Splines and wavelets for medical imaging

Michael Unser

Biomedical Imaging Group,
Swiss Federal Institute of Technology Lausanne
Email: Michael.Unser@epfl.ch; http://bigwww.epfl.ch/

1. INTRODUCTION

This chapter is meant as an introduction to the four papers on splines and wavelets that are reprinted here [P1-P4]. We give a general overview the subject, highlighting some of the main uses of splines (and wavelets) in biomedical imaging. We also comment on the appropriateness of these techniques in this particular context.

Both types of representations are useful when one wishes to consider image data as a continuum rather than a discrete array of pixels. Such a continuous modeling of the data is often required in medical imaging. Interpolation, in particular, plays a crucial role at various stages of processing [18, 24]. For instance, it is present—explicitly or not—for tomographic reconstruction, irrespective of the type of algorithm used (filtered backprojection, inverse Fourier or iterative reconstruction). Another important area is medical image visualization; this involves simple 2D operations such as image zooming, panning, rotation, or 3D manipulations, such as reslicing or maximum intensity projection [21], which are often used by radiologists. Interpolation models are also required for performing various types of image registrations [8, 33]; these include intra-modal registration for rigid-body motion compensation [P3], inter-modal registration of CT, PET and MR data sets of a same subject [10, 25], as well as elastic matching for stereotaxic normalization or distortion correction. Considering an image as a continuously–defined function is often desirable for feature extraction, in particular, contour detection. Likewise, wavelets have been found to be well-suited for multi-scale processing, noise filtration of medical images [36], statistical data analysis (fMRI and PET) [31], and feature extraction (e.g., texture).

Almost anyone involved with medical imaging has been using splines, knowingly or nor: the most commonly-used interpolation algorithm—bilinear interpolation—is equivalent to fitting the image with a spline of degree 1. Splines, however, start revealing their true power as one moves to higher degrees for they provide the best cost-performance tradeoff [24]. Since quality is a major concern with medical images, it therefore makes good sense to use them in applications.

Even though splines and wavelets are often treated as separate topics, they are closely related, as will be shown here. We will use a progressive approach to explain the important ideas at a general level, without too many details. Our starting point will be the spline interpolation problem which is discussed in Section 2. Next, we will consider the possibility of coarsening the spline grid, which leads to the powerful idea of multiresolution analysis explained in Section 3. Finally in Section 4, we

will uncover the wavelets which provide a non-redundant decomposition of an image across scale. We will also emphasize the rather special role of splines in wavelet theory and applications.

2. SPLINE INTERPOLATION

A polynomial spline of degree n is a piecewise polynomial function of degree n with pieces that are patched together such as to guarantee the continuity of the function and of its derivatives up to order n-1. Because of this latter smoothness constraint, a spline has only one degree of freedom per polynomial segment (cf. B-spline representation); this is much less than the (n+1) coefficients that are required to represent an unconstrained polynomial of degree n. The price to pay, of course, is that the model fitting process is not local. An additional ingredient that is essential for efficient processing is that the structure of the model be regular; this is the case with the *uniform splines* for which the knot are equally spaced.

Schoenberg, who invented splines, also proved that any uniform spline can be represented unambiguously in terms of a B-spline expansion, which uses shifted B-spline basis functions [20]. This B-spline representation turns out to be most convenient for performing signal and image processing—for a recent overview of the subject, see [27]. The leading idea behind B-spline processing [P1-P2] is that the fitting process (as well as many other spline operations) can be accomplished most efficiently by digital filtering without any need for matrix manipulations.

When working with multi-dimensional data sets like images or volumes, the underlying spline model is usually chosen to be separable. Specifically, one represents the p-dimensional spline function

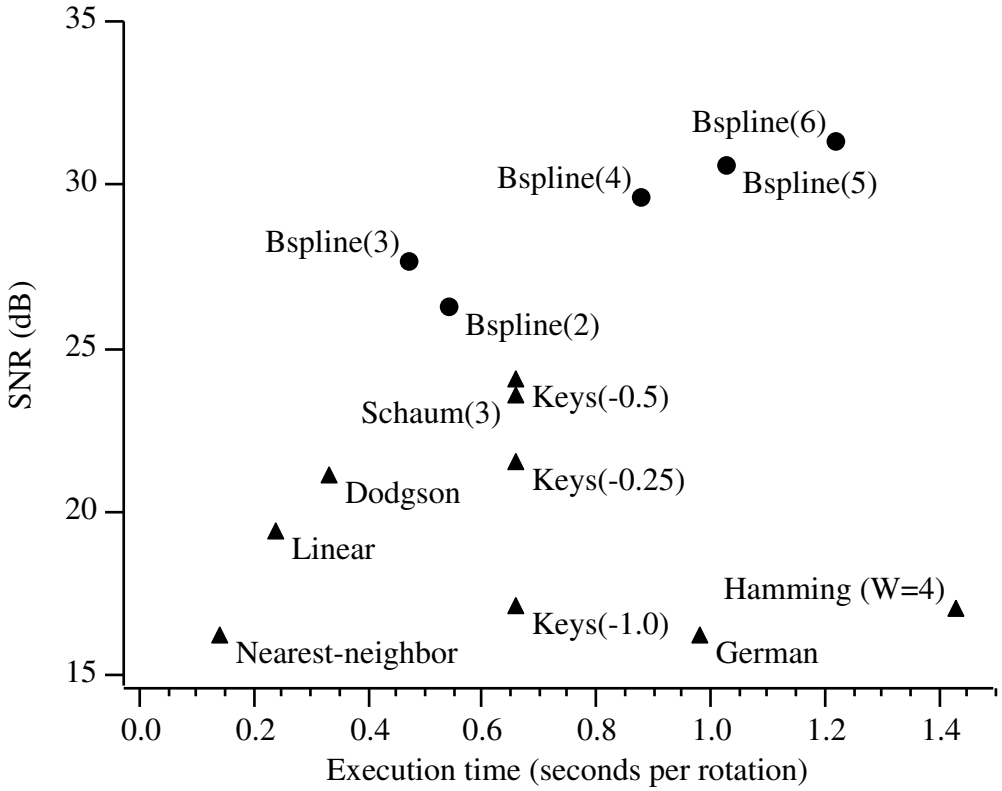

Fig. 1: Cost-performance comparison of various image interpolation algorithm. The experiment involves the application to an image of 15 successive rotation of 24 degrees each.

which has the form of a discrete convolution. Thus, one can determine the $c(\mathbf{k})$'s by deconvolving this equation, which is the key idea behind the algorithms described in [P1-P2]. Since the basis functions are chosen to be separable, the deconvolution process is performed in a separable fashion by successive one-dimensional filtering along the various dimensions of the data. The so-called direct spline filter is implemented recursively as described in [P2].

One remarkable property of the interpolation model (1) is that the underlying interpolator can have infinite support even though it is implemented using finite B-spline basis functions. This is possible only because the $c(\mathbf{k})$'s in (1) are different from the initial pixel values $f(\mathbf{k})$. Thanks to the prefiltering step, the interpolation model is not strictly local anymore for $n \geq 2$, which is one of the reasons why spline interpolators tend to perform better than the conventional techniques that use short kernel interpolating functions. Other more profound reasons for their better performance are found by looking at their approximation theoretic properties [6, 7, 32].

The graph in Fig. 1 shows the results of a cost performance comparison of some commonly used interpolators in medical imaging. We compared the following algorithms: (a) nearest neighbor interpolation (spline of degree 0), (b) bilinear interpolation (spline de degree 1), (c) Dodgson's quadratic interpolator [12], (d) several versions of Keys' cubic interpolator [13]— often considered to be the state-of-the art method in the field [18], (e) Shaum's cubic Lagrange interpolator [19], (f)

windowed sinc interpolation using a Hamming window, and (g) higher-order spline polynomial interpolation using the technique described in [P1-P2]. The experiment, which was designed to magnify distortions, consisted in applying a succession of 15 rotations to some test image. The plot in Fig. 1 displays the final signal-to-noise ratio versus the total CPU time. It is clear from this comparison that splines give the highest quality results and that their cost-benefit ratio is quite favorable. More details on this comparison can be found in [24]. The superiority of splines for interpolation has also been confirmed by other researchers [17, 16].

As a motivation for further reading on splines, we end this section with a listing of their most attractive properties:

- One can always obtain a continuous representation of a discrete signal by fitting it with a spline in one or more dimensions. The fit may be exact (interpolation) or approximate (least-squares or smoothing splines). Spline fits are usually preferable to other forms of representations (e.g., Lagrange polynomial interpolation) because they have a lesser tendency to oscillate (minimum curvature property) [1].
- The primary reason for working with the B-spline representation is that the B-splines are compactly supported. They are the shortest functions with an order of approximation $L=n+1$ [5]. This short support property is a key consideration for computational efficiency. Their simple analytical form also greatly facilitates manipulations [P1-P2].
- Splines are smooth and well-behaved functions (piecewise polynomials). Splines of degree n are $(n-1)$ continuously differentiable. As a result, splines have excellent approximation properties. Precise convergence rates and error estimates are available [6].
- The family of polynomial splines provides design flexibility. By increasing the degree n, we can progressively switch from the simplest piecewise constant ($n=0$) and piecewise linear ($n=1$) representations to the other extreme, which corresponds to a bandlimited signal model ($n \to +\infty$) [3].

3. MULTIRESOLUTION

Other attractive reasons for working with splines are their multiresolution properties. In the context of multi-scale processing, it is advantageous to consider a series of fine-to-coarse spline approximations of the image. The corresponding spline model at resolution $h = 2^i$ is

$$s_h(\mathbf{x}) = \sum_{\mathbf{k} \in Z^D} c_h(\mathbf{k}) \varphi(\mathbf{x} / h - \mathbf{k}). \qquad (4)$$

It has essentially the same form as (1), except that the B-spline basis functions are now enlarged by a factor of h and spaced accordingly. This situation is illustrated in Fig. 2.

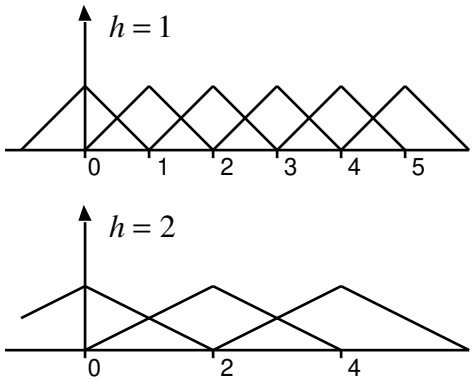

Fig.2: Basis functions for the linear splines at resolutions $h = 2^0$ and $h = 2^1$.

Clearly, for $h > 1$, the model (4) can no longer be exact; however, it can be specified such that the quadratic error with the initial image is minimized.

Fig. 3: Example of a cubic spline image pyramid. Each level is represented by its sample values at the grid point; these are in one-to-one correspondence with the B-spline coefficients.

This type of multiresolution approximation leads to image pyramids. These are typically constructed iteratively by successive reduction by a factor of two in each direction (cf. Fig. 3). The "reduce" operator is obtained by prefiltering the image and re-sampling thereafter—the design of the optimal prefilter is considered in [P1-P2].

Image pyramids are very useful tools for image processing. Their primary application is multi-scale processing where one typically starts by solving a coarse version of the initial problem using reduced size arrays, and then progressively refines the solution as one moves to finer resolutions. This always results in tremendous acceleration of the algorithm because most iterations are performed using very small image copies. Often this also improves robustness because of the smoothing effect of the pyramid which reduces the likelihood of getting trapped in local optima.

Image registration is a good example where this multiresolution strategy is advantageous. A complete description and evaluation of such an algorithm is given in [P3]. The method can handle both rigid-body and affine transformations and it takes advantage of the spline model at all levels. Thanks to the consistent design and to the use of a higher-order interpolation model (cubic splines), it is among

the most precise alignment techniques available and yet fast enough to be competitive. This algorithm has been applied to the registration of a variety of biomedical images in 2D and 3D. These include high-resolution electron micrographs for correlation averaging, and medical images such as PET, SPECT, CT, and MRI. It is also well suited for motion compensation in image sequences of the retina (fundus camera) or fMRI data.

4. WAVELETS

The connection between multiresolution analysis and *wavelets* can be made by considering the error images (or residues) in a *dyadic* image pyramid (cf. Fig. 4a)—"dyadic" means that, at each step, the size of the image is reduced by a factor of two in each dimension. Since the lower resolution approximation was obtained by projecting the image onto some lower resolution subspace, the residual information has less degrees of freedom than the number of displayed error pixels. It is therefore possible to represent these residues in a non-redundant way using wavelet basis functions.

To simplify the discussion, we consider a one-dimensional signal $s_{2^{i-1}}(x)$ at scale $h = 2^{i-1}$, as specified by (4) with $p = 1$, and its approximation $s_{2^i}(x)$ at the next coarser scale $h = 2^i$. The approximation error can thus be written as

$$s_{2^{i-1}}(x) - s_{2^i}(x) = \sum_{k \in Z} d_i(k) \psi_{i,k}(x), \tag{5}$$

where $\psi_{i,k}(x) = 2^{-i/2} \psi(x / 2^i - k)$ are the wavelet basis functions at the resolution i; they are obtained by translating a dilated version (by a factor 2^i) of a basic wavelet template $\psi(x)$. To illustrate the concept, we show in Fig. 4b the wavelet coefficients d_1 (here, $i = 1$) of the error image in Fig. 4a. Note that, in 2D, the wavelet coefficients are organized in three quadrants corresponding to the tensor product wavelets $\varphi(x)\psi(y)$, $\psi(x)\varphi(y)$ and $\psi(x)\psi(y)$, respectively, with φ as in (4). In this way, we have obtained a one-to-one decomposition of the image into the sum of its lower resolution approximation plus the residual information coded by the wavelet coefficients. This decomposition process can be iterated on the lower resolution approximation of the image to yield a multi-level wavelet decomposition, as exemplified in Fig. 4c.

Similarly, one obtains a full wavelet decomposition of an arbitrary function $f(x)$ by carrying out the decomposition over all scales. In one dimension, the full wavelet expansion of f reads

$$f(x) = \sum_{i \in Z} \sum_{k \in Z} d_i(k) \psi_{i,k}(x); \tag{6}$$

it involves an additional summation over scales from $h \to 0$ ($i = -\infty$) to $h \to +\infty$ ($i = +\infty$). Mathematically, the wavelet functions $\{\psi_{i,k}\}_{i \in Z, k \in Z}$ form a basis of L_2 (the space of square-integrable functions), which also means that the wavelet representation is one-to-one. This can also be seen with the wavelet decomposition in Fig. 4c; it has as many coefficients as there are pixels in the image in Fig. 4a and is fully reversible. Usually, the wavelet transform will exhibit many more zero (or small coefficient) values as the original pixel representation of the image. This can be used advantageously for image coding [22], as well as for threshold-based noise reduction [35].

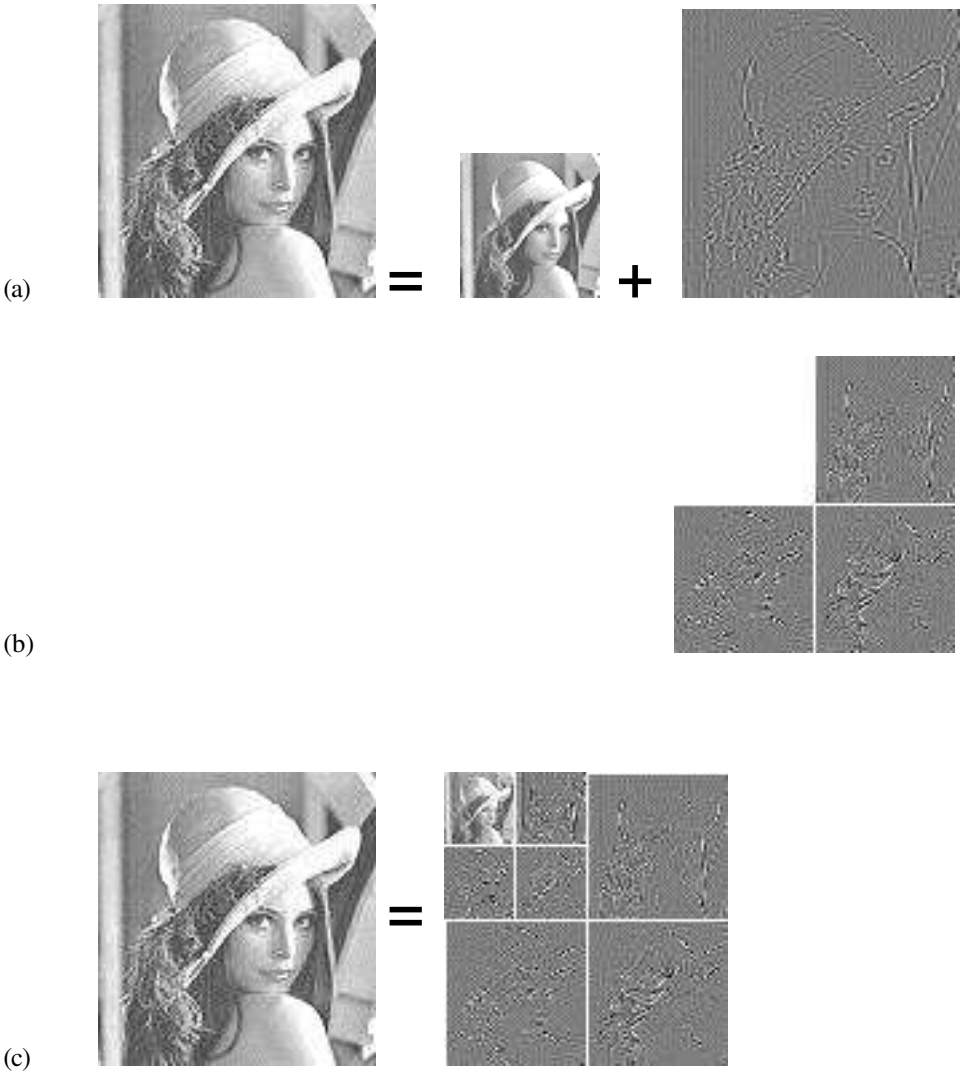

Fig. 4: From pyramids to wavelets. (a) The initial image is decomposed into its lower resolution approximation plus a residual. (b) Instead of the pixel-based representation of the residual, it is more advantageous to use a non-redundant representation in terms of wavelet basis functions—in 2D, there are 3 tensor-product wavelets, each corresponding to one quadrant. (c) The process may be iterated to yield a multi-level wavelet decomposition.

The choice of the wavelet in (5) (or (6)) is obviously quite constrained, even though the solution is not unique. In the case of polynomial splines, the wavelets, which are splines as well, have been characterized explicitly [29]. The most prominent examples are the Battle-Lemarié wavelets [4, 14], which are orthogonal, and the compactly supported B-spline wavelets [9, 28]—the wavelet analogs of the classical B-splines.

There are many applications of wavelets in biomedical imaging, most of which are reviewed in [P4]. The primary ones are medical image compression, tomographic reconstruction, image enhancement of mammograms, noise reduction in a variety of medical images including MRI, and the

statistical analysis of functional images of the brain (PET and fMRI). Some detailed algorithm descriptions can be found in [2].

Various types and brands of wavelets have been described in the literature [11, 23, 34, 15], many of which are not splines. It is important to note, however, that each wavelet contains a spline convolution factor (associated with the so-called regularity factor) which is fully responsible for the more involved mathematical properties of wavelets (vanishing moments, order of approximation, and regularity). This is the reason why splines play a very special role in wavelet theory [23]. Moreover, spline wavelets have many features that are particularly relevant and desirable for medical imaging [26] and that make them stand apart from other the wavelets available. Their most important properties are:

- *Closed-form representation*. This is a rather unique property; all other wavelet bases are defined indirectly through an infinite recursion.
- *Simplicity of manipulation*. This is especially true for mathematical operations such as differentiation.
- *Symmetry*. Symmetry (resp. anti-symmetry) is crucial for feature detection; without it, the zero-crossing (resp. maximum) locations are shifted with respect to the corresponding image feature.
- *Shortest and most regular scaling function of order* L. The order and degree of differentiability (regularity) of the basis functions are key mathematical parameters. Having the shortest functions reduces the amount of computation.
- *m-scale relation.* B-splines satisfy a two scales relation for any integer m, and not just for powers of two as is usually the case in wavelet theory. This turns out to be advantageous for designing fast, non-dyadic wavelet algorithms [30].
- *Best approximation properties.* For a given order L, splines tend to give smaller scale-truncated approximation errors than other standard wavelet families. Splines have been shown to be "π-times" better (in some precise approximation theoretic sense) than Daubechies wavelets [6].
- *Optimal time-frequency localization.* B-spline wavelets have been shown to converge to modulated Gaussians [28]. They therefore offer close-to-optimal time-frequency localization in the sense specified by the uncertainty principle. This makes them the atoms of choice for time-frequency (or space-frequency) analysis.

Reprinted papers

[P1] M. Unser, A. Aldroubi and M. Eden, "B-spline signal processing: Part I—theory," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 821-833, 1993.

[P2] M. Unser, A. Aldroubi and M. Eden, "B-spline signal processing: Part II—efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 834-848, 1993.

[P3] P. Thévenaz, U.E. Ruttimann and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Processing*, vol. 7, no. 1, pp. 27-41, 1998.

[P4] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 626-638, 1996.

Additional references

[1] J.H. Ahlberg, E.N. Nilson and J.L. Walsh, *The Theory of splines and their applications*. New York: Academic Press, 1967.

[2] A. Aldroubi and M. Unser, *Wavelets in Medicine and Biology*. Boca Raton, FL: CRC, 1996.

[3] A. Aldroubi, M. Unser and M. Eden, "Cardinal spline filters: Stability and convergence to the ideal sinc interpolator," *Signal Processing*, vol. 28, no. 2, pp. 127-138, 1992.

[4] G. Battle, "A block spin construction of ondelettes. Part I: Lemarié functions," *Commun. Math. Phys.*, vol. 110, pp. 601-615, 1987.

[5] T. Blu and M. Unser, "Minimum support interpolators with optimum approximation properties," in Proc. *Int. Conf. Image Processing*, Santa Barbara, CA, 1998, vol. III, pp. 242-245.

[6] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques: Part II—wavelets," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2796-2806, 1999.

[7] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques: Part I—interpolators and projectors," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2783-2795, 1999.

[8] L.G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.

[9] C.K. Chui and J.Z. Wang, "On compactly supported spline wavelets and a duality principle," *Trans. Amer. Math. Soc.*, vol. 330, no. 2, pp. 903-915, 1992.

[10] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens and G. Marchal, "Automated multi-modality image registration based on information theory," in *Information Processing in Medical Imaging*, Y. Bizais, Ed., pp. 263–274, 1995.

[11] I. Daubechies, *Ten lectures on wavelets.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[12] N.A. Dodgson, "Quadratic interpolation for image resampling," *IEEE Trans. Image Processing*, vol. 6, no. 9, pp. 1322-1326, 1997.

[13] R.G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 6, pp. 1153-1160, 1981.

[14] P.-G. Lemarié, "Ondelettes à localisation exponentielle," *J. Math. pures et appl.*, vol. 67, no. 3, pp. 227-236, 1988.

[15] S. Mallat, *A wavelet tour of signal processing*. San Diego: Academic Press, 1998.

[16] E.H.W. Meijering, W.J. Niessen and M.A. Viergever, "Piecewise polynomial kernels for image interpolation: A generalization of cubic convolution," in Proc. *Int. Conference on Image Processing*, Kobe, Japan, 1999,

[17] J.L. Ostuni, A.K.S. Santha, V.S. Mattay, D.R. Weinberger, R.L. Levin and J.A. Frank, "Analysis of interpolation effects in the reslicing of functional MR-images," *J. Computer Assisted Tomography*, vol. 21, no. 5, pp. 803-810, 1997.

[18] J.A. Parker, R.V. Kenyon and D.E. Troxel, "Comparison of interpolating methods for image resampling," *IEEE Trans. Med. Imaging*, vol. MI-2, no. 1, pp. 31-39, 1983.

[19] A. Schaum, "Theory and design of local interpolators," *CVGIP: Graphical Models and Image Processing*, vol. 55, pp. 464-481, 1993.

[20] I.J. Schoenberg, "Contribution to the problem of approximation of equidistant data by analytic functions," *Quart. Appl. Math.*, vol. 4, pp. 45-99, 112-141, 1946.

[21] S. Schreiner, C.B. Paschal and R.L. Galloway, "Comparison of projection algorithms used for the construction of maximum intensity projection images," *J. Computer Assisted Tomography*, vol. 20, no. 1, pp. 56-67, 1996.

[22] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 41, no. 12, pp. 3445-3462, 1993.

[23] G. Strang and T. Nguyen, *Wavelets and filter banks*. Wellesley, MA: Wellesley-Cambridge, 1996.

[24] P. Thévenaz, T. Blu and M. Unser, "Image interpolation and resampling," in *Handbook of Medical Image Processing*, in press.

[25] P. Thévenaz and M. Unser, "Spline pyramids for inter-modal image registration using mutual information," in Proc. *Wavelet Applications in Signal and Image Processing V*, San Diego, CA, September 16-19, 1997, vol. Proc. SPIE, Vol. 3169, pp. 236-247.

[26] M. Unser, "Ten good reasons for using spline wavelets," in Proc. *SPIE Conf. Wavelet Applications in Signal and Image Processing V*, San Diego, CA, August 6-9, 1997, vol. Proc. SPIE 3169, pp. 422-431.

[27] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22-38, 1999.

[28] M. Unser, A. Aldroubi and M. Eden, "On the asymptotic convergence of B-spline wavelets to Gabor functions," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 864-872, 1992.

[29] M. Unser, A. Aldroubi and M. Eden, "A family of polynomial spline wavelet transforms," *Signal Processing*, vol. 30, no. 2, pp. 141-162, 1993.

[30] M. Unser, A. Aldroubi and S.J. Schiff, "Fast implementation of the continuous wavelet transform with integer scales," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3519-3523, 1994.

[31] M. Unser, P. Thévenaz, C. Lee and U.E. Ruttimann, "Registration and statistical analysis of PET images using the wavelet transform," *IEEE Engineering in Medicine and Biology*, vol. 14, no. 5, pp. 603-611, 1995.

[32] M. Unser, P. Thévenaz and L. Yaroslavsky, "Convolution-based interpolation for fast, high-quality rotation of images," *IEEE Trans. Image Processing*, vol. 4, no. 10, pp. 1371-1381, 1995.

[33] P.A. van den Elsen, E.-J.D. Pol and M.A. Viergever, "Medical image matching: A review with classification," *IEEE Engineering in Medicine and Biology*, vol. 12, no. 1, pp. 26-39, 1993.

[34] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[35] J.B. Weaver, X. Yansun, D.M. Healy, Jr. and L.D. Cromwell, "Filtering noise from images with wavelet transforms," *Magnetic Resonance in Medicine*, vol. 21, no. 2, pp. 288-95, 1991.

[36] J.B. Weaver, X. Yansun, D.M. Healy and J.R. Driscoll, "Wavelet-encoded MR imaging," *Magnetic Resonance in Medicine*, vol. 24, no. 2, pp. 275-87, 1992.