IMAGE DECONVOLUTION USING A GAUSSIAN SCALE MIXTURES MODEL TO APPROXIMATE THE WAVELET SPARSENESS CONSTRAINT

Yingsong Zhang, Nick Kingsbury

Signal Processing & Communication Group, Dept. of Engineering, University of Cambridge
Cambridge CB2 1PZ, UK

ABSTRACT

This paper proposes to use an extended Gaussian Scale Mixtures (GSM) model instead of the conventional ℓ_1 norm to approximate the sparseness constraint in the wavelet domain. We combine this new constraint with subband-dependent minimization to formulate an iterative algorithm on two shift-invariant wavelet transforms, the Shannon wavelet transform and dual-tree complex wavelet transform (DTCWT). This extented GSM model introduces spatially varying information into the deconvolution process and thus enables the algorithm to achieve better results with fewer iterations in our experiments.

Index Terms— Image restoration, Wavelet transforms

1. INTRODUCTION

In this paper we consider improved methods for image deconvolution, using the Gaussian Scale Mixture (GSM) model to approximate the sparseness constrants with shift-invariant wavelet frames.

Normally, the image measurement process can be represented by a known stationary linear filter followed by the addition of white noise of zero mean and variance σ^2:

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

where $\mathbf{x}$ and $\mathbf{y}$ are vectors containing the original image and the observed image, respectively; $\mathbf{n}$ represents independent white noise. These vectors have $N = N_1 \times N_2 \times \ldots \times N_K$ components, where N_k stands for the number of samples along dimension k. $\mathbf{H}$ is a square, block-circulant matrix that approximates the convolution (blurring function).

Traditional frequently used approaches to deconvolution are based on the least squares (LS) formulation using the differences between the measured and deconvolved images as the cost function during optimization. It is well known that the LS formulation usually leads to an ill-posed problem that needs extra regularization. Under appropriate assumptions, the regularized deconvolution problem can be derived using a Bayesian framework and controlled by different prior assumptions. Hence we can get the MAP estimate of $\mathbf{x}$ by minimising a cost function

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 - \sigma^2 \log(p(\mathbf{x})) \tag{2}$$

where $p(\mathbf{x})$ is the regularising prior.

Wavelet-based deconvolution is a relatively recent technique in the area of deconvolution. It minimises the cost function w.r.t the wavelet coefficients $\mathbf{w}$:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{H}\mathbf{M}\mathbf{w} - \mathbf{y}\|^2 - \sigma^2 \log(p(\mathbf{w})) \tag{3}$$

Thanks to CSC scholarship for funding my PhD study.

where $\mathbf{M}$ is the inverse wavelet transform and $-\log(p(\mathbf{w}))$ is called the wavelet domain penalty function.

State-of-art wavelet-based deconvolution algorithms mostly exploit the sparseness penalty [1, 2, 3], because the wavelet coefficients tend to give an economical representation of the image [4, 3], especially when there are no other obvious options for alternative penalties. Direct measurement of sparseness is the ℓ_0 norm which counts the number of the nonzero elements. However, the ℓ_0 norm results in an NP-complex minimisation problem, because the only known method to search for the exact solution is combinatorial [5]. Hence, the ℓ_1 norm is used most often in practice. Indeed, the minimisation of the ℓ_1 norm promotes sparsity, which has been known for many years, and put to use in a wide range of applications [6]. Donoho [5] showed that the minimal ℓ_1-norm near-solution is the sparsest near-solution for most large underdetermined systems of equations.

However, the ℓ_1 norm is not spatially adaptive. It is often connected with marginal statistics of wavelet coefficients. It may well depict the wavelet coefficients in a subband, but it does not describe the correlations within a small patch. Better results can be expected if a prior that can depict spatially varying statistics is used.

Extending the work of Vonesch & Unser [1], we formulate the wavelet-regularized subband-dependent minimization problem with a GSM prior instead of the traditional ℓ_1 norm and compare two shift-invariant wavelets, the Shannon wavelet and dual-tree complex wavelet transform (DTCWT), both of which provide shift invariance [1, 7].

We follow the same notation conventions as in [1, 7]:

Name	Type	Representation
$\mathbf{W}$	matrix	forward wavelet transform
$\mathbf{W}_j$	matrix	forward wavelet transform in a given subband j with all other subbands set to zero
$\mathbf{M}$	matrix	inverse wavelet transform
$\mathbf{M}_j$	matrix	inverse wavelet transform in the given subband j with all other subbands set to zero
$\mathbf{P}_j$	matrix	$\mathbf{P}_j = \mathbf{M}_j\mathbf{W}_j$, the transfer function of subband j
$\mathbf{w}$	vector	wavelet coefficients of $\mathbf{x}$, $\mathbf{w} = \mathbf{W}\mathbf{x}$
$\mathbf{w}_j$	vector	wavelet coefficients of $\mathbf{x}$ in subband j, $\mathbf{w}_j = \mathbf{W}_j\mathbf{x}$
$\mathbf{A}_j$	matrix	$(\mathbf{A}_j)_{ii} = \begin{cases} A_{ii}, & \text{if } w_i \text{ is in subband } j \\ 0, & \text{otherwise} \end{cases}$

2. THE ℓ_{02} PENALTY

Besides ℓ_1 norms, other spareseness approximations can be used. Here we use the penalty

$$-\log(p(\mathbf{w})) = \sum_i \frac{|w_i|^2}{\mathrm{E}(|w_i|^2) + \epsilon^2} \approx \|\mathbf{w}\|_0 \tag{4}$$

where $\mathrm{E}(|w_i|^2)$ is the expectation of $|w_i|^2$, and ϵ^2 is a small positive constant. When $\mathrm{E}(|w_i|^2)$ is zero, the penalty of w_i will be so large that w_i will be forced to be near zero. In a Bayesian view, this penalty corresponds to a Gaussian pdf $p(\mathbf{w})$ that is proportional to $\exp\{-\frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w}\}$ where $\mathbf{A}$ is a diagonal matrix with A_{ii}^{-1} being ϵ^2 plus the expected variance of the zero-mean w_i, such that

$$\mathbf{w} \sim \mathcal{N}(0, \mathbf{A}^{-1}) \tag{5}$$

For convenience, we call this penalty the ℓ_{02} penalty.

2.1. Relation with the GSM model

The ℓ_{02} penalty can be considered as a simple extension to the GSM model. The GSM model describes the non-stationary behaviour of the wavelet coefficients of natural signal [8, 9], using:

$$w_i = \sqrt{z_i} r_i \tag{6}$$

It assumes that each coefficient w_i is specified by a stationary zero mean Gaussian probability density function r_i and a spatially fluctuating variance z_i. i is a position index.

It has been shown that for varying z_i this model can successfully simulate the high kurtosis and longer tails of the marginal distributions in addition to the bow-tie shape of the marginal histograms of wavelet coefficients of natural signals [10, 9].

The ℓ_{02} penalty can be obtained by simply setting $z_i = A_{ii}^{-1}$, where A_{ii} varies in different positions.

2.2. Discussion on the ℓ_{02} penalty

When comparing with the ℓ_1 norm, the ℓ_{02} penalty has the following advantages:

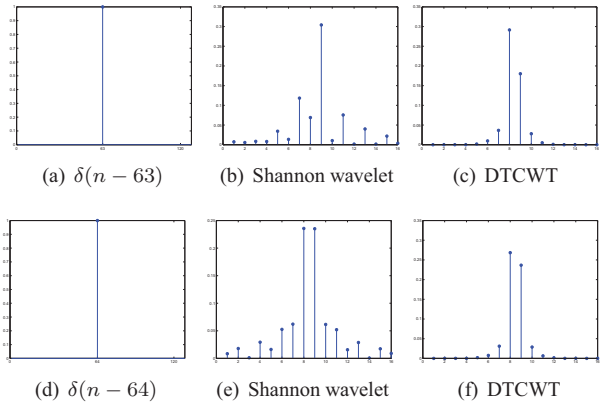
- It is closer to the ℓ_0 norm than the ℓ_1 norm.
- It does not need to adjust the regularisation parameter λ. When using ℓ_1 norms, the prior pdf is actually $\exp(\lambda\|\mathbf{x}\|_1)$, where λ is a statistical parameter of the prior distribution of wavelet coefficients. On the contrary, the penalty we propose here does not have such an explicit parameter.
- It is smooth and differentiable.
- It corresponds to an adaptive GSM model. $\mathrm{E}(|w_i|^2)$ is localised and spatially adaptable.

Successful application of the ℓ_{02} penalty relies on the estimation of $\mathrm{E}(|w_i|^2)$. $\mathrm{E}(|w_i|^2)$ represents the expectation of the image structure [4], which should be shift independent, i.e. a small shift in the signal should not result in dramatically change in $\mathrm{E}(|w_i|^2)$. When applied with a non-redundant wavelet basis, which is well-known to suffer from the oscillation problem of shifting [11](Figure 1), the estimation of $\mathrm{E}(|w_i|^2)$ needs to be handled very carefully.

3. ALGORITHM

We combine the ℓ_{02} penalty with the Shannon wavelet based subband emphasis factor α_j derived by Vonesch & Unser [1], which results in the following cost function for each subband j:

$$\begin{aligned}\tilde{J}_j(\mathbf{x}) =& \alpha_j \|\mathbf{P}_j \left[\mathbf{x}^{(n)} + \alpha_j^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})\right] - \mathbf{P}_j\mathbf{x}\|^2 \\ &+ \sigma^2(\mathbf{W}_j\mathbf{x})^T\mathbf{A}(\mathbf{W}_j\mathbf{x}) + C_j(\mathbf{x}^{(n)}, \mathbf{y})\end{aligned} \tag{7}$$



(a) $\delta(n-63)$ (b) Shannon wavelet (c) DTCWT

(d) $\delta(n-64)$ (e) Shannon wavelet (f) DTCWT

Fig. 1. The distribution of $|w_i|$ in the high-pass subband on level 3. When the signal has a shift, the distribution of the DTCWT wavelet coefficients magnitudes maintains a relatively stable pattern around the singularity compared with that of the Shannon wavelet, which is a shift-invariant non-redundant Discrete Wavelet Transform (DWT). This is because the DTCWT makes use of the complex modulus operation, which helps to reduce the variance of $|w_i|^2$.

where $\mathbf{x}^{(n)}$ is the current estimate of $\mathbf{x}$, $C_j(\mathbf{x}^{(n)}, \mathbf{y}) = \alpha\|\mathbf{P}_j\mathbf{x}^{(n)}\|^2 + \|\mathbf{P}_j\mathbf{y}\|^2 - \|\mathbf{P}_j\mathbf{H}\mathbf{x}^{(n)}\|^2$ does not dependent on $\mathbf{x}$, and $\alpha_j = \max_{\|\mathbf{v}\|^2=1, \mathbf{v}\in\mathbf{W}_j}\|\mathbf{H}\mathbf{v}\|^2$. It results in the following update rule:

- The Landweber iteration:

$$\mathbf{z}_j^{(n)} = \alpha_j\mathbf{w}_j^{(n)} + \mathbf{W}_j\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$$

- The denoising operation, which projects the new coefficients back into the space dominated by the prior $\mathbf{A}$ by soft-thresholding $\mathbf{z}_j^{(n)}$:

$$\mathbf{w}_j^{(n+1)} = \left(\alpha_j\mathbf{I} + \sigma^2\mathbf{A}_j\right)^{-1}\mathbf{z}_j^{(n)} \tag{8}$$

Rewriting this in element-wise form, we have

$$w_i^{(n+1)} = \frac{\mathrm{E}(|w_i|^2) + \epsilon^2}{\alpha_j(\mathrm{E}(|w_i|^2) + \epsilon^2) + \sigma^2} z_i^{(n)} \tag{9}$$

where $w_i^{(n+1)}$ and $z_i^{(n)}$ are in the subband j. Then the spatial-domain estimate is gained by $\mathbf{x}^{(n+1)} = \mathbf{M}\mathbf{w}^{(n+1)}$.

The subband emphasis factor α_j derived by Vonesch & Unser [1] was modified to work with the DTCWT instead of the Shannon basis [7].

3.1. Estimation and update of the prior A

$\mathbf{A}$ represents the expected variance of the wavelet coefficients; however, the exact information of $\mathrm{E}(|w_i|^2)$ is usually unavailable. Thus a coarse estimate of $\mathbf{A}$ from the blurred image has to be used instead: the wavelet variance estimates are given by the energy of the wavelet coefficients of an estimate of the original image; and the diagonal entries A_{ii} are the reciprocals of the estimated variances plus ϵ^2.

Since the estimated image is often contaminated by artifacts and noise, this straightforward approach fails to obtain a satisfactory estimate of $\mathbf{A}$. An intuitive approach to overcome such difficulties is to denoise the coefficients before we calculate $\mathbf{A}$, which will give

us a smoother prior. Because $\mathbf{w}$ is sparse, we may estimate $\hat{\mathbf{w}}$ by bivariate wavelet shrinkage [13] of $\mathbf{w}$, and we use $1/(|\hat{w}_i|^2 + \epsilon^2)$ for each term A_{ii} in $\mathbf{A}$.

Meanwhile, we also adopt a heuristic complementary approach, which is to update the estimate of $\mathbf{A}$ every several iterations, because the restored image structure gets closer to the true one. This updating approach can be viewed as being similar to the iteratively re-weighted least squares (IRLS) algorithm [12]. The penalty function $\|\mathbf{w}\|_0$ is replaced by a sequence of least squares problems:

$$-\log(p_n(\mathbf{w})) = \sum_i \frac{1}{|\hat{w}_i^{(n)}|^2 + \epsilon^2} |w_i|^2. \tag{10}$$

4. EXPERIMENTAL RESULTS

We have space for only one example here. Results are shown in Figure 2 and 3.

4.1. Experiment settings

For comparative purposes, we performed the experiments on the standard test image, Cameraman. We convolved the image with a 9×9 uniform blurring kernel, and added white Gaussian noise to the results in order to replicate the experimental setup of Vonesch and Unser [1] and others. We use the Blurred Signal-to-Noise Ratio (BSNR) to define the noise level over N pels: $\mathrm{BSNR} = 10\log_{10}\frac{\|\mathbf{Hx}\|^2 - N(\overline{\mathbf{Hx}})^2}{N\sigma^2}$. In our experiments, the BSNR was set to 40dB. We also adopted Improvement in Signal-to-Noise Ratio (ISNR, same as SERG in [1]) to evaluate each estimate $\mathbf{x}^{(n)}$: $\mathrm{ISNR}(\mathbf{x}^{(n)}) = 10\log_{10}(\frac{\|\mathbf{y}-\mathbf{x}\|^2}{\|\mathbf{x}^{(n)}-\mathbf{x}\|^2})$. For each test case, we used the same initial estimate as [1], which was obtained using the under-regularised Wiener-type filter: $\mathbf{x}^{(0)} = (\mathbf{H}^T\mathbf{H} + 10^{-3}\sigma^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}$. We averaged the ISNR results over 30 noise realisations, see Figure 2. The numerical results of the fast thresholded Landweber (FTL) algorithm are quoted from Vonesch and Unser's paper [1].

4.2. Implementation details

As described above, before the iterations start, it is necessary to estimate $\mathbf{A}$. In our experiments, we choose to use the bivariate shrinkage denoising rule [13] to calculate $\hat{w}_i$, and hence $A_{ii}^{-1} = \mathrm{E}(|\hat{w}_i|^2) + \epsilon^2$. This is because this shrinkage algorithm exploits the interscale dependencies between wavelet coefficients while being relatively simple. In order to keep the influence of different denoising schemes to a minimum, the denoising is implemented on the DTCWT coefficients of $\mathbf{x}^{(n)}$ and then we inverse transform the denoised coefficients back to be $\hat{\mathbf{x}}^{(n)}$ for estimating $\mathbf{A}$.

For estimation of $\mathrm{E}(|w_i|^2)$ when applying the algorithm based on the Shannon wavelet, we use the following method to reduce the uncertainty in estimating $\mathrm{E}(|w_i|^2)$ introduced by the oscillation problem of the Shannon wavelet:

1. apply undecimated Shannon wavelet transform on the denoised $\hat{\mathbf{x}}^{(k)}$ to obtain undecimated coefficients $\bar{\mathbf{w}}$;
2. partition the undecimated coefficients $\bar{\mathbf{w}}$ into regions g_i corresponding to decimated Shannon wavelet coefficients w_i;
3. take the mean of $|\bar{w}_j|^2$ in region g_i as $\mathrm{E}(|w_i|^2)$

For convenience, we call it the undec-mean estimate. The undec-mean estimates are shown in the middle column of Figure 3. We compare the undec-mean estimate with the direct estimate (Figure 3, the left-hand column), which is given by $\mathrm{E}(|w_i|^2) \approx |w_i^{(k)}|^2$. The undec-mean estimate of $\mathrm{E}(|w_i|^2)$ preserves more image features than the direct estimate (Figure 3, the first row). However, the method causes the energy of singularities near region boarders to disperse over a small neighbourhood. This energy dispersion problem puts limits on the performance of this algorithm.

We update the estimate of $\mathbf{A}$ after every 5 iterations in our experiments. The results show that updating with the direct estimate causes the results to deteriorate instead of continuing improve (Figure 2 (a)), while updating with the undec-mean estimate improves the convergence speed and overall performance in terms of ISNR (Figure 2 (b)).

When applying with the DTCWT, we use the direct estimate of $\mathrm{E}(|w_i|^2)$, because the DTCWT has a relatively stable pattern around the singularity (Figure 1) due to its improved subsampling lattice, and we also want to avoid the energy dispersion problem. We update the estimate of $\mathbf{A}$ after every 5 iterations. We show the estimates of $\mathrm{E}(|w_i|^2)$ in the right column of Figure 3.

In order to show the evolution of the estimate of $\mathrm{E}(|w_i|^2)$ in each subband and compare the difference of the estimates, we scale all the subband images in the same level by a common factor. Hence there are 4 scale factors corresponding to each level respectively in Figure 3. The coarest-scale images are scaled such that their maxima equal 1.

5. DISCUSSION AND CONCLUSION

With the Gaussian scale mixtures model, the modified FTL algorithm produces a better result with fewer iterations than the original FTL algorithm [1], primarily due to the spatially varying knowledge introduced by the Gaussian scale mixture prior. However, good results are only achieved when the energy oscillation problem of Shannon wavelets is counteracted by complementary approaches, e.g. the undec-mean estimate used in our experiments, or the random shift approach [14]. We also observed that the updating of $\mathbf{A}$ only improved the result by a very small amount after the 15th iteration in Figure 2(b).

When using the DTCWT as the wavelet frame instead of the Shannon wavelet bases, we found that this algorithms outperformed the modified FTL algorithm by nearly 0.7 dB. The direct estimation method we used in the estimation of $\mathbf{A}$ keeps the estimate as local as possible. This direct estimate is applicable with the DTCWT primarily due to the complex modulus operation, which gives near shift-invariance to the estimate.

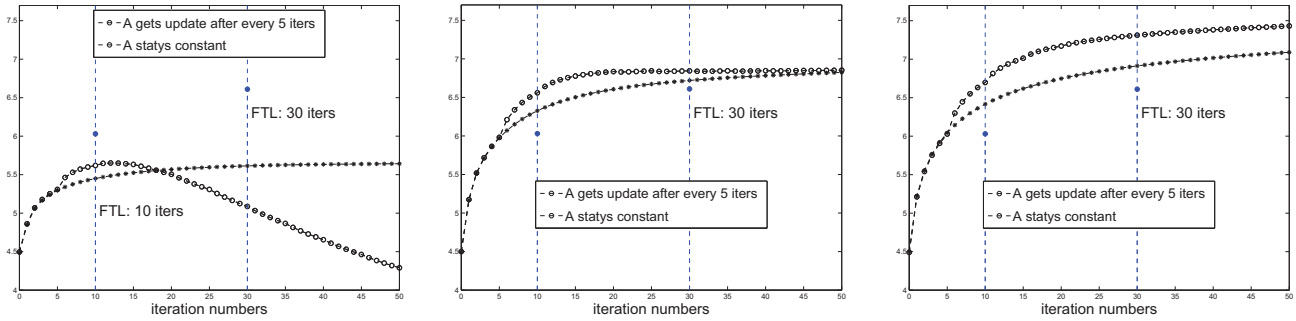
6. REFERENCES

[1] Cédric Vonesch and Michael Unser, "A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539–549, 2008.

[2] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[3] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," *Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. Proceedings., 1999 IEEE International Conference on*, vol. 6, 1999.

[4] P. de Rivaz and N.G. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," *Image Process-*



(a) Shannon wavelets with direct estimates of $\mathbf{A}$. (b) Shannon wavelets with undec-mean estimate of $\mathbf{A}$. (c) DTCWT with direct estimates of $\mathbf{A}$.

Fig. 2. ISNR as a function of the iteration number when applying our iteration rule. Blue-filled circles are the markers for the results of the FTL algorithm [1]. The initial estimate of $\mathbf{A}$ was obtained from the under-regularised Wiener filtered result to start the iteration. $-o-$ marks the ISNR curve when $\mathbf{A}$ gets updates every five iterations, while $-*-$ marks the ISNR curve when $\mathbf{A}$ gets no updates after the iteration begins. Note here that the result of Shannon wavelet (b) used the undec-mean estimate of $\mathbf{E}(|w_i|^2)$ to counteract the energy oscillation problem and thus had a better result than (a).

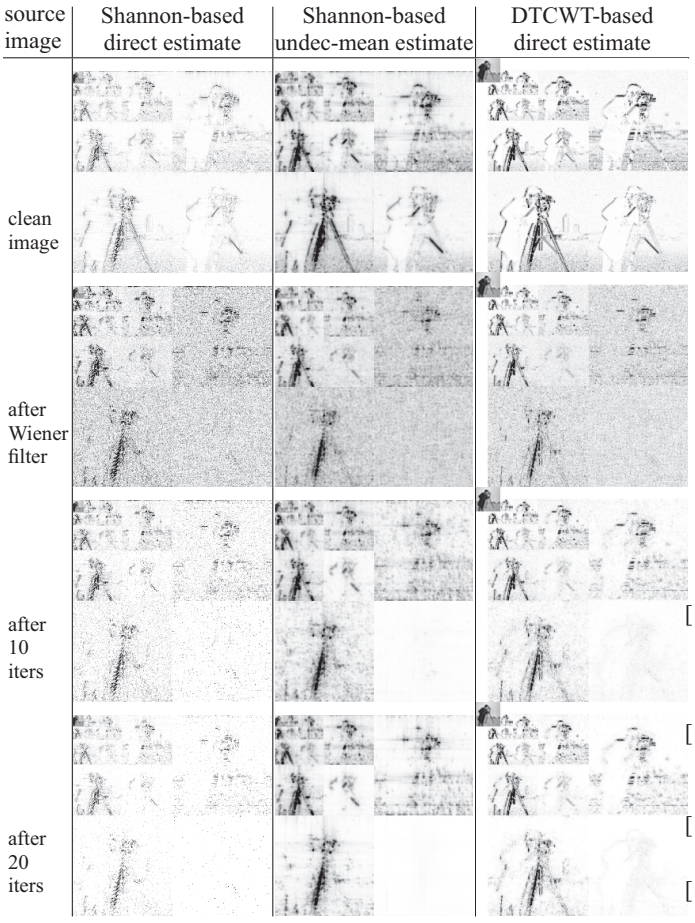


Fig. 3. Estimates of $\sqrt{\mathrm{E}(|w_i|^2)}$ of blurred Cameraman with SNR=40dB. The estimates evolve as the iteration goes. Undec-mean estimates improve the estimation of $\mathrm{E}(|w_i|^2)$, however, they also cause energy dispersion over the neighbourhood. All the subband images in the same level are scaled by a common factor.

ing, 2001. Proceedings. 2001 International Conference on, vol. 2, 2001.

[5] D.L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution," *Comm. Pure Appl. Math*, vol. 59, no. 7, pp. 907–934, 2006.

[6] I. Daubechies, M. Fornasier, and I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints," *J. Fourier Anal. Appl., to appear*, 2007.

[7] Yingsong Zhang and N.G. Kingsbury, "A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis," *Proceedings of the International Conference on the Engineering in Medicine and Biology*, 2008.

[8] M.A. Miller and N.G. Kingsbury, "Image denoising using derotated complex wavelet coefficients," *IEEE Transactions on Image Processing*, Sept.

[9] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *Image Processing, IEEE Transactions on*, vol. 12, no. 11, pp. 1338–1351, 2003.

[10] L. Parra, C. Spence, and P. Sajda, "Higher-order statistical properties arising from the non-stationarity of natural signals," *Advances in Neural Information Processing Systems*, vol. 13, pp. 786–792, 2001.

[11] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury, "The dual-tree complex wavelet transform," *Signal Processing Magazine, IEEE*, vol. 22, no. 6, pp. 123–151, 2005.

[12] M.R. Osborne and M. Robert, *Finite algorithms in optimization and data analysis*, Wiley, 1985.

[13] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 50, no. 11, pp. 2744–2756, 2002.

[14] R.R. Coifman and D.L. Donoho, "Translation-invariant denoising," *Wavelets and Statistics*, vol. 103, pp. 125–150, 1995.