Wavelet based noise reduction of four-dimensional data with application to MRI

Jun Yu and Nils Östlund

Research Report

Centre of Biostochastics

Wavelet based noise reduction of four-dimensional data with application to MRI

Jun Yu[1]

Centre of Biostochastics
Swedish University of Agricultural Sciences, Umeå, Sweden

Nils Östlund

Department of Biomedical Engineering and Informatics
University Hospital, Umeå, 901 81 Sweden

Abstract

This paper is devoted to the development of noise reduction methods for spatial-temporal signals with applications in magnetic resonance imaging. A noise reduction algorithm for 4D MRI signals, based on the wavelet transform and Gaussian scale mixtures, is proposed here. Simulation study shows that the new method is capable to improve the performance of noise reduction in higher dimensions.

Keywords: Noise reduction, wavelet shrinkage, 4-D MRI data, Gaussian scale mixtures.

---

[1] E-mail address to the correspondence author: jun.yu@slu.se

1 Introduction

Wavelet methods have become a widely spread tools in signal and image processing tasks [18]. Wavelet based noise reduction methods are often called wavelet shrinkage methods and was introduced by Donoho and Johnstone [7], Donoho et al. [8], and Bruce and Gao [3]. The main idea of a shrinkage method is to transform the data by using a type of wavelet transform, remove noise from the wavelet coefficients by shrinking them, and then reconstruct a denoised signal or image by applying the inverse wavelet transform. Normal images can be sparsely represented in the wavelet domain with a few large coefficients, while the noise (often assumed white) is represented with many small coefficients. By shrinking the small coefficients more than the larger coefficients, the method is able to reduce noise.

How to apply the shrinkage procedure has gained much interest since the technique was introduced. Many different shrinkage functions, adjusted to different types of noise, have been developed. Improvements of the method is now primarily achieved by using redundant transform domains (such as undecimated wavelets, ridgelets, curvelets, see e.g. [5, 12, 18, 30]) and/or by exploiting the inter-coefficient dependencies of the wavelet transform (e.g. [22, 23]).

Wavelet denoising is primarily used for one and two-dimensional data. When the data has more than two dimensions the noise reduction is often performed on two dimensional slices of the data. Since many of the state of the art methods use undecimated wavelet transforms and/or the inter-coefficient dependencies, more information can be obtained by performing the noise reduction on the whole data set (three or higher dimensional) simultaneously. This could be done at the expense of more things to estimate and higher demands on the computer power.

There are special demands in noise reduction of MRI. While normal images are judged by their visual appearance, introducing artifacts and smoothing are not accepted in MRI images. Artifacts introduced by a noise reduction algorithm could be misinterpreted as clinical findings. A review of wavelet denoising in MRI was provided by Pižurica et al. [21].

The main objective of this paper is to introduce a wavelet shrinkage method for four dimensional MRI data, which implies 3-dimensional MRI images plus the time dimension. We will evaluate the performance using this wavelet based noise reduction for DCE-MRI, through simulation studies.

2 Methods

2.1 Statistical image modelling

The goal of image denoising and restoration is to relieve human observers from the task of distinguishing various artifacts from the underlying image, by reconstructing a plausible estimate of the original image form the distorted or noisy observation. A prior probability model for both the noise and the uncorrupted images is of central importance for this application.

Due to the high dimensionality of the signal, statistical modelling of natural images is a challenging task. Two basic assumptions are commonly made in order to reduce dimensionality: locality and homogeneity. The locality indicates that the probability structure is defined locally, typically by a Markov assumption. That is, the probability density of a pixel, conditioned on a set of neighbours, is independent of the pixels beyond the neighbourhood. The spatial homogeneity implies that the distribution of values in a neighbourhood is the same for all such neighbourhoods, regardless of absolute spatial position. Markov random field (MRF) model is commonly simplified by assuming the distributions are Gaussian. This last assumption, however, is problematic for image modelling, where the complexity of local structures is not well described by the Gaussian densities, especially when the signal-to-noise ratio is lower.

The power of statistical image models can be substantially improved by transforming the signal from the original domain to a new representation. Over the past decades, decomposing images with a set of multiscale band-pass oriented filters (such as wavelet decomposition) has become standard to image processing. It is effective at decoupling the high-order statistical features of natural images. It has been shown that the marginal distributions of wavelet coefficients are highly kurtotic and can be modelled by heavy-tailed distributions. A variety of parametric models has been proposed, including the generalised Laplacian [10, 13, 27], the Bessel K [28], the multivariate Student's t-distribution [31], the α-stable family [20], and the Cauchy distribution [24]. Within the subbands of a representation, the kurtotic behaviours of coefficients allow one to remove noise using a point nonlinearity. Such approaches have become quite popular in image denoising, as described in the next subsection.

In addition to the non-Gaussian marginal behaviour, the responses of band-pass filters exhibit important non-Gaussian joint statistical behaviour. In particular, even when they are second-order decorrelated, the coefficients of similar position, orientation and scale are highly correlated [26, 36]. The dependencies between wavelet coefficients have been investigated, and it is found [4, 26, 34] that large amplitude coefficients are sparsely distributed through-

out the image and tend to occur in clusters and the standard deviation of a coefficient scales roughly linearly with the amplitude of nearby coefficients. Vannucci and Corradi [32, 33] studied the covariance structure of wavelet coefficients within and across scales and suggested a Bayesian approach to the wavelet shrinkage. Furthermore, the dependency between local coefficients and the associated marginal behaviours can be modelled using a random field with a spatially fluctuating variance. A particularly useful example arises from the product of a Gaussian vector and a hidden scalar multiplier, known as a *Gaussian scale mixture* (GSM) [1]. GSM distributions represent an important family of the elliptically symmetric distributions, which are those that can be defined as functions of a quadratic norm of the random vector. Several studies have assumed that the local variance is governed by a continuous multiplier variable [16, 19, 34, 35]. This model can capture the strongly leptokurtotic behaviour of the marginal densities of natural image wavelet coefficients, as well as the correlation in their local amplitudes.

2.2   Wavelet domain denoising

For two-dimensional data such as MRI images, the 2-D discrete wavelet transform (DWT) [18] translates the image content into an approximation subband and a set of detail subbands at different orientations and resolution scales. Typically, the bandpass content at each scale is divided into three orientations subbands characterised by horizontal, vertical and diagonal directions. The approximation subband consists of the scaling coefficients and the detail subbands are composed of wavelet coefficients. Here we consider a undecimated wavelet transform [18] where the number of the wavelet coefficients is equal at each scale to the original size.

There are several properties of the wavelet transform which make this representation attractive for denoising, for instance:

- multiresolution – image details of different sizes are analysed at the appropriate resolution scales

- sparsity – the majority of the wavelet coefficients are small in magnitude

- edge detection – large wavelet coefficients coincide with image edges

- edge clustering – the edge coefficients within each subband tend to form spatially connected clusters

- edge evolution across scales – the coefficients that represent image edges tend to persist across the scales.

Wavelets have been used for denoising in many medical imaging applications, see [21] and the references therein. In general a wavelet shrinkage method compares empirical wavelet coefficients with some threshold and sets them towards zero if their magnitudes are less than the threshold value. The threshold acts like an oracle, which distinguishes between significant and insignificant wavelet coefficients. The coefficients at coarsest scale (the scaling coefficients) are typically left intact, while coefficients at all other scales (wavelet coefficients) are thresholded via wavelet shrinkage, the idea of which is to heavily suppress those coefficients that represent noise and to retain the coefficients that are more likely to represent the actual signal or image discontinuities. For any shrinkage scheme to be effective, an essential property is that the magnitude of the signal should be larger than that of existing noise. The shrinkage operator is typically a piecewise linear and monotonically nondecreasing function. Thus, in practice, the shrinkage operator will not introduce artifact.

Multiscale representations provide a useful tool for representing the structures of signals/images. The widely used orthonormal or biorthogonal wavelet transform leads to successful implementation in image compression. But the results are far from optimal for other applications such as denoising and detection. This is mainly due to that the DWT is critically sampled (the number of coefficients is equal to the number of image pixels) and therefore the loss of the translation-invariance property in the DWT, leading to disturbing visual artifacts (such as aliasing or ringing). A widely followed solution to this problem is to use basis functions designed for orthogonal or biorthogonal systems, but to reduce or eliminate the decimation of the subbands [2, 6, 9, 12, 25].

Once the constraint of critical sampling has been dropped, however, there is no need to limit oneself to these basis functions. Significant improvement comes from the use of representations with a higher degree of redundancy, as well as increased selectivity in orientation [5, 23, 29].

Here we are going to use the undecimated wavelet transform for representing the four-dimensional MRI data. Figure 1 presents an illustrative scheme of such a decomposition together with the wavelet shrinkage.

The method developed by Portilla et al. [23] is still one of the best performing noise reduction methods for MRI data. It uses scale mixture of Gaussians as a model for the wavelet coefficients.

2.3 Gaussian Scale Mixtures

The GSM model has been used successfully to describe the statistics of local clusters of multiscale subband coefficients, which can include spatial neigh-
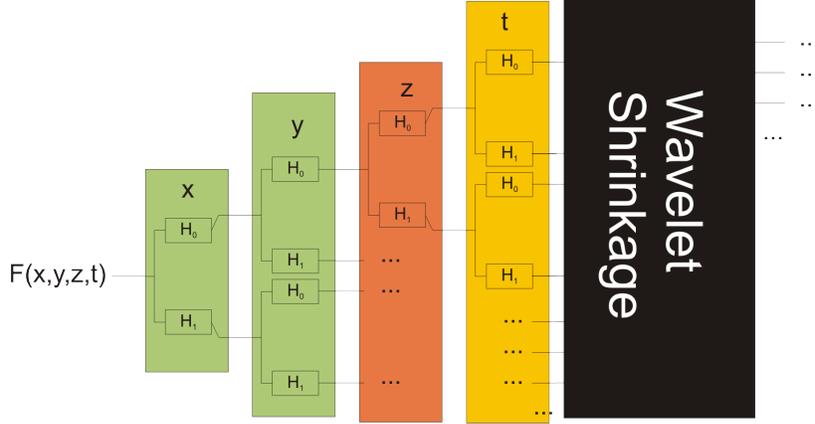
Figure 1: Illustration of 4D undecimated wavelet transform and shrinkage

bours as well as coefficients in adjacent scale and orientation subbands (e.g. [11, 23, 35]). By definition, a GSM density is an infinite mixture of zero-mean Gaussian variables with covariances related by multiplicative scaling. It can emulate many of the non-Gaussian statistical behaviours observed in neighbours of subband coefficients. In addition, the underlying Gaussian structure leads to relatively simple parameter learning and inference procedures.

A GSM random vector $\mathbf{x}$ is defined as the product of a zero-mean Gaussian vector $\mathbf{u}$ and an independent positive scalar variable z:

$$\mathbf{x} \stackrel{d}{=} \sqrt{z} \cdot \mathbf{u}, \tag{1}$$

where $\mathbf{x}, \mathbf{u} \in \mathcal{R}^N$, $z \in \mathcal{R}^+$ independent of $\mathbf{x}$, and N is the dimensionality of $\mathbf{x}$ and $\mathbf{u}$. In our case this dimensionality is the size of the neighbourhood of coefficients clustered around the reference coefficients. In general, the neighbourhood may include coefficients from other subbands at nearby scales and orientations. The density of a GSM vector $\mathbf{x}$ is given as

$$
\begin{aligned}
p(\mathbf{x}) &= \int p(\mathbf{x}|z) p_z(z) dz \\
&= \int \frac{1}{\sqrt{(2\pi z)^N |\Sigma|}} \exp\{-\frac{\mathbf{x}^T \Sigma^{-1} \mathbf{x}}{2z}\} p_z(z) dz.
\end{aligned}
\tag{2}
$$

It is determined by the covariance matrix Σ of $\mathbf{u}$ and the mixing density $p_z(z)$. The conditions under which a random vector may be represented by a GSM have been studied [1]. As a family of probability distributions, GSM includes many common kurtotic distributions, such as α-stable family (including

5

Cauchy distribution), the generalised Gaussian family, and the symmetrised Gamma family. For instance, if z follows an inverse Gamma distribution, the resulting GSM density reduces to a multivariate Student's t-distribution [1, 37]. As theoretical properties of the GSM family, it includes that GSM densities are symmetric about zero and Gaussian when conditioned on z, and they have leptokurtic marginal densities.

2.4 Model wavelet coefficients by GSM

By using the wavelet transform, we decompose the image into pyramid subbands at different scales and orientations. Denote by $x_c^{s,o}(n,m)$ the wavelet coefficient at scale s, orientation o, spatial position $(2^s n, 2^s m)$. Denote also by $\mathbf{x}^{s,o}(n,m)$ a neighbourhood of coefficients clustered around this reference coefficient $x_c^{s,o}(n,m)$. For notational simplicity, we drop hereafter the superscripts s,o and indices (n,m). We assume that this neighbourhood $\mathbf{x}$ (of dimension N) is characterised by a GSM model as in (1). Assume also that the image is corrupted by independent additive white Gaussian noise. Thus a vector corresponding to a neighbourhood of N observed coefficients of the pyramid representation can be modelled by the GSM ([17, 23]):

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon} = \sqrt{z} \cdot \mathbf{u} + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{y}$ is a neighbourhood of N observed wavelet coefficients and $\boldsymbol{\epsilon}$ zero-mean Gaussian. Note that based on the assumptions the three random components on the righthand side of (3) are mutually independent.

Both $\mathbf{u}$ and $\boldsymbol{\epsilon}$ are zero mean Gaussian with covariance matrices $\Sigma_{\mathbf{u}}$ and $\Sigma_{\boldsymbol{\epsilon}}$, respectively. Hence, the conditional distribution of the observed neighbourhood vector $\mathbf{y}$ given z is also a zero-mean Gaussian, with covariance matrix

$$\Sigma_{\mathbf{y}|z} = z\Sigma_{\mathbf{u}} + \Sigma_{\boldsymbol{\epsilon}}, \tag{4}$$

and density

$$p(\mathbf{y}|z) = \frac{1}{\sqrt{(2\pi)^N |z\Sigma_{\mathbf{u}} + \Sigma_{\boldsymbol{\epsilon}}|}} \exp\left\{-\frac{\mathbf{y}^T (z\Sigma_{\mathbf{u}} + \Sigma_{\boldsymbol{\epsilon}})^{-1}\mathbf{y}}{2}\right\}. \tag{5}$$

The neighbourhood noise covariance $\Sigma_{\boldsymbol{\epsilon}}$ can be estimated separately from some regions where signals are (almost) not existing. It can also be obtained, as pointed in [23], by decomposing a delta function in two dimension $\sigma\sqrt{D_x D_y}\delta(n,m)$ into pyramid subbands, where (D_x, D_y) is the image dimension. Similar decomposition can be done in four dimension. This delta signal

6

has the same power spectrum as the noise, but it is free from random fluctuations. Elements of $\Sigma_{\boldsymbol{\epsilon}}$ may then be computed directly as sample covariances. This procedure is easily generalised for non-white noise, by replacing the delta function with the inverse Fourier transform of the square root of the noise power spectral density. Note that the entire procedure may be performed off-line, as it is signal-independent.

Taking expectation over z on (4) it follows that

$$\Sigma_{\mathbf{y}} = E[z]\Sigma_{\mathbf{u}} + \Sigma_{\boldsymbol{\epsilon}}. \tag{6}$$

By setting $E[z] = 1$ without loss of generality, we obtain

$$\Sigma_{\mathbf{u}} = \Sigma_{\mathbf{y}} - \Sigma_{\boldsymbol{\epsilon}}. \tag{7}$$

Thus, given the noise covariance $\Sigma_{\boldsymbol{\epsilon}}$, the signal covariance $\Sigma_{\mathbf{u}}$ can be computed from the observation covariance $\Sigma_{\mathbf{y}}$ according (7).

In order to estimate the center coefficient x_c from the observed noisy neighbourhood $\mathbf{y}$, the Bayes least squares estimator (BLSE), which minimises the expected square error, can be applied. BLSE is the conditional mean of x_c given $\mathbf{y}$ [15]:

$$
\begin{aligned}
E[x_c|\mathbf{y}] &= \int x_c p(x_c|\mathbf{y}) dx_c \\
&= \int \int_0^\infty x_c p(x_c, z|\mathbf{y}) dz dx_c \\
&= \int \int_0^\infty x_c p(x_c|\mathbf{y}, z) p(z|\mathbf{y}) dz dx_c \\
&= \int_0^\infty p(z|\mathbf{y}) E[x_c|\mathbf{y}, z] dz,
\end{aligned}
\tag{8}
$$

under the assumption of uniform convergence in order to exchange the order of integration. According to (8), this BLS estimator can be computed by the weighted average of the Bayes least squares estimate of x_c when conditioned on z, with weights of the posterior density $p(z|\mathbf{y})$. The detailed treatment of each component is referred to [23].

Furthermore, as for the prior density $p_z(z)$ of the multiplier, the commonly used Jeffreys' prior is chosen here. This noninformative prior distribution on the parameter space is proportional to the square root of the determinant of the Fisher information. It was suggested by Jeffreys [14], which tries (in the spirit of invariance) to treat all parameter values equitably. The key feature is that it is invariant under reparameterisation of the parameter vector, which makes it of special interest for use with scale parameters.

2.5 Denoising algorithm

Based on the methodologies described in the previous subsections, we have the denoising algorithm:

1. Decompose the image into subbands (Fig. 1)

2. For each subband (except the lowpass residual):

   - Compute neighbourhood noise covariance $\Sigma_{\boldsymbol{\epsilon}}$ from the image-domain noise covariance as discussed in Subsection 2.4
   - Estimate noisy neighbourhood covariance $\Sigma_{\mathbf{y}}$ by observations
   - Estimate $\Sigma_{\mathbf{u}}$ from $\Sigma_{\boldsymbol{\epsilon}}$ and $\Sigma_{\mathbf{y}}$ using (7)
   - For each neighbourhood, compute $E[x_c|\mathbf{y}]$ using (8)

3. Reconstruct the denoised image from the processed subbands and the lowpass residual

3 Simulation study

For evaluating the performance of our wavelet-based noise reduction method for 4D signals, we have conducted a simulation study. The data were generated from the real DCE-MRI signals and each contaminated with synthesised additive Gaussian white noise at five different variances which providing five different signal-to-noise ratios.

3.1 Setups: Data generation

a) Fitting a parametric model to a real DCE-MRI data (Fig. 2)

b) The estimated parameters were assumed to be true and new data was created with that spatial distribution of parameters

c) Gaussian noise with five different SNR levels was added to the "true" signal obtained from a)-b)

d) Restrict the data to the tumour region ($64 \times 56 \times 10 \times 56$) with some margins for avoiding the edge effect

Based on the data that we generated, we performed our noise reduction algorithm on the whole 4D data set as well as 3D slices of the data.

8
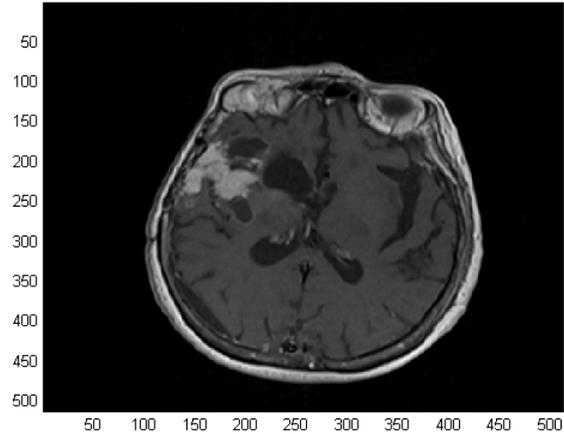
Figure 2: An anatomic brain image

3.2 Setups: Denoising

1. Analysis was performed in MatLab on the HPC2N cluster¡a href="http://www.hpc2n.umu.se/"¿HPC2N cluster¡/a¿

2. Noise reduction was performed on the whole data set (4D), as well as 3D slices of the data

3. The proposed algorithm for 4D wavelet transform: non-decimated, Daubechies "db4", 2 scales

4. The proposed algorithm for noise reduction in 4D using GSM

5. Results were evaluated on noise reduction using RMSE and compared with smoothing

4 Results and conclusions

The simulation results (Fig. 3) show that our new algorithm outperforms the common smoothing technique. It improves substantially the noise reduction at every tested noise levels, on both 3D and 4D signals. Furthermore, it has also shown that one can gain more by utilising the full 4D signals.

One drawback of this denoising algorithm is the computational time when
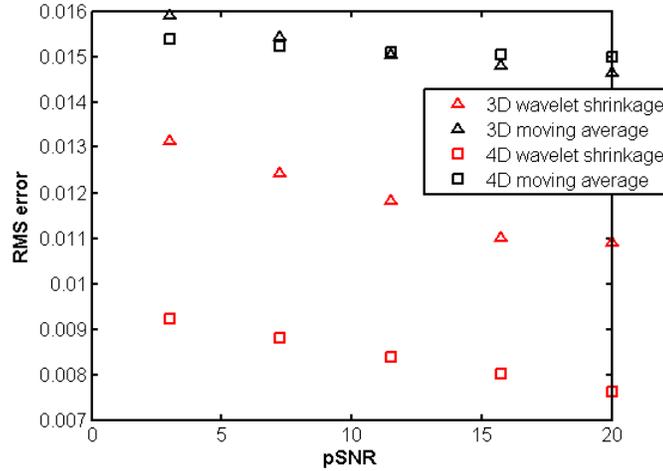
Figure 3: RMSE comparison between wavelet shrinkage and moving average for both 3D and 4D signals at different SNR levels

the dimension is increasing. This is partly solved by using the HPC2N clusters at Umeå University. Next step? One would think the following topics:

- Different wavelets for spatial dimension and time series

- New method based on Rician noise when the SNR is low

- Maximum likelihood estimator for noise variance

- Method for estimating a prior for the multiplier by maximising the joint likelihood

- Error propagation in DCE-MRI for investigating how the noise reduction will affect the uncertainty in estimation of parameters presented in the physiological models

Acknowledgment

References

[1] D.F. Andrews and C.L. Mallows. Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B*, 36:99–102, 1974.

[2] A. Bijaoui and M. Giudicelli. Optimal image addition using the wavelet transform. *Experimental Astronomy*, 1:347–363, 1991.

[3] A. Bruce and H.Y. Gao. Understanding waveshrink: Variance and bias estimation. *Biometrika*, 83:727–745, 1996.

[4] R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Processing*, 8(12):1688–1701, December 1999.

[5] G.Y. Chen and B. Kégl. Image denoising with complex ridgelets. *Pattern Recognition*, 40:578–585, 2007.

[6] R.R. Coifman and D.L. Donoho. Translation-invariant denoising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, Lecture Notes in Statistics. Springer-Verlag, 1995.

[7] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[8] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57:301–369, 1995.

[9] P. Dutilleux. An implementation of the "algorithm à trous" to compute the wavelet transform. In J.M. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase-Space*, pages 298–304. Springer, New York, 1989.

[10] P. Gehler and M. Welling. Products of "edge-perts". In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Proc. Advances in Neural Information Processing Systems*, pages 419–426, 2006.

[11] J. Guerrero-Colon, L. Mancera, and J. Portilla. Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids. *IEEE Trans Image Processing*, 17(1):27–41, January 2008.

[12] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet

transform. In J.M. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase-Space*, pages 286–297. Springer, New York, 1989.

[13] J. Huang and D. Mumford. Statistics of natural images and models. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, 1999.

[14] H. Jeffreys. An invariant form for the prior probability in estimation problems. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, volume 186, pages 453–461, 1946.

[15] E.L. Lehmann. *Theory of Point Estimation*. John WIley & Sons, 1983.

[16] S.M. LoPresto, K. Ramchandran, and M.T Orchard. Wavelet image coding based on a new generalized gaussian mixture model. In *Proc. Data Compression Conference*, Snowbird, UT, March 1997.

[17] S. Lyu and E.P. Simoncelli. Statistical modeling of images with fields of gaussian scale mixtures. In *Advances in Neural Computation Systems*, pages 945–952. NIPS, Canada, May 2006.

[18] S. Mallat. *A Wavelet Tour of Signal Processing - The Sparse Way*. Academic Press, third edition, 2009.

[19] M.K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Trans. Signal Processing*, 6:300–303, December 1999.

[20] L. Parra, C. Spence, and P. Sajda. Higher-order statistical properties arising from the non-stationarity of natural signals. In *Proc. Advances in Neural Information Processing Systems*, volume 13, 2000.

[21] A. Pižurica and W. Philips. Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Trans Image Processing*, 15(3):654–665, 2006.

[22] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy. A joint inter- and intrascale statistical model for bayesian wavelet based image denoising. *IEEE Trans. Image Processing*, 11(5):545–557, 2002.

[23] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans Image Process.*, 12(11):1338–1351, 2003.

[24] L. Sendur and I.W. Selesnick. Bivaraite shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Trans Signal Processing*, 50(11):2744–2756, November 2002.

[25] M. Shensa. Discrete wavelet transforms: Wedding the à trous and mallat algorithms. *IEEE Trans Signal Processing*, 40:2464–2482, 1992.

[26] E.P Simoncelli. Statistical models for images: compression, restoration and synthesis. In *Proc 31st Asilomar Conf on Signals, Systems and Computers*, volume 1, pages 673–678, November 1997.

[27] E.P. Simoncelli and E.H. Adelson. Noise removal via bayesian wavelet coring. In *Proc. Third IEEE Int'l Conf. Image Processing*, volume 1, pages 379–382, Lausanne, Switzerland, September 1996.

[28] A. Srivastava, X. Liu, and U. Grenander. Universal analytical forms for modeling image probabilities. *IEEE Trans Pattern Analysis and Machine Intelligence*, 24(9):1200–1214, September 2002.

[29] J.L. Starck, E.J. Candès, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Trans Image Processing*, 11(6):670–684, 2002.

[30] J.L Starck, F. Murtagh, and J.M. Fadili. *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity.* Cambridge University Press, 2010.

[31] Y.W. Teh, M. Welling, S. Osindero, and G.E. Hinton. Energy-based models for sparse overcomplete representations. *Journal of Machine Learning Research*, 4:1235–1260, 2003.

[32] M. Vannucci and F. Corradi. Covariance structure of wavelet coefficients: theory and models in a bayesian perspective. *Journal of Royal Statistical Society, Ser. B*, 61:971–986, 1999.

[33] M. Vannucci and F. Corradi. Modeling dependence in the wavelet domain. In P. Müller and B. Vidakovic, editors, *Bayesian Inference in Wavelet-Based Models*, number 141 in Lecture Notes in Statistics, pages 173–186. Springer, 1999.

[34] M.J. Wainwright and E.P. Simoncelli. Scale mixtures of gaussians and the statistics of natural images. In S.A. Solla, T.K. Leen, and K.R. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 855–861. MIT Press, Cambridge, MA, 2000.

[35] M.J. Wainwright, E.P. Simoncelli, and A.S. Willsky. Random cascades on wavelet trees and their use in modeling and analyzing natural imagery. *Appl. Comput. Harmon. Anal.*, 11(1):89–123, July 2001.

[36] B. Wegmann and C. Zetzsche. Statistical dependence between orientation filter outputs used in an human vision based image code. In *Proc. Visual Comm. Image Processing*, volume 1360, pages 909–922, Lausanne, Switzerland, 1990.

[37] M. Welling, G.E. Hinton, and S. Osindero. Learning sparse topographic representations with products of student t-distribution. In *Proc. Advances in Neural Information Processing Systems*, pages 1359–1366, 2002.