Journal of
Microscopy

Journal of Microscopy, Vol. 255, Issue 3 2014, pp. 123–127 doi: 10.1111/jmi.12151
Received 6 March 2014; accepted 3 June 2014

Wavelets: on the virtues and applications of the mathematical microscope

MICHAEL UNSER
Biomedical Imaging Group, EPFL, Lausanne, Switzerland

Key words. Deconvolution, denoising, wavelets, image analysis, multiresolution, computational imaging.

Summary

The paper provides a short introduction to wavelets and discusses their main applications in microscopy and biological imaging.

Wavelets as a virtual microscope

Wavelets offer a powerful way of decomposing signals or images into their elementary constituents across scale (multiresolution decomposition). They provide a one-to-one representation in very much the same way as the Fourier transform does, with the fundamental difference that the wavelet-basis functions are jointly localized in space and frequency (Daubechies, 1988; Mallat, 1989; Unser & Aldroubi, 1996; Mallat, 2009).

There is a striking analogy between the wavelet transform and a microscope. To keep the discussion simple, we shall focus on the one-dimensional scenario where the input image $f(x)$ is a function of the space variable $x \in \mathbb{R}$. The wavelet transform involves a series of dyadic magnification factors $a = 2^i$, with $i \in \mathbb{Z}$. It corresponds to a mathematical microscope whose point-spread function (PSF) ϕ can be dilated (or contracted when $i < 0$) at will by powers of two, like

$$\phi_i(x) = \phi(x/2^i).$$

There the virtual PSF is a reference function ϕ (scaling function or wavelet) to be specified in the sequel. The leading concept behind wavelets is to observe $f(x)$ by forming correlations with ϕ_i (or convolutions with its space-reversed version) and sampling the data at the appropriate rate. Specifically, the observation at resolution level i and location index $k \in \mathbb{Z}$ is given by

$$\langle f, \phi_{i,k} \rangle = \int_{\mathbb{R}} f(x)\phi_i(x - 2^i k)\mathrm{d}x,$$

with the notational convention

$$\phi_{i,k}(x) = \phi_i(x - 2^i k)$$

where the corresponding sampling step $a = 2^i$ is matched to the size of the integration window ϕ_i (virtual PSF at resolution 2^i). The underlying principle is that the virtual PSF is engineered to maximize the intake of information; first, within a given resolution level i (scale of the microscope) by ensuring that the shifted replicates of ϕ_i – that is, $\phi_{i,k}$ with $k \in \mathbb{Z}$ – are orthogonal to one another, and, second, by avoiding redundancies across scale (global inter-scale orthogonality). The latter requirement translates into the virtual PSF being bandpass (that is, a wavelet denoted by $\phi = \psi$) rather than the more traditional lowpass solution (denoted by $\phi = \varphi$) that would better fit the description of a physical microscope.

The Haar transform: from Legos to wavelets

The fundamental idea in the theory of wavelets is to construct a series of fine-to-coarse approximations $\{f_i\}_{i\in\mathbb{Z}}$ of a function $f(x)$ and to exploit the structure of the multiresolution approximation errors, which are orthogonal across scale. Here, we shall illustrate the concept by taking f_i (the best approximation of f at resolution i) to be a piecewise-constant function represented by the (Lego-like) expansion

$$f_i(x) = \sum_{k\in\mathbb{Z}} c_i[k]\varphi_{i,k}(x), \tag{1}$$

where the basis functions $\varphi_{i,k}$ are adjacent rectangular functions of size 2^i. Specifically,

$$\varphi_{i,k}(x) = \varphi\left(\frac{x - 2^i k}{2^i}\right) = \begin{cases} 1, & \text{for } x \in [2^i k, 2^i(k+1)) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The optical analogy is that of a microscope with ideal optics whose resolution is determined by the size of its CCD-type (charge coupled device) detector (rectangular window of size 2^i), while the spatial location is encoded by k. An example of such a sequence of approximations is shown in Fig. 1, where each $c_i[k]$ is given by the height of the corresponding

This work was founded in part by the Swiss National Science Foundation under Grant 200020-121763.
Correspondence to: Michael Unser, Biomedical Imaging Group, EPFL, CH-1015 Lausanne, Switzerland. Tel +41 21 693 51 75; e-mail: michael.unser@epfl.ch

© 2014 The Authors
Journal of Microscopy © 2014 Royal Microscopical Society

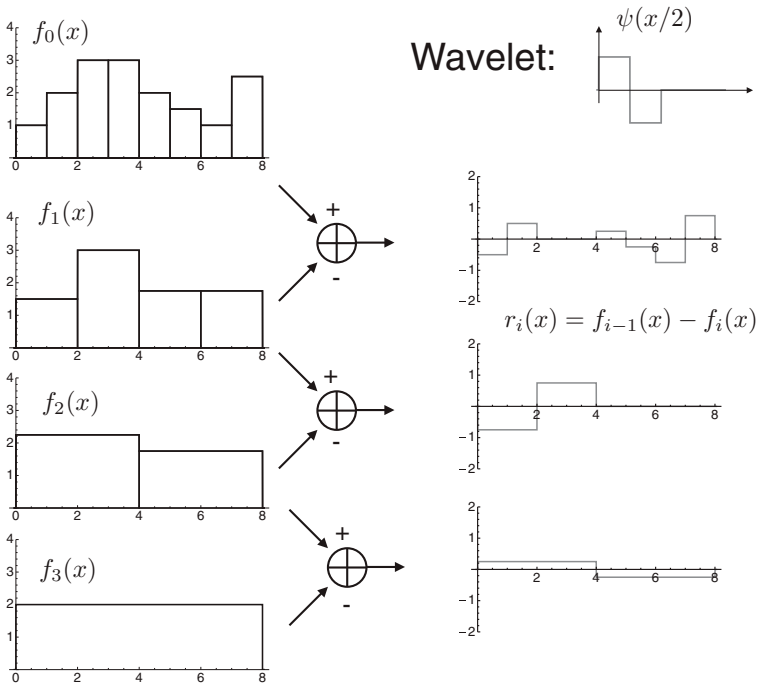

Fig. 1. Multiresolution signal analysis using piecewise-constant basis functions with a dyadic scale progression. Left: multiresolution pyramid. Right: error signals between two successive levels of the pyramid.

piecewise-constant segment of the signal. For the purpose of demonstration, we select the initial signal as $f(x) = f_0(x)$.

The primary template $\varphi = \varphi_{0,0}$ is called the *scaling function*. It satisfies the following three properties, which are key to the construction of a wavelet basis of $L_2(\mathbb{R})$ (the space of finite-energy functions) (Unser & Blu, 2003):

(1) Orthonormality: $\langle \varphi, \varphi(\cdot - k) \rangle = \delta_k$ for all $k \in \mathbb{Z}$.
(2) Two-scale relation:

$$\varphi(x/2) = \sum_{k \in \mathbb{Z}} h[k] \varphi(x - k), \tag{3}$$

where the sequence h is the so-called *refinement mask*.
(3) Partition of unity: $\sum_{k \in \mathbb{Z}} \varphi(x - k) = 1$.

In practice, a given brand of wavelets (e.g. Haar, Daubechies, splines) is summarized by its refinement mask h which uniquely specifies φ via the solution of (3). The rectangular functions defined by (2) with $i = 0$ are orthogonal within a given scale simply because they are nonoverlapping. Their refinement mask is $h = (1, 1)$, which translates into what we jokingly refer to as the *Lego-Duplo* relation[1]

$$\varphi(x/2) = 1 \times \varphi(x) + 1 \times \varphi(x - 1). \tag{4}$$

[1] The Duplos are the large-scale counterpart of the Lego building blocks. The main point for the analogy with wavelets is that Legos and Duplos are compatible; they can be combined to build complex shapes. The enabling property is that a Duplo is equivalent to two smaller Lego placed next to each other, as expressed by (4).

They also satisfy the partition of unity, as can be checked by setting $c_i[k] = 1$ in (1).

By considering the rescaled version of such a basis, we specify the signal subspace V_i at resolution i as

$$V_i = \left\{ f_i(x) = \sum_{k \in \mathbb{Z}} c_i[k] \varphi_{i,k}(x) : c_i \in \ell_2(\mathbb{Z}) \right\} \subset L_2(\mathbb{R})$$

which, in our example, contains all the finite-energy functions that are piecewise constant on the intervals $\left[2^i k, 2^i(k+1)\right)$ with $k \in \mathbb{Z}$. The two-scale relation (3) implies that the basis functions at scale $i = 1$ are contained in V_0 (our original signal space in Fig. 1) and, by extension, in V_i for $i \leq 0$. This translates into the general inclusion property $V_{i'} \subset V_i$ for any $i' > i$, which is fundamental to the theory.

Let us now consider the multiresolution approximation of $f_0(x)$, as illustrated in Fig. 1. Given the sequence of fine-scale coefficients $c_0[k]$, we construct the best piecewise-constant approximation $f_1(x)$ at scale 1 specified by its coefficients $c_1[k]$ in (1) with $i = 1$. The minimum-error solution is obtained by taking the mean of two consecutive samples. The process is repeated iteratively until one reaches the bottom of the pyramid, as shown on the left-hand side of Fig. 1. The description of this coarsening algorithm is

$$c_i[k] = \frac{1}{2} c_{i-1}[2k] + \frac{1}{2} c_{i-1}[2k+1] = \big(c_i * \tilde{h}\big)[2k], \tag{5}$$

© 2014 The Authors

which is run recursively for $i = 1, \ldots, i_{\max}$, where $i_{\max}$ denotes the bottom level of the pyramid.

The key observation for uncovering the wavelets is that the residuals $r_i(x) = f_{i-1}(x) - f_i(x) \in V_{i-1}$ on the right of Fig. 1 exhibit a characteristic sign-alternation pattern which is due to the fact that the two consecutive samples $(c_{i-1}[2k], c_{i-1}[2k+1])$ are at an equal distance from their mean value $c_i[k]$. This suggests rewriting the residuals as

$$r_i(x) = f_{i-1}(x) - f_i(x) = \sum_{k\in\mathbb{Z}} d_i[k]\psi_{i,k}(x), \tag{6}$$

where the characteristic oscillating pattern is encoded in the wavelets $\psi_{i,k} = \psi\left(\frac{x-2^i k}{2^i}\right)$, which are rescaled and shifted replicates of a single template: the Haar wavelet

$$\psi(x) = \begin{cases} 1, & \text{for } x \in [0, 1/2) \\ -1 & \text{for } x \in [1/2, 1) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

This function is shown on the upper right of Fig. 1. The wavelet coefficients $d_i[k]$ are given by the consecutive half-differences

$$d_i[k] = \frac{1}{2}c_{i-1}[2k] - \frac{1}{2}c_{i-1}[2k+1] = \big(c_i * \tilde{g}\big)(2k). \tag{8}$$

More generally, since the wavelet template $\psi_{1,0}$ at scale $i = 1$ belongs to V_0, we can write that

$$\psi(x/2) = \sum_{k\in\mathbb{Z}} g[k]\varphi(x-k), \tag{9}$$

which is the wavelet counterpart of the two-scale relation (3). In the present example, we have that $g[k] = (-1)^k h[k]$, which is a general relation that is characteristic of an orthogonal design. Likewise, in order to gain generality, we have chosen to express the decomposition Eqs (5) and (8) (fast wavelet-transform algorithm) in terms of discrete convolution (filtering) and down-sampling operations where the corresponding Haar analysis filters are $\tilde{h}[k] = \frac{1}{2}h[-k]$ and $\tilde{g}[k] = \frac{1}{2}(-1)^k h[-k]$. The Hilbert-space interpretation of this approximation process is that $r_i \in W_i$, where W_i is the orthogonal complement of V_i in V_{i-1}; that is, $V_{i-1} = V_i + W_i$ with $V_i \perp W_i$, as a consequence of the orthogonal projection theorem.

Finally, we close the loop by observing that

$$\begin{aligned} f_0(x) &= f_{i_{\max}}(x) + \sum_{i=1}^{i_{\max}} \Big(\underbrace{f_{i-1}(x) - f_i(x)}_{r_i(x)} \Big) \\ &= \sum_{k\in\mathbb{Z}} c_{i_{\max}}[k]\varphi_{i_{\max},k}(x) + \sum_{i=1}^{i_{\max}} \sum_{k\in\mathbb{Z}} d_i[k]\psi_{i,k}(x), \end{aligned} \tag{10}$$

which provides an equivalent, one-to-one representation of the signal in an orthogonal wavelet basis, as illustrated in Fig. 2. More generally, we can push the argument to the limit and apply the decomposition to any finite-energy function

$$\forall f \in L_2(\mathbb{R}), \quad f = \sum_{i\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_i[k]\psi_{i,k}, \tag{11}$$

where $d_i[k] = \langle s, \tilde{\psi}_{i,k}\rangle$ and $\{\tilde{\psi}_{i,k}\}_{i,k\in\mathbb{Z}}$ is a suitable (bi-)orthogonal wavelet basis with the property that $\langle \tilde{\psi}_{i,k}, \psi_{i',k'}\rangle = \delta_{k-k',i-i'}$.

Remarkably, the whole process described above – with the exception of the explicit right-hand side of (2), (7) and the central expressions in (5) and (8) – is completely generic and applicable for any other wavelet basis of $L_2(\mathbb{R})$ that is specified in terms of a wavelet filter g and a scaling function φ (or, equivalently, an admissible refinement filter h). The bottom line is that the wavelet decomposition and reconstruction algorithm is fully described by the four digital filters $(h, g, \tilde{h}, \tilde{g})$ that form a perfect-reconstruction filterbank (Vetterli & Kovacevic, 1995). The Haar transform is associated with the shortest-possible filters. Its only downside is that the basis functions are discontinuous and that the scale-truncated error decays no faster than the first power of the sampling step $a = 2^i$ (first order of approximation).

The concept is generalizable to higher dimensions by using tensor-product basis functions, which practically amounts to running the 1-D wavelet-decomposition algorithm (Eqs (5) and (8)) along every dimension of the data. Specifically in 2-D, one has to consider the separable scaling function $\varphi(x)\varphi(y)$ and the tensor-product wavelets $\varphi(x)\psi(y)$, $\psi(x)\varphi(y)$ and $\psi(x)\psi(y)$, which results in three complementary wavelet channels (horizontal, vertical and diagonal) for each scale. The process of computing the separable wavelet transform of an image is illustrated in Fig. 3. The important point is that the wavelet coefficients in Fig. 3(d) are in one-to-one correspondence with the image in Fig. 3 a (orthonormal transform). Also, a large proportion of these coefficients are very small – and hence, negligible. This explains why wavelets are so effective for image coding. Indeed, wavelets are the basis of the JPEG 2000 compression standard (Christopoulos *et al.*, 2000).

Another twist is to consider a redundant wavelet representation that does not involve sub-sampling at coarser scales. This leads to the concept of a *wavelet frame* which consists of a union of wavelet bases corresponding to different shifts of the data. While using a wavelet frame requires more storage, it can have advantages for data processing because the representation is intrinsically shift-invariant (Unser, 1995), while a decomposition in a wavelet basis is only approximatively so.

Applications in microscopy

We conclude the presentation with a brief discussion of successful uses of wavelets in microscopy. Most of these applications exploit the property that the wavelet transform of a

© 2014 The Authors

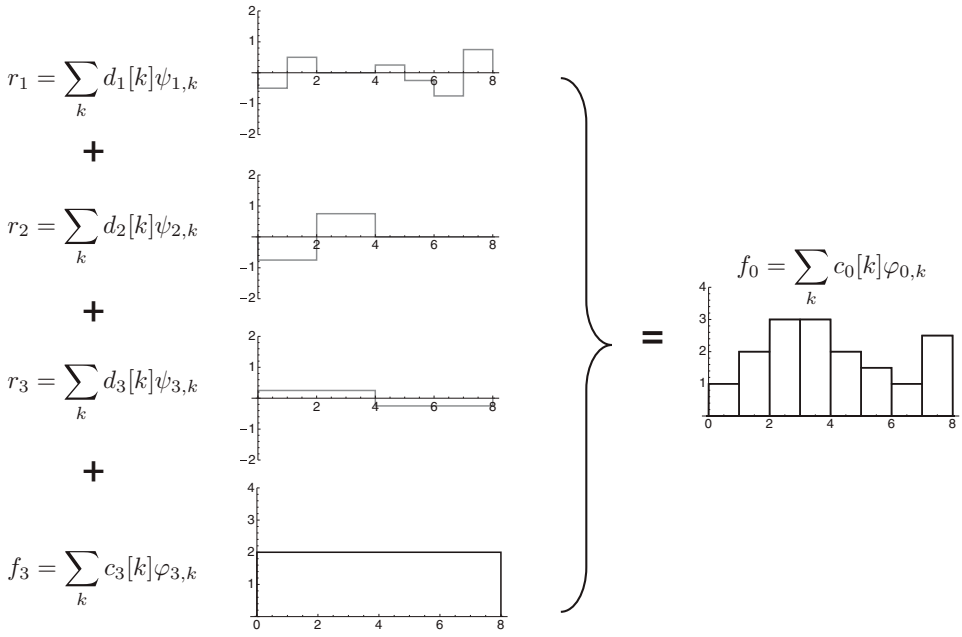

Fig. 2. Decomposition of a signal into orthogonal scale components. The error signals $r_i = (f_{i-1} - f_i)$ between two successive signal approximations are expanded using a series of properly scaled wavelets.

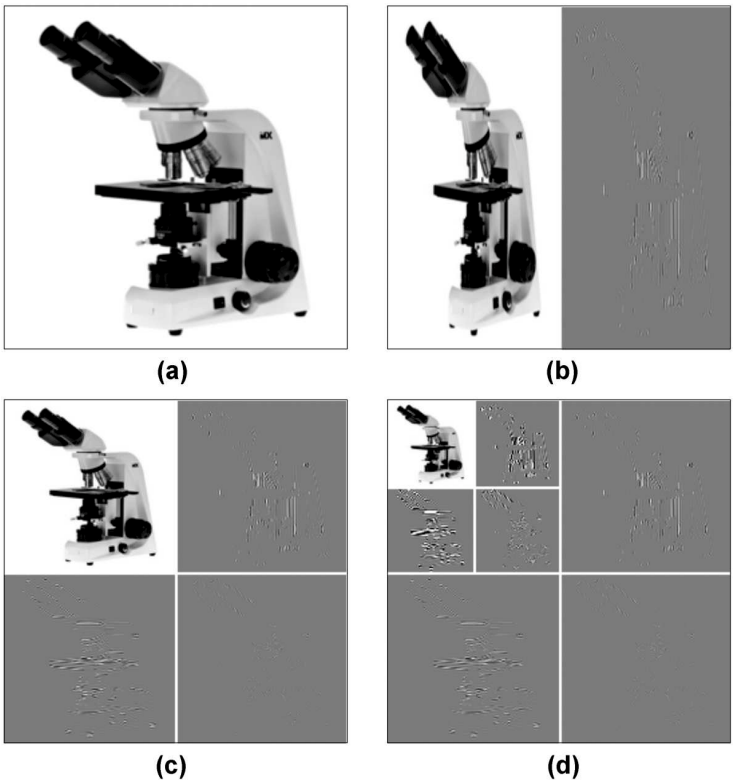

Fig. 3. Separable cubic-spline wavelet decomposition of an image. (a) Original 256×256 image. (b) First pass of (horizontal) 1-D wavelet decomposition: the rows are split into halves. (c) Second pass of (vertical) 1-D wavelet decomposition: the columns are split into halves. (d) Full 2-D wavelet transform obtained by iterating the splitting process on the lower-resolution version of the image. The wavelet coefficients are displayed so that the amplitudes of small absolute value appear in mid grey, the negative values in darker grey and the positive values bright.

© 2014 The Authors

typical, noise-free image is *sparse* in the sense that it exhibits a few large wavelet coefficients (principally around edges or in heavy textured areas), while the majority of coefficients (over smooth image regions) have tiny amplitudes so that they can be discarded without noticeable effect on the quality of the reconstruction.

- Multiscale particle detection: Computing the wavelet transform is equivalent to applying a bank of matched filters that are tuned to different scales. Generally, a wavelet that is (quasi-)isotropic and well localized will act as a good spot detector. The bandpass character of the wavelet is very helpful in suppressing slowly varying background structures so that there is no need for additional preprocessing. Moreover, it is possible to combine the output from multiple scales to improve the robustness of the detection (Olivo-Marin, 2002).
- Denoising: The guiding principle here is to exploit the dichotomy between the signal that is concentrated in a few large coefficients and the measurement noise that is spread out evenly in the wavelet domain. Noise is suppressed by 'killing' the small wavelet coefficients by suitable thresholding and reconstructing the signal thereafter (Weaver *et al.*, 1991; Donoho, 1995). A more-elaborate version for microscopy takes into account the photon-limited nature of the noise (Poisson statistics) and applies a self-tuning strategy to optimally adjust the wavelet-domain thresholding functions for minimum-error reconstruction (Luisier *et al.*, 2010, 2011).
- 3-D deconvolution of fluorescence micrographs: Deconvolution is a delicate inversion process that can easily result in unwanted noise amplification. The standard remedy for keeping the noise under control is to impose some 'regularization' constraint on the solution. In the wavelet approach, one favours a solution with a sparse wavelet transform by penalizing the ℓ_1-norm of the wavelet coefficients (Vonesch & Unser, 2008). The corresponding optimization problem is solved iteratively by using a variant of the ISTA (iterative shrinkage-thresholding algorithm) (Figueiredo & Nowak, 2003). The process can be accelerated substantially by using a multilevel strategy with iteration parameters that are adapted to the specific structure of the problem (Vonesch & Unser, 2009). The principle of wavelet-domain regularization is applicable to other inverse problems in imaging, including tomographic reconstruction (Daubechies *et al.*, 2004).
- Extended depth-of-field: The aim here is to reconstruct an in-focus image by combining the data from a focal series of images (z-stack) with a limited depth of field (optical sectioning). This is achieved through a careful merging of the wavelet transforms of the individual images to produce a single composite that is sharp everywhere (Forster *et al.*, September 2004).

References

Christopoulos, C., Skodras, A.S. & Ebrahimi, T. (2000) The JPEG2000 still image coding system: an overview. *IEEE Trans. Consum. Electron.*, **16**(4), 1103–1127.

Daubechies, I. (1988) Orthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math*, **41**(7), 909–996.

Daubechies, I., Defrise, M., & De Mol, C. (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, **57**(11), 1413–1457.

Donoho, D.L. (1995) De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, **41**(3), 613–627.

Figueiredo, M.A.T. & Nowak, R.D. (2003) An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, **12**(8), 906–916.

Forster, B., Van De Ville, D., Berent, J., Sage, D. & Unser, M. (2004) Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. *Microsc. Res. Tech.*, **65**(1–2), 33–42.

Luisier, F., Blu, T. & Unser, M. (2011) Image denoising in mixed Poisson-Gaussian noise. *IEEE Trans. Image Process.*, **20**(3), 696–708.

Luisier, F., Vonesch, C., Blu, T. & Unser, M. (2010) Fast interscale wavelet denoising of Poisson-corrupted images. *Signal Process.*, **90**(2), 415–427.

Mallat, S.G. (1989) A theory of multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intel.*, **11**(7), 674–693.

Mallat, S. (2009) *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd edn. Academic Press, San Diego.

Olivo-Marin, J.C. (2002) Extraction of spots in biological images using multiscale products. *Pattern Recogn.*, **35**(9), 1989–1996.

Unser, M. & Aldroubi, A. (1996) A review of wavelets in biomedical applications. *Proceed. IEEE*, **84**(4), 626–638.

Unser, M. (1995) Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Process.*, **4**(11), 1549–1560.

Unser, M. & Blu, T. (2003) Wavelet theory demystified. *IEEE Trans. Signal Process.*, **51**(2), 470–483.

Vetterli, M. & Kovacevic, J. (1995) *Wavelets and Subband Coding*. Prentice Hall, Englewood Cliffs, NJ.

Vonesch, C. & Unser, M. (2008) A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Trans. Image Process.*, **17**(4), 539–549.

Vonesch, C. & Unser, M. (2009) A fast multilevel algorithm for wavelet-regularized image restoration. *IEEE Trans. Image Process.*, **18**(3), 509–523.

Weaver, J.B., Yansun, X., Healy Jr., D.M. & Cromwell, L.D. (1991) Filtering noise from images with wavelet transforms. *Magn. Reson. Med.*, **21**(2), 288–295.

© 2014 The Authors