameter γ is estimated as given in Achim [11]. This denoising algorithm is utilised within (4) through the denoising of the individual vector elements of $\theta^t_{\mathbf{q}}$ denoted as $\theta^t_{\mathbf{q},i}$ at position i in the vector and $\theta^{t+1}_{\mathbf{x},i}$ similarly defined (see (7)). The subsequent denoising algorithms are described given the measured wavelet coefficient $v = \theta^t_{\mathbf{q},i}$ and the clean wavelet coefficient estimate for the next iterations as $\hat{w} = \theta^{t+1}_{\mathbf{x},i}$ i.e. v is denoised to form $\hat{w}$ for all elements in the vectors.

3.1. Comparative Methods

Two comparative methods were implemented to benchmark the performance of the proposed system: Soft Thresholding (ST) and the Amplitude-Scale-Invariant Bayes Estimator (ABE).

3.1.1. Soft Wavelet Thresholding: ST

Soft wavelet Thresholding (ST) is a simple thresholding function proposed by Donoho and Johnstone [12] that takes a threshold $\mathcal{T}$ and modifies each wavelet coefficient according to

$$\hat{w} = \eta(v) = \text{sign}(v)(|v| - \mathcal{T}) \cdot \mathbb{1}_{(|v|>\mathcal{T})} \tag{19}$$

where $\mathbb{1}_{(\cdot)}$ denotes the indicator function. The threshold $\mathcal{T}$ is calculated as the Mth largest magnitude value of $\theta^{\mathbf{t}}_{\mathbf{q}}$ at each iteration [3].

3.1.2. Amplitude-Scale-Invariant Bayes Estimator: ABE

An additional denoising function for wavelet coefficient shrinkage is the so called Amplitude-Scale-Invariant Bayes Estimator proposed by Figueiredo and Nowak [13]. In the framework of the AMP reconstruction algorithm, the new estimate $\theta^{t+1}_{\mathbf{x}}$ based on the previous estimate $\theta^t_{\mathbf{x}}$ is given by:

$$\hat{w} = \eta\left(v\right) = \frac{\left(v^2 - 3\sigma^2\right)_+}{v}, \tag{20}$$

where σ^2 is the noise variance of the present iteration t and $(\cdot)_+$ is a function such that $(u)_+ = 0$ if $u \leq 0$ and $(u)_+ = u$ if $u > 0$.

Table 1 summarises the denoising and Onsager terms of the ABE, ST and Cauchy-MAP methods ($\eta\left(\cdot\right)$ and $\eta'\left(\cdot\right)$).

4. RESULTS

Figure 1 shows the normalised Mean Squared Error (NMSE) results for the proposed method and the two comparative methods (ST and ABE). This figure shows the averaged results across all 591 test images. The dataset of 591 images is obtained from "pixel-wise labelled image database v2" at http://research.microsoft.com/en-us/projects/objectclassrecognition. This dataset was also used for AMP reconstruction experiments by Tan et al. [4] and Som and Schniter [5] and similarly used a top left patch (in this case of 128×128 pixels).

Figure 2 visualises the results of the three considered denoising algorithms within the wavelet based AMP framework. This figure shows an original image together with the first four reconstruction iterations for the three denoising methods. This figure reflects the extremely fast convergence of the proposed method (also reflected in the NMSE results shown in figure 1).

Table 2 shows quantitative final results after 30 iterations (averaged over all 591 images). This table highlights the improvement in NMSE associated with the proposed method compared to ST and ABE. Also, although the speed of the proposed method is slightly slower than the ST and ABE methods this is significantly offset by the faster convergence of the Cauchy-MAP based approach.

Although state of the art denoising algorithms have been implemented within an AMP framework (such as BM3D [7] within D-AMP [6]), the first derivative Onsager term used in these cases is only an approximation (rather than an exact analytical function as developed in our work). Additionally, it is shown within the D-AMP structure that wavelet thresholding (such as the Cauchy-MAP) is two orders of magnitude faster converging than the state of the art methods such as BM3D. Furthermore, our proposed method is specifically adapted to heavy tailed distributions and therefore is be applicable to signals having this characteristic (such as ultrasound images etc.). This will be investigated in future work.

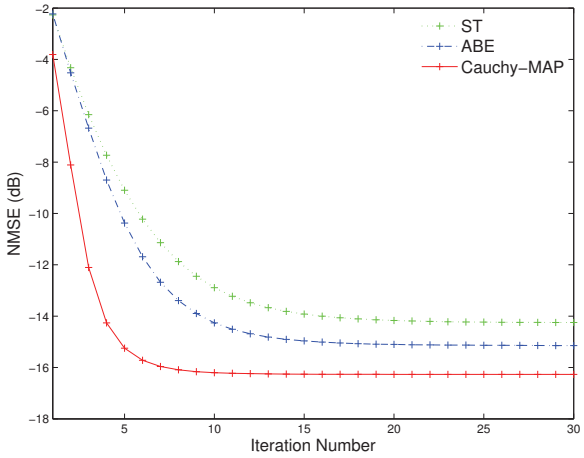

Fig. 1. Normalised Mean Squared Error (NMSE) vs iteration number. Results averaged over 591 input image patches (top left 128×128 from each image). Measurement ratio (δ) is 0.1831. The measurements were noiseless.

4.1. Discussion

The input for each reconstruction experiment was the top left hand 128×128 patch from each of the images in the dataset. The input signal $\mathbf{x}$ was a vectorised version of this image patch of length 16384.

The wavelet transform $\mathcal{W}$ was implemented to 7 levels using the **sym4** filters with critical sampling at each stage and symmetric extension (to preserve transform orthogonality and ensure $\mathcal{W}^{-1} = \mathcal{W}^T$, essential for efficient computation of the iterations given in (4),(5) and (6)). The **sym4** filters were chosen as they are approximately symmetric and have a filter length which gives the best result

Table 1. Summary of $\eta(\cdot)$ and $\eta'(\cdot)$ functions for use in AMP reconstruction where v is the ith element of $\theta^t_{\mathbf{q}}$ to be denoised i.e. $v = \theta^t_{\mathbf{q},i}$

Algorithm	$\eta(\cdot)$	$\eta'(\cdot)$
Cauchy-MAP	$v/3 + s + t$ (15)	$1/3 + s' + t'$ (17)
ABE	$\frac{(v^2-3\sigma^2)_+}{v}$ (20)	$\mathbb{1}_{\left(v^2>3\sigma^2\right)} \cdot \left(1 + 3\left(\frac{\sigma}{v}\right)^2\right)$
ST	$\text{sign}(v)(\lvert v\rvert - \mathcal{T}) \cdot \mathbb{1}_{(\lvert v\rvert>\mathcal{T})}$ (19)	$\mathbb{1}_{(\lvert v\rvert>\mathcal{T})}$

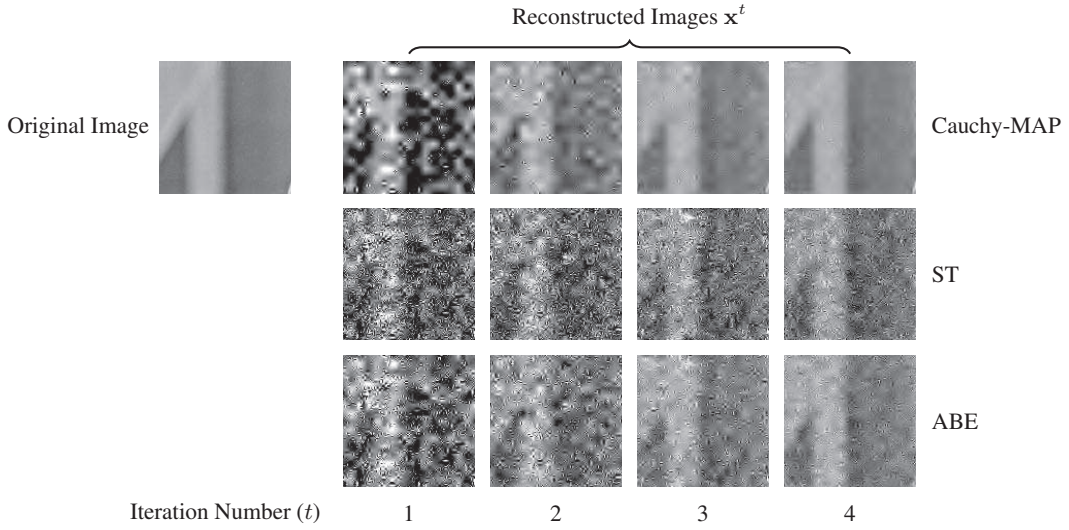

Fig. 2. Iterations of ABE, ATE and Cauchy Denoising Methods within AMP reconstruction framework

Table 2. Runtime and NMSE results averaged over 591 input image patches (top left 128×128 from each image) the measurement ratio (δ) is 0.1831. The measurements were noiseless.

Algorithm	**NMSE(dB)**	**Runtime (secs)**
Cauchy-MAP	-16.27	5.33
ABE	-15.15	4.77
ST	-14.25	4.57

(in terms of NMSE). They also offer a good compromise between compact support and frequency localisation. This wavelet also gave considerably better qualitative (less blocking artefacts) and quantitative results (in terms of NMSE) compared to the Haar wavelets implemented by Tan et al. [4] and Som and Schniter [5]. The measurement ratio ($\delta = M/N$) was 0.1831 for all the experiments. This gave comparable NMSE results to previous work. The algorithms were implemented and run on a Macbook Pro (2015) with a 2.7GHz Intel Core i5 processor with 8GB of RAM within a Matlab R2014a environment. The image patch size was small due to the large space requirements of the $M \times N$ and $N \times N$ matrices ($\mathbf{\Phi}$ and $\mathcal{W}$) respectively. This was in common with previous work [4, 5] and future work will address scalability issues from a theoretical and implementation standpoint.

Future investigations will also focus on characterising the performance of the Cauchy based denoising method over a greater range of measurement ratios and extending this to a better performing bivariate method [8] comparable to the AMP-Wiener method implemented by Tan et al. [4].

5. CONCLUSION

The Approximate Message Passing (AMP) method is an efficient and effective compressive sensing signal reconstruction algorithm. We have used AMP in a two dimensional image based framework integrated with an orthogonal wavelet transform. The key transformation within the AMP framework is a denoising function. A novel application of a heavy tailed Cauchy prior MAP denoising application is introduced. This denoising function and its derivative are used with the AMP framework in order to achieve extremely fast convergence compared to previously implemented methods (over twice as fast for a typical image).

6. REFERENCES

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

[2] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] D. L. Donoho, A. Maleki and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 914 – 918, 2009.

[4] J. Tan, Y. Ma and D Baron, "Compressive imaging via approximate message passing with image denoising," *Signal Processing, IEEE Transactions on*, vol. 63, no. 8, pp. 2085–2092, 2015.

[5] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3439–3448, 2012.

[6] C.A. Metzler, A. Maleki and R.G. Baraniuk, "From denoising to compressed sensing," *arXiv preprint arXiv:1406.4175*, 2014.

[7] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative ltering." *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, August 2007.

[8] A. Achim, P. Tsakalides and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773 – 1784, August 2003.

[9] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications. New York: Wiley*. Wiley, New York, 1995.

[10] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1601–1611, June 1998.

[11] A. Achim, D. Herranz and E.E. Kuruoğlu, "Astrophysical image denoising using bivariate isotropic cauchy distributions in the undecimated wavelet domain," vol. 2, pp. 1225–1228, 2004.

[12] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, August 1994.

[13] M. A. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: an empirical bayes approach using jeffrey's noninformative prior," *Image Processing, IEEE Transactions on*, vol. 10, no. 9, pp. 1322–1331, 2001.