ℓ_p-Multiresolution Analysis: How to Reduce Ringing and Sparsify the Error

Arrate Muñoz, *Student Member, IEEE*, Thierry Blu, *Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—We propose to design the reduction operator of an image pyramid so as to minimize the approximation error in the ℓ_p-sense (not restricted to the usual $p = 2$), where p can take non-integer values. The underlying image model is specified using shift-invariant basis functions, such as B-splines. The solution is well-defined and determined by an iterative optimization algorithm based on digital filtering. Its convergence is accelerated by the use of first and second order derivatives. For p close to 1, we show that the ringing is reduced and that the histogram of the detail image is sparse as compared with the standard case, where $p = 2$.**

***Index Terms*—Banach spaces, multiresolution, non-Euclidean norms, splines.**

I. Introduction

MULTIRESOLUTION analysis is a simple yet very powerful concept which goes back to the pioneering works of Rosenfeld [1] and Burt and Adelson [2]. Instead of a fixed size pixel array, one considers a hierarchical image description at multiple resolution levels; typically, a series of fine-to-coarse approximations which are stored in a pyramid data structure. Such pyramids are extremely useful for speeding up computations. In fact, there are multiscale versions of most image processing algorithms. The main advantages of multiscale processing are the following.

- *Computational speed:* Since there are much fewer pixels at the coarser levels of the pyramid, iterative algorithms that switch between resolution levels require less computation and have faster convergence.
- *Spatial resolution adaptation:* Many image processing algorithms operate on very localized neighborhoods and it makes good sense to adapt the resolution in an optimal fashion. This is especially true with iterative schemes which proceed by successive refinement—here the resolution should be linked to the step size of the algorithm.
- *Increased robustness:* In the context of iterative algorithms, the smoothing effect of the pyramid reduces the likelihood of getting trapped in local extrema.
- Analogies can be made with the hierarchical organization of the human primary visual cortex.

Manuscript received July 19, 2001; revised February 14, 2002. This work was supported by the Swiss National Science Foundation under Grant 2100-053540. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nasser Kehtarnavaz.

The authors are with the Biomedical Imaging Group, STI/IOA BM, Swiss Federal Institute of Technology Lausanne (EPFL), CH-1015 Lausanne EPFL, Switzerland (e-mail: arrate.munoz@epfl.ch; thierry.blu@epfl.ch; michael.unser@epfl.ch).

Publisher Item Identifier S 1057-7149(02)04799-1.

Multiresolution analysis also plays a central role in the theory of the wavelet transform, which provides a nonredundant representation of images across scales. Here too, the applications in image processing and computer vision are numerous, and often very successful [3]–[5].

One of the earliest and most popular examples of pyramid is due to Burt and Adelson [2]. Their Gaussian filtering, however, produces excessive smoothing, which leads to some loss of image details. Higher-quality image approximation can be obtained by designing a reduction filter that is optimum in the least-squares sense, or by using the lowpass branch of a wavelet decomposition algorithm [6], [7]. Another option is to use spline pyramids that minimize either the ℓ_2 or the L_2-approximation error [8], [9]. These latter representations are especially attractive for continuous/discrete multiscale processing. The nice feature of these pyramids is that they can all be implemented using a combination of filters and sampling rate converters. Of course, the critical aspect here is filter design—a standard requirement is the biorthogonality of the reduction and expansion operators [10]. Unfortunately, simplicity also comes at a price and these pyramids suffer from limitations that are inherent to linear methods; in particular, edge blurring (when the smoothing is too strong, e.g., the Gaussian pyramid), aliasing (when it is not enough), and ringing artifacts (when the filters have a sharp cutoff). Thus, the selection of a suitable multiresolution model is essentially a question of compromise: higher order spline or wavelet approximations generally yield better energy compaction but they also give rise to larger Gibbs oscillations as the functions become more and more bandlimited [11].

An attractive alternative to linear pyramids is to go nonlinear. Several authors have proposed to replace the linear pyramid filters by nonlinear ones including the median and morphological operators [12]–[16]. However, these so-called morphological pyramids are generally not meant to provide a continuous/discrete representation. Nonlinear filters can also introduce distortions that make the reduced images visually unpleasant.

In this paper, we will pursue another approach and introduce spline pyramids that are optimal for ℓ_p-norms. Note that the choice of a spline model in this context is equivalent to specifying the expansion mechanism, i.e., polynomial spline interpolation. Thus, the challenge is to come up with a corresponding reduction operator that produces visually pleasant results without aliasing and with minimal ringing artifacts. The information lost by the reduction operator should be as little as possible. Normally, this corresponds to difference images which are sparse—contain a lot of small-valued elements.

The paper is organized as follows. In Section II, we present a Banach-space formulation of the problem and prove that the

1057-7149/02$17.00 © 2002 IEEE

solution is well defined. In Section III, we propose a digital filtering-based procedure that computes the solution iteratively. The computational overhead of the iterative approach is small, and we expect the generation of the pyramid to account for only a very small part of the total effort in a typical multiscale algorithm. In Sections IV and V, we compare the approximations obtained for different ps and orders of the approximation functions, respectively. We end with a discussion of our results in Section VI.

II. Multiresolution Subspaces of ℓ_p

In this section, we present the theoretical basis of our method. We justify the choice of our continuous/discrete model and show that our approximation problem has a well-defined solution.

A. Definitions and Notation

The $\|\cdot\|_{\ell_p}$ norm of a sequence $c = \{c_k\}_{k\in\mathbb{Z}}$ is defined as

$$\|c\|_{\ell_p} = \left(\sum_{k\in\mathbb{Z}} |c_k|^p\right)^{1/p} \tag{1}$$

with $1 \le p < \infty$ and the special case $\|c\|_{\ell_\infty} = \max_{k\in\mathbb{Z}} |c_k|$.

The z-transform of a signal $s(k)$, $k \in \mathbb{Z}$ is denoted by

$$S(z) = \sum_{k\in\mathbb{Z}} s(k) z^{-k}.$$

If we make $z = e^{j2\pi f}$, we recover the Fourier transform.

The symbol $\downarrow N$ denotes the downsampling operator by the integer factor N; it is defined as

$$s_{\downarrow N}(k) = s(Nk), \qquad \forall k \in \mathbb{Z}.$$

The dual operator $\uparrow N$ represents upsampling by the integer factor N

$$s_{\uparrow N}(k) = \begin{cases} s\left(\dfrac{k}{N}\right), & \text{if } N \text{ divides } k, \\ 0, & \text{elsewhere.} \end{cases}$$

B. Approximation Signal Model

Our signal model [cf. (2)] is similar to the ones encountered in wavelet theory. For simplicity, we will present the theory in 1-D. The extension to multiple dimensions is straightforward through the use of tensor product basis functions. The use of a separable model implies that the expansion mechanism is separable as well; the reduction mechanism, on the other hand, will not be separable unless we are dealing with the classical case $p = 2$ (least squares approximation).

Specifically, we choose to represent all signals in terms of shifted basis functions, which are typically sampled B-splines. A discrete signal, e.g., $s(k)$, will always denote the samples on the finest grid. Its coarser level approximation $\tilde{s}(k)$ at resolution N will use basis functions $\varphi_l(k)$ that are a translated version of some template h_N: $\varphi_l(k) = h_N(k - Nl)$

$$\tilde{s}(k) = \sum_{l\in\mathbb{Z}} c_l h_N(k - Nl) = [c]_{\uparrow N} * h_N(k). \tag{2}$$

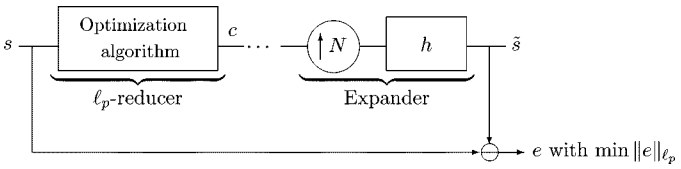

Fig. 1. Reduction/expansion system for an integer scaling factor N. Reduction: the signal is reduced by a factor N so as to minimize the ℓ_p-norm of the error. Expander: upsampling and filtering, as specified by the approximation model [see (2)].

In other words, $\tilde{s}(k)$ is entirely specified by its coefficients c_l. These are the quantities that are stored in the pyramid; at each level there is exactly one such number per node. The template h_N should be interpreted as an expansion (or interpolation) filter which maps the coefficients sequence c_l (coarse level of the pyramid) to the finest resolution level on which the signals are defined (cf. the right hand side of Fig. 1). We are using the subscript N in h_N to indicate that the expansion filter depends on N (typically, a spline interpolator with an expansion factor N). In the sequel, we will sometimes leave out this dependence to simplify the notation. The corresponding approximation space is

$$V_N = \left\{ \tilde{s}(k) = \sum_{l\in\mathbb{Z}} c_l h_N(k - Nl) \colon c \in \ell_p \right\}. \tag{3}$$

It is clearly convex and N-integer shift-invariant, i.e., $s(k) \in V_N$ iff $s(k + N) \in V_N$.

For our formulation, it is essential that V_N be a closed subspace of ℓ_p to ensure a well-defined solution of our approximation problem. This will be the case if $\{h(k - Nl)\}_{k\in\mathbb{Z}}$ forms a p-stable (or p-Riesz) basis (for a similar definition in the continuous L_p framework cf. [17])

$$\forall c \in \ell_p, \quad A \cdot \|c\|_{\ell_p} \le \left\| \sum_{l\in\mathbb{Z}} c_l h(k - Nl) \right\|_{\ell_p} \le B \cdot \|c\|_{\ell_p} \tag{4}$$

with $0 < A, B < \infty$.

This norm equivalence implies that $h \in \ell_p$ (by letting $c_l = \delta_l$) and that ℓ_p and V_N are isomorphic Banach spaces.

This above condition is ensured by the following theorem that is proven in the Appendix.

Theorem 1: If $h \in \ell_1$ and $\{h(k - lN)\}_{l\in\mathbb{Z}}$ is a p_0-Riesz basis for some $1 \le p_0 \le \infty$ then it is also a p-Riesz basis for $1 \le p \le \infty$.

Consequently, if h is in ℓ_1 and generates a Riesz basis in the conventional ℓ_2-sense, then it is automatically also p-stable for any p. The following result by Aldroubi *et al.* [18] gives a simple way to check if $h \in l_1$ generates a Riesz basis or not.

Theorem 2: $\{h(k - lN)\}_{l\in\mathbb{Z}}$ is a ℓ_2-Riesz basis if and only if

$$0 < \alpha \le \sum_{i=0}^{N-1} \left| H\left(e^{j2\pi(f-i)/N}\right) \right|^2 \le \beta < +\infty.$$

C. Discrete/Continuous Multiresolution; B-Splines as Basis Functions

The filter $h = h_N$ is obtained by sampling at the integers a continuous basis function $\varphi(x) \in L_2(\mathbb{R})$ dilated by N, i.e., $h_k = \varphi(k/N)$. We make this choice because we want the interpolated version of our approximated signal to belong to the space $\text{span}\{\varphi((x/N)-l)\}_{l\in\mathbb{Z}}$, i.e., $\tilde{s}(x) = \sum_{l\in\mathbb{Z}} c_l\varphi((x/N)-l)$, while its samples $\tilde{s}(k)$ belong to $\text{span}\{h(k-Nl)\}_{l\in\mathbb{Z}}$ [cf. (3)]. This yields a consistent discrete/continuous signal representation. The advantage of this joint model is the possibility of applying continuously defined operators commonly used in image processing such as derivatives or geometrical transformations. B-splines are examples of continuous basis functions that we consider appropriate for this multiresolution analysis. The main reasons for this choice are their maximal order of approximation for a given support (key consideration for computational efficiency) [19]; splines are also smooth and well-behaved (piecewise polynomials) and their simple analytic form facilitates their manipulation [20]. In addition, they satisfy a two-scale relation which makes them appropriate for multiscale processing [8]. Finally, as the coefficients of the filter h_k for the B-splines belong to ℓ_1 and generate a Riesz basis of ℓ_2 [21], they satisfy the conditions on h to have V_N be a closed subspace of ℓ_p (see Section III-B).

D. Projection Theorem in Banach Spaces

Given the discrete signal $s \in \ell_p$, we would like to find the approximation $\tilde{s} \in V_N \subset \ell_p$ that minimizes the error $\|s-\tilde{s}\|_{\ell_p}$. The projection theorem in Banach spaces (i.e., ℓ_p) [22] states that, since V_N is a convex closed subspace of ℓ_p for any $s \in \ell_p$, there exists $\tilde{s} \in V_N$ such that

$$\|s-\tilde{s}\|_{\ell_p} = d(s, V_N) = \inf_{s_V\in V_N} \|s-s_V\|_{\ell_p}. \tag{5}$$

Thus, $\tilde{s}$ is the best approximation of s in V_N, in the ℓ_p-sense. We denote $\tilde{s} = P_{V_N}s$. For $1 < p < \infty$, $\tilde{s}$ is unique. Unicity is lost for $p = 1$ and $p = \infty$; nevertheless, all the minima are global, ensuring that the solution to our approximation problem is well-defined. Thus, our initial problem of calculating the minimum error approximation translates into calculating the coefficients c_l in (2) that describe the projection $\tilde{s}$.

III. Optimal Approximation

In this section, we refer to the state-of-the-art algorithm to calculate ℓ_2-projections. We then turn to more general ℓ_p-projections and present a novel iterative algorithm based on digital filtering.

A. Optimal Approximation in ℓ_2

We will start by presenting the solution of Aldroubi *et al.* [18]. For $p = 2$, our space is a Hilbert space, i.e., a Banach space with an inner product. In that case, the calculation of $P_{V_N}s$ takes the simpler form

$$\tilde{s}(k) = P_{V_N}s(k) = \sum_{l\in\mathbb{Z}} \left\langle s(.), \mathring{h}(k-Nl)\right\rangle h(k-Nl) \tag{6}$$

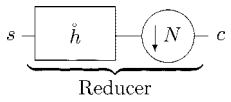

Fig. 2. Optimal reducer for ℓ_2-norms: antialiasing filter and downsampler.

where $\mathring{h}$ is the (unique) dual function of h, i.e., $\mathring{h} \in V_N$ and $\langle\mathring{h}(k-Nl), h(n-Nl)\rangle = \delta(k-n)$ (biorthonormality).

The coefficients c_l of the orthogonal projection of the input signal $s \in \ell_1$ onto V_N are given by

$$c_l = \left[s * \mathring{h}^T\right]_{\downarrow N}(l) \tag{7}$$

where

$$\mathring{H}(z) = \frac{NH(z^{-1})}{\sum_{k=0}^{N-1} H\left(e^{j(2\pi k/N)}z\right) H\left(e^{j(2\pi k/N)}z^{-1}\right)}.$$

The corresponding reduction/expansion digital filtering system is shown in Fig. 1. In this particular case, the reduction is implemented via a prefilter $\mathring{h}$ followed by a downsampler as shown in Fig. 2. Note that the Riesz condition ensures that the filter $\mathring{h}$ exists and is always well-defined.

B. Optimal Approximation in ℓ_p

Now, we deal with the general case of finding an optimal ℓ_p-approximation. The difficulty of working in Banach spaces is the lack of an inner product. Practically, this means that the solution cannot be computed by a one step linear algorithm. In this section, we develop an iterative optimization procedure that takes advantage of linear filtering and of the calculation of first and second order derivatives.

The ℓ_p-norm of the approximation error $e = s - \tilde{s}$ is a convex function of the coefficients c_l, which ensures that its local minima are also global. This is because of the constitutive definition of a norm (esp., triangle inequality and semilinearity) and because the error e depends linearly on c. The consequence is that a gradient-based optimization algorithm with adaptive steps will always converge to the global minimum. However, since for $p = 1$ the norm of the error e is only piecewise differentiable, we must be prepared to encounter some difficulties (slower convergence) as p gets close to 1.

1) Theoretical Derivation of the Optimization Algorithm: To speed up convergence, we propose a robust optimization algorithm, the formulation of which is Hessian-based. The idea behind it is to optimize the coefficients c_l each in turn and to express the norm of the error as a second order polynomial which is easily minimized. The update formula for the vector of coefficients is then derived.

Mathematically, we justify our algorithm as follows: If we fix an index l_0 in the expression for $\tilde{s}$, we have

$$\begin{aligned}\tilde{s}(k) &= \sum_{l\in\mathbb{Z}} c_l h(k-Nl) \\ &= c_{l_0}h(k-Nl_0) + \underbrace{\sum_{l\neq l_0} c_l h(k-Nl)}_{\eta(k)}.\end{aligned} \tag{8}$$

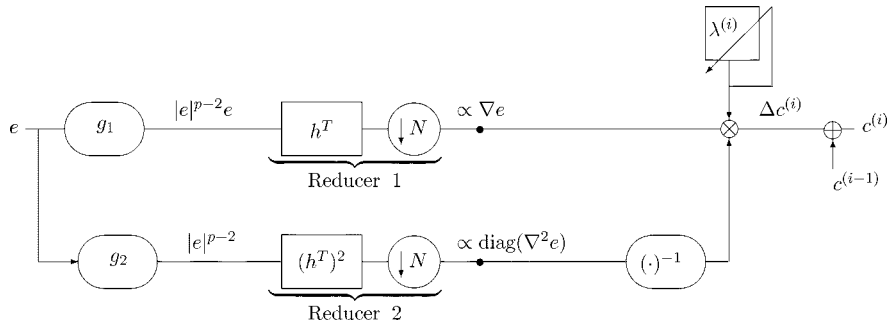

Fig. 3. Optimal reducer for ℓ_p-norms: Reducer 1: Gradient estimation, Reducer 2: Diagonal of the Hessian estimation. The value of λ that minimizes the error at each iteration is calculated using a line search algorithm. The filters are reversed versions of the original ones: $h^T(k) = h(-k)$.

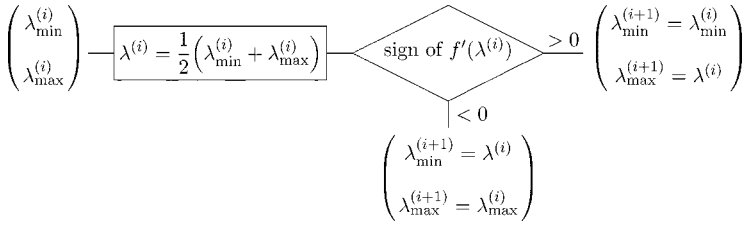

Fig. 4. One iteration of the algorithm that evaluates λ optimal.

Then, we rewrite the norm of the error as

$$\begin{aligned}\|e\|_{\ell_p}^p &= \|s - \tilde{s}\|_{\ell_p}^p \\ &= \sum_{k\in\mathbb{Z}} |e(k)|^{p-2}(s(k) - c_{l_0}h(k - Nl_0) - \eta(k))^2. \quad (9)\end{aligned}$$

We minimize the last expression as a function of c_{l_0} considering $|e(k)|^{p-2}$ as independent of c_{l_0}, in order to get the update formula. Thus, if we know $c^{(i)} = (\ldots, c_{-1}^{(i)}, c_0^{(i)}, \ldots c_{l_0}^{(i)}, \ldots)$, we obtain the update vector of coefficients $\Delta c^{(i+1)}$ by calculating for each index l_0

$$\Delta c_{l_0}^{(i)} = c_{l_0}^{(i+1)} - c_{l_0}^{(i)} = -\frac{\sum_{k\in\mathbb{Z}} |e(k)|^{p-2}e(k)h(k - Nl_0)}{\sum_{k\in\mathbb{Z}} |e(k)|^{p-2}h^2(k - Nl_0)} \quad (10)$$

and then $\tilde{s}^{(i+1)} = \tilde{s}^{(i)} + \sum_{l\in\mathbb{Z}} \Delta c_l^{(i)} h(k - Nl)$.

We now show that this algorithm can also be interpreted as a gradient-based or quasi-Newton search procedure. The partial derivative of the norm of the error $\|e\|_{\ell_p}^p$ with respect to c_l is

$$\frac{\partial \|e\|_{\ell_p}^p}{\partial c_l} = -\sum_{k\in\mathbb{Z}} g_1(e(k))h(k - Nl)$$

with $g_1(x) = p|x|^{p-2}x$. The second order partial derivative is

$$\frac{\partial^2 \|e\|_{\ell_p}^p}{\partial c_l \partial c_n} = \sum_{k\in\mathbb{Z}} g_2(e(k))h(k - Nl)h(k - Nn)$$

with $g_2(x) = p(p-1)|x|^{p-2}$; these define the entries of the (infinite dimensional) Hessian matrix H.

The update formula for the usual Hessian algorithm [23] takes the form $\Delta c^{(i)} = -H^{-1}\nabla e$ where ∇e is the gradient (vector of partial derivatives) and H is the Hessian (matrix of second order partial derivatives). Here, the Hessian matrix is essentially diagonal dominant because $h(k)$ is decaying away from the origin. It is therefore legitimate to use the following simplified update formula

$$\Delta c^{(i)} = -\lambda(\mathrm{diag}\, H)^{-1}\nabla e \quad (11)$$

where we have also introduced a step size λ.

If we make $\lambda = p - 1$, we have the equivalence with formula (10). In the following of the paper, we will call "Hessian fixed" the algorithm described by (11) with $\lambda = p - 1$.

We will see that the advantage of the Hessian over the gradient-based methods is its efficiency, especially when p gets close to 1. It costs slightly more per iteration because the diagonal of the Hessian has to be evaluated in addition to the gradient, but we will show how to compute it efficiently, using filtering and downsampling.

2) Implementation of the Optimization Algorithm: We describe now the modular structure of the optimization algorithm designed to calculate the coefficients c_l of the ℓ_p-approximation signal. The implementation uses two reduce operations (Fig. 3) followed by an expander (Fig. 1). The update vector $\Delta c^{(i)}$ is obtained from the error in three steps. First, gradient estimation (Fig. 3—upper branch), then, inverse of the diagonal of the Hessian estimation (Fig. 3—lower branch). Those are finally combined and multiplied by the step size λ to provide the update vector $\Delta c^{(i)}$. The diagram of Fig. 1 shows how to recompute the error at the given iteration.

The value of the step size λ in (11) can be made optimal in the sense of minimizing the error as much as possible at each step. The idea is to remark that we are minimizing

$$f(\lambda) = \left\| e^{(i)} - \lambda \sum_k u(l)h(k - Nl) \right\|_{\ell_p} \quad (12)$$

with $u = (\mathrm{diag}\, H)^{-1}\nabla e$ according to (11). In practice, we estimate an upper ($\lambda_{\max}$) and lower ($\lambda_{\min}$) bound for λ. This

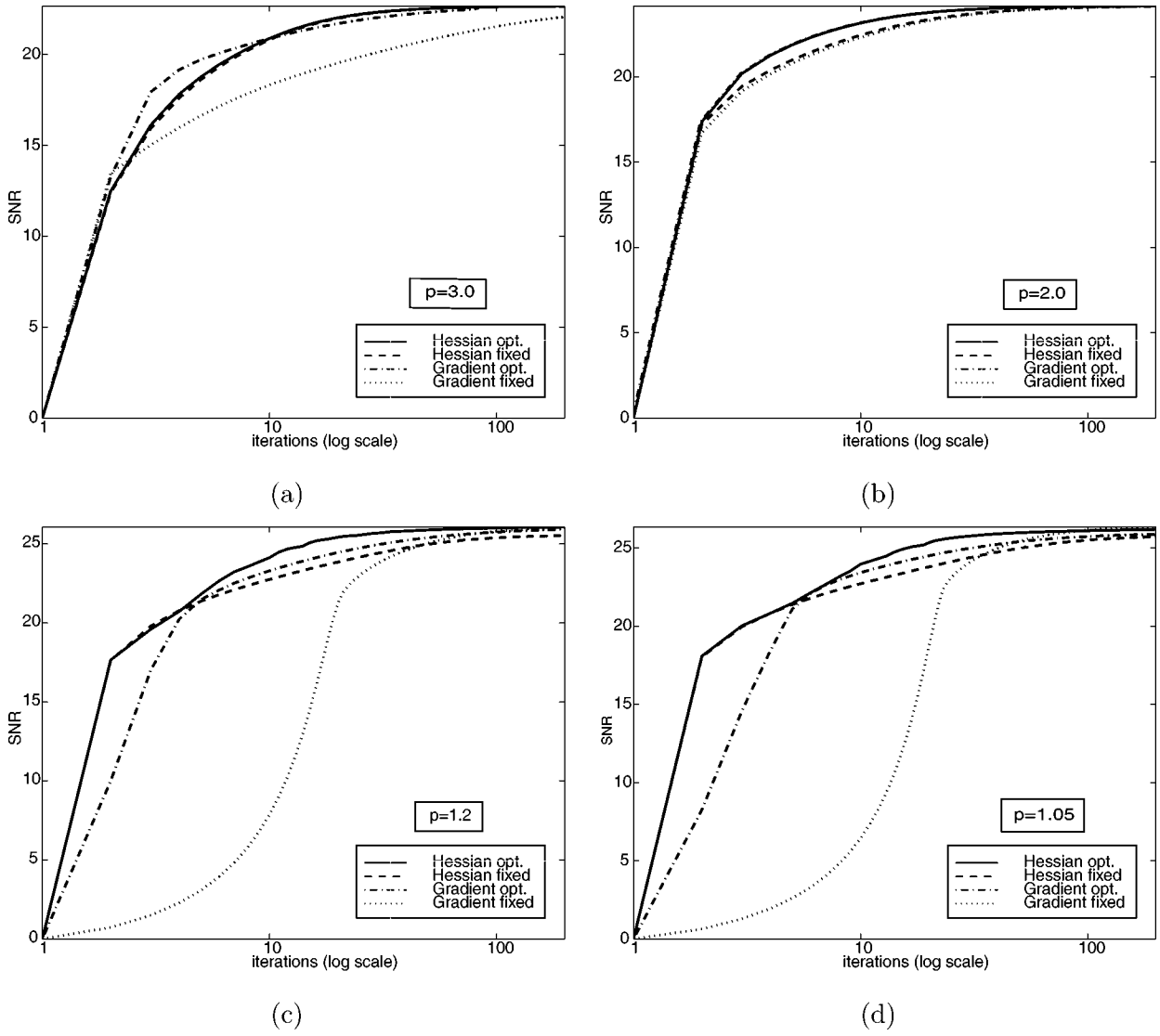

Fig. 5. Convergence of the algorithm for different values of p. Each graph illustrates the decrease of the criterium (i.e., the increase of the SNR_{ℓ_p}) as a function of the number of iterations. The axes are logarithmic. Results for: (a) $p = 3.0$ and $\lambda_{\mathrm{fixed}} = 0.000\,45$, (b) $p = 2.0$ and $\lambda_{\mathrm{fixed}} = 0.12$, (c) $p = 1.2$ and $\lambda_{\mathrm{fixed}} = 1.0$, and (d) $p = 1.05$ and $\lambda_{\mathrm{fixed}} = 2.0$.

value is optimized by using a line search algorithm that reduces by two the length of the interval $[\lambda_{\min}, \lambda_{\max}]$ at each step. Fig. 4 describes one iteration of the algorithm. The search for the optimal λ is made acceptable in terms of computational overhead by choosing initial bounds close to the optimal, i.e., those calculated in the preceding iteration. If $f'(\lambda_{\min}^{(0)}) < 0$ and $f'(\lambda_{\max}^{(0)}) > 0$ then the convergence is ensured by the convexity of $f(\lambda)$. Note that for $p = 2$ we have an exact formula to calculate λ optimal that amounts to minimizing a second order polynomial. We observe from Tables I and II that the line search algorithm requires approximately 10–14 iterations when $p < 2$ in order to yield an optimal step size λ. The test image is the one in Fig. 6—top. Cubic splines are chosen for the interpolation and the scale is reduced by a factor of two.

Fig. 5 represents a typical example of convergence of the algorithm for the minimization of the ℓ_p-approximation error. The test signal and the parameters are the same as for the example given above. Each graph shows the decrease of the criterion [increase of the $\mathrm{SNR}_{\ell_p} = -20\log(\|s-\tilde{s}\|_{\ell_p}/\|s\|_{\ell_p})$ dB], measured as a function of the number of iterations for different values of p. The results are shown in decibels. Four variants of our algorithm are compared: Either gradient or Hessian-based with the parameter λ calculated in an optimal fashion; gradient-based with λ fixed; Hessian-based as given by (11) with $\lambda = p-1$. In all cases, our initial guess was $c^{(0)} = 0$. We observe in Fig. 5(a) that for high values of p ($p = 3.0$ in this case), the performances of the Hessian-based and gradient optimal algorithms are very similar. The convergence of the gradient-based algorithms with λ fixed is worse. Fig. 5(b) is a special case as we deal with the convergence of the least-squares approximations. Here, the diagonal of the Hessian is constant and independent of the input, which implies that the Hessian and gradient-based algorithms that use the same strategy for determining λ are equivalent. On the other hand, we observe a slightly faster increase of the $\ell_p - \mathrm{SNR}$ for the algorithms with λ optimal over the other ones. Fig. 5(c) and (d) demonstrate the behavior of the algorithms with ps close to 1 ($p = 1.2$ and $p = 1.05$, respectively). Here, the algorithms with λ optimal converges in less iterations than

TABLE I
GRADIENT ALGORITHM: AVERAGE (AND STANDARD DEVIATION) NUMBER OF ITERATIONS OF THE LINE SEARCH ALGORITHM (FIG. 4)

Gradient	mean $\pm$ stdv
ℓ_3-optimal	3.36 ± 1.82
ℓ_2-optimal	1.0 ± 0.0
$\ell_{1.2}$-optimal	9.83 ± 1.27
$\ell_{1.05}$-optimal	8.74 ± 1.84

TABLE II
HESSIAN ALGORITHM: AVERAGE (AND STANDARD DEVIATION) NUMBER OF ITERATIONS OF THE LINE SEARCH ALGORITHM (FIG. 4)

Hessian	mean $\pm$ stdv
ℓ_3-optimal	10.92 ± 0.89
ℓ_2-optimal	1.0 ± 0.0
$\ell_{1.2}$-optimal	14.41 ± 1.43
$\ell_{1.05}$-optimal	14.29 ± 1.29

their counterparts in a global sense. We observe a slower decrease of the error in the first iterations of the gradient method with λ fixed, due to the conservative step size we have chosen to ensure convergence when we are close to the solution. The Hessian algorithm moves as fast as the version that uses the λ optimal at the beginning but afterwards; it gets slower as the diagonal of the Hessian gets larger for p close to 1 [when $p = 1$ the denominator of (10) becomes $\sum_{k\in\mathbb{Z}} |c(k)|^{-1} h^2(k - Nl)$]. Note that this behavior is not intuitive.

In conclusion, we recommend the Hessian-based, fixed step-size algorithm when $p \geq 2$ as it gives almost the same performance as the one that uses line search. For p close to 1, the algorithms that take advantage of λ optimal are more robust at the price of an added computational cost. Note that each iteration has a complexity comparable to that of the ℓ_2-projection. What makes the ℓ_p-algorithm computationally more expensive is the number of iterations required for reaching the solution. Based on the results in Fig. 5, we may conclude that 10–20 iterations are necessary. However, we found empirically that if instead of $c^{(0)} = 0$ we start with the ℓ_2-solution, we save of the order of ten iterations for $p < 2$.

C. Generation of Image Pyramids

If the basis functions φ used to specify h satisfy a two-scale relation, then the dyadic multiresolution for the linear case has the nestedness property of the vector spaces: $\cdots \subset V_{2^i} \subset V_{2^{i-1}} \subset \cdots \subset V_1$ [18], [24]. The discrete wavelet transform that minimizes the L_2-norm exploits this nestedness by computing the projection at one scale from the previous finer approximation. This hierarchical approach is not appropriate here. In principle, one should always go back to the finest scale to compute the coarse level approximations because of the nonlinear structure of the reduction operator.

In Fig. 6, we illustrate this distinction. We have generated the pyramid on the left hand side using the optimal $\ell_{1.2}$ approach (we take the finest resolution image as initial image to calculate all coarser approximations). The pyramid on the right hand side is suboptimal in the sense that each coarse-level approximation is computed from the previous finer level approximation. The reconstruction error is measured by the SNR_{ℓ_p} as defined before (here, $p = 1.2$). It is evident that the error of the approximation at level 1 is the same in both cases (26.11 dB) as we start from the same image. As expected, the error is slightly larger for the suboptimal [18.74 dB (level 2) and 14.53 dB (level 3)] than for optimal [18.99 dB (level 2) and 14.77 dB (level 3)] pyramid. In practice, these differences are not significant and it is quite justifiable to use the step-wise suboptimal approach to minimize computation. Nevertheless, in the following tests we did not use the recursive downsampling approach but the direct one.

IV. COMPARISON OF APPROXIMATIONS FOR DIFFERENT ps

In this section, we characterize the pyramid decompositions for different values of p. We observe two effects: The ringing is reduced and the histogram of the error gets sparser as p gets closer to 1.

A. Ringing

To illustrate the reduction of the ringing, we show in Fig. 7 the low-resolution approximations (reduced 1 : 4) of the image from Fig. 6—top, as a function of p. Here, the basis functions are cubic B-splines and the images are interpolated back to the original size. Observe the overshooting (ringing) for high ps in Fig. 7(a) and (b). It appears around the nucleus and border of the cells. On the other hand, in Fig. 7(c) and (d) the images are much less textured. Subjectively, these approximations are more pleasant visually because the regions are more nearly homogeneous.

B. Histogram Sparcity

Now, we center our attention on the study of the histograms corresponding to the detail images. Ideally, we would like our error image to be as sparse as possible, with an histogram presenting a high peak at zero. This would indicate that a large portion of the image is reproduced in the low-resolution approximation. With this idea in mind, we compare in Figs. 8 and 9 the histograms of the detail images for different values of p for a series of biomedical images. In all cases, the sparsest detail histograms correspond to ps close to 1, indicating that the gray value in the original image is more frequently kept in the low resolution approximation than for larger ps. For images in which the amount of noise (due to the characteristics of the image modality) is moderate or low, the height of the peak at zero is impressive.

Furthermore, combining the visual information from the approximated images and the detail histogram, we can derive the following conclusions: The near zero values in the detail histogram correspond to "large" objects in the original image where the term "large" is relative to the current scale. In other words, the "large" objects and background are kept in the approximation image while "small" objects are retained in the detail images. In addition, we benefit from an excellent preservation of the shape structures for low ps. The images are more blurred for high p (see Fig. 7).

The height of the peak at zero and the spread of the detail histogram depend heavily on the characteristics of the image.

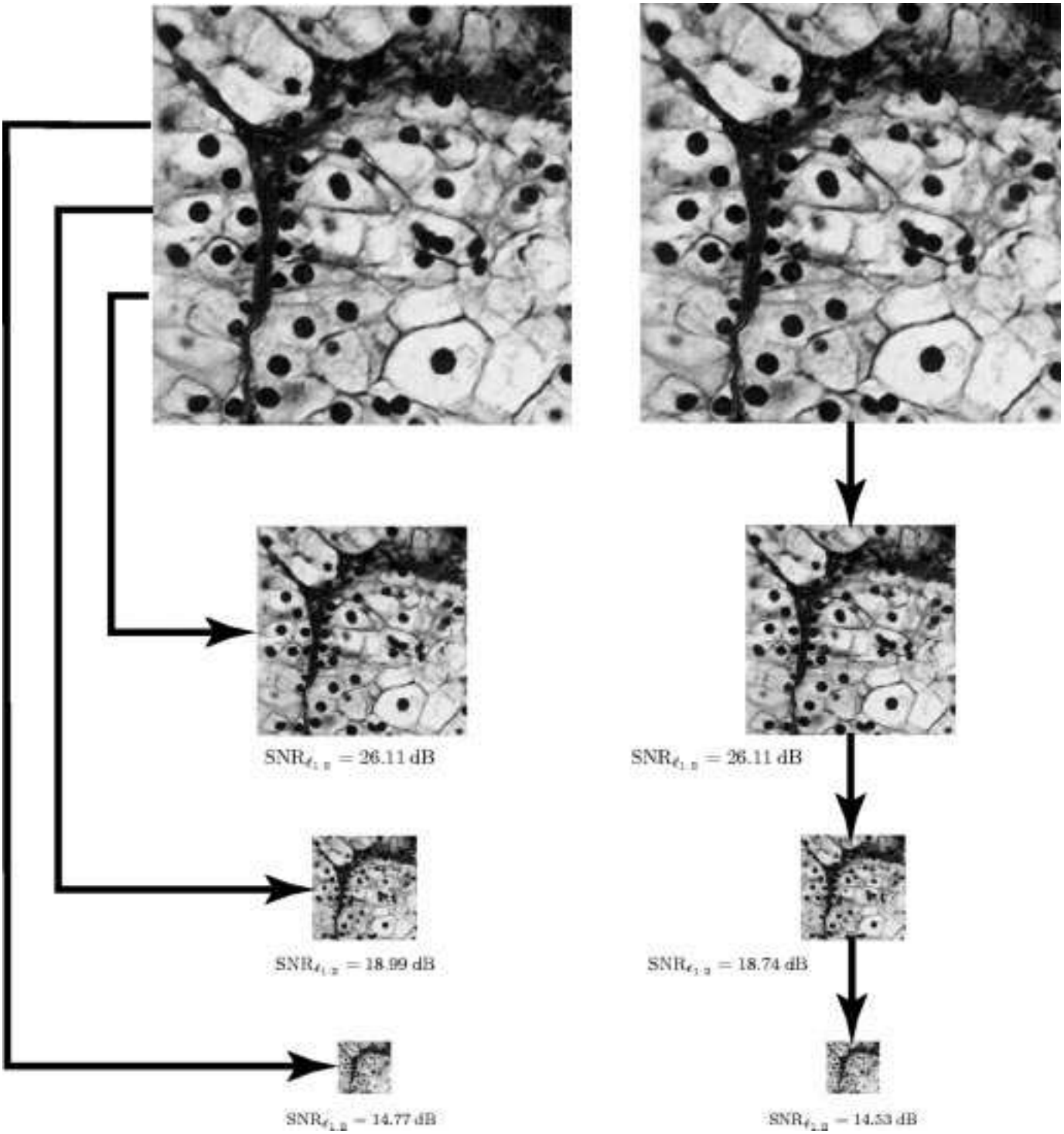

Fig. 6. Optimal versus suboptimal $\ell_{1.2}$ pyramid.

The peak is higher (resp., lower), if there is more (resp., less) edge information in the original image. The averaging effect characteristic of the least-squares approximation leads to a detail histogram with a Gaussian appearance. As p grows, the averaging gets even more accentuated. The spread increases with the degree of uniformity of the original histogram, independently of p.

Quantitative results for the images in Fig. 7 are given in Table III. Each column correspond to the results coming from a different ℓ_p-approximation. The table displays the values of the SNR_{ℓ_p}, the Kullback–Leibler distance between the histograms of the original and low resolution images and the entropy of the residues for each of the calculated ℓ_p-approximations.

We observe that the results are consistent: The minimum ℓ_p-error (maximum SNR_{ℓ_p}) is achieved for the corresponding ℓ_p-approximation in each case. The results of the Kullback–Leibler distance clearly indicate that the histogram of the image is best preserved for values of p close to 1. The entropy of the difference image also tends to get smaller for p close to 1.

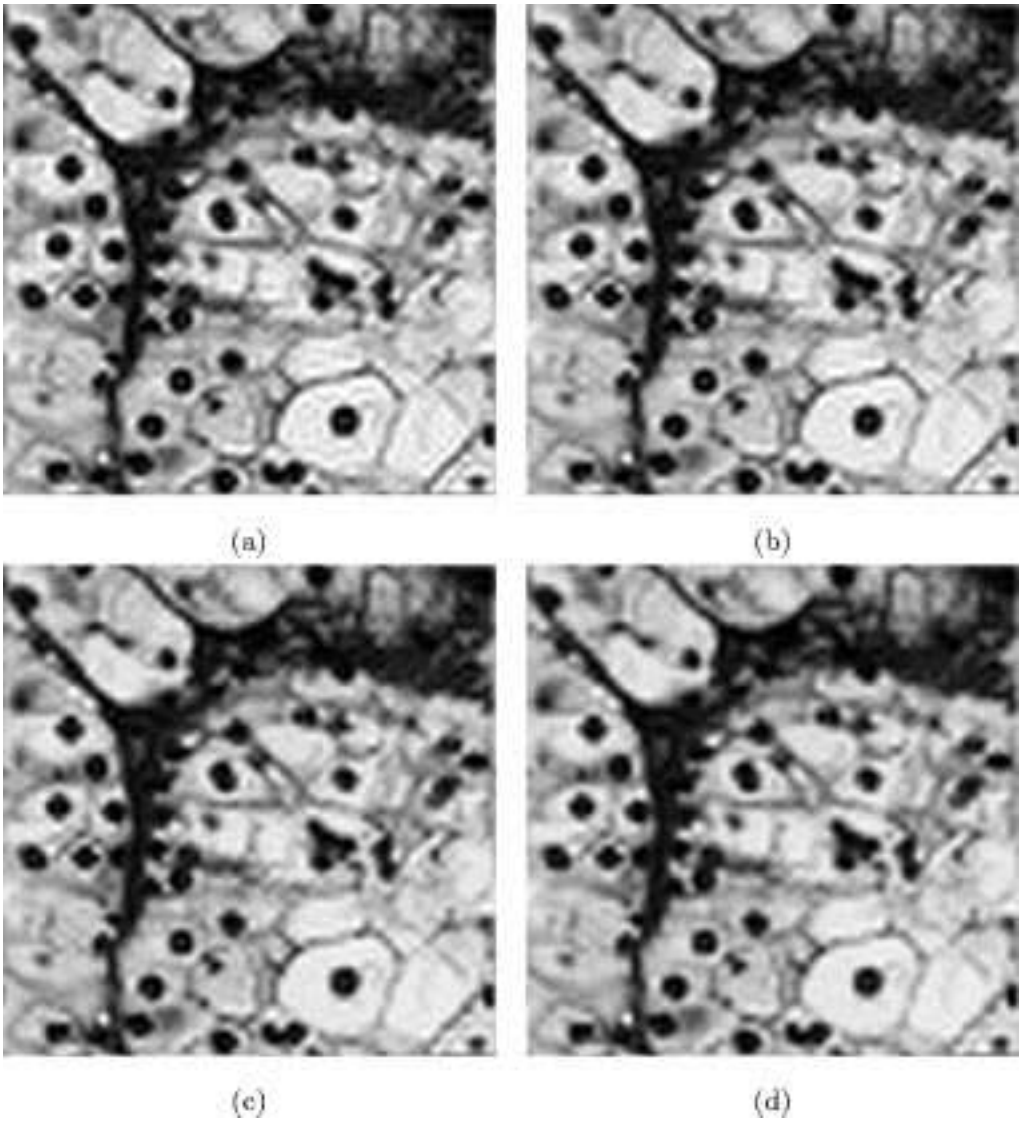

Fig. 7. Expanded version of the approximations calculated using cubic splines at scale 4 for different values of p. (a) ℓ_3-optimal, (b) ℓ_2-optimal, (c) $\ell_{1.2}$-optimal, and (d) $\ell_{1.05}$-optimal.

Note that the last results concerning the Kullback–Leibler and entropy measures are nothing but manifestations of the fact that ℓ_1-approximation tries to preserve the original image values.

Thus, our conclusion is that the most promising scheme is the ℓ_1-approximation because of the following properties.

- Preservation of the structure shapes at different scales which is appropriate for object detection.
- Reduction of ringing and spurious textures.
- The point structures are presented only at the finest scales of the detail images. This may be an advantage for some applications, for example, in the detection of microcalcifications on mammographies.

V. Comparison of Approximations for Different Orders

We now examine the choice of the degree of our spline basis functions. Mainly, we are concerned with the tradeoff between quality of approximation and computational complexity.

For our experiments we have chosen to compare approximations (reduction 1 : 4) calculated using B-splines of degrees 0, 1, 3, and 5 as basis functions. The results are shown in Figs. 10 and 11 for $p = 2$ and $p = 1.05$, respectively. We observe that the blocking artifacts typical of piecewise constant spline approximation [Figs. 10(a) and 11(a)] disappear for higher order splines. Most observers will also agree that the subjective quality of spline approximations with ps close to 1 (Fig. 11) is better than $p = 2$ (Fig. 10) for all degrees. Note that for the least-squares case the ringing gets visibly accentuated as the spline degree increases, while this is less the case for $p = 1$.

In Table IV, we give the SNR_{ℓ_p} and entropy of the difference image that correspond to the ℓ_1 and ℓ_2 approximations calculated for different spline degrees. We observe that we have lower values of the entropy (maximum SNR_{ℓ_p}) with ps close to 1 when using spline basis of the same degree. The minimum ℓ_2-error is reached for the higher order splines. This finding is consistent with the standard theory of splines [25], [26]: as the degree

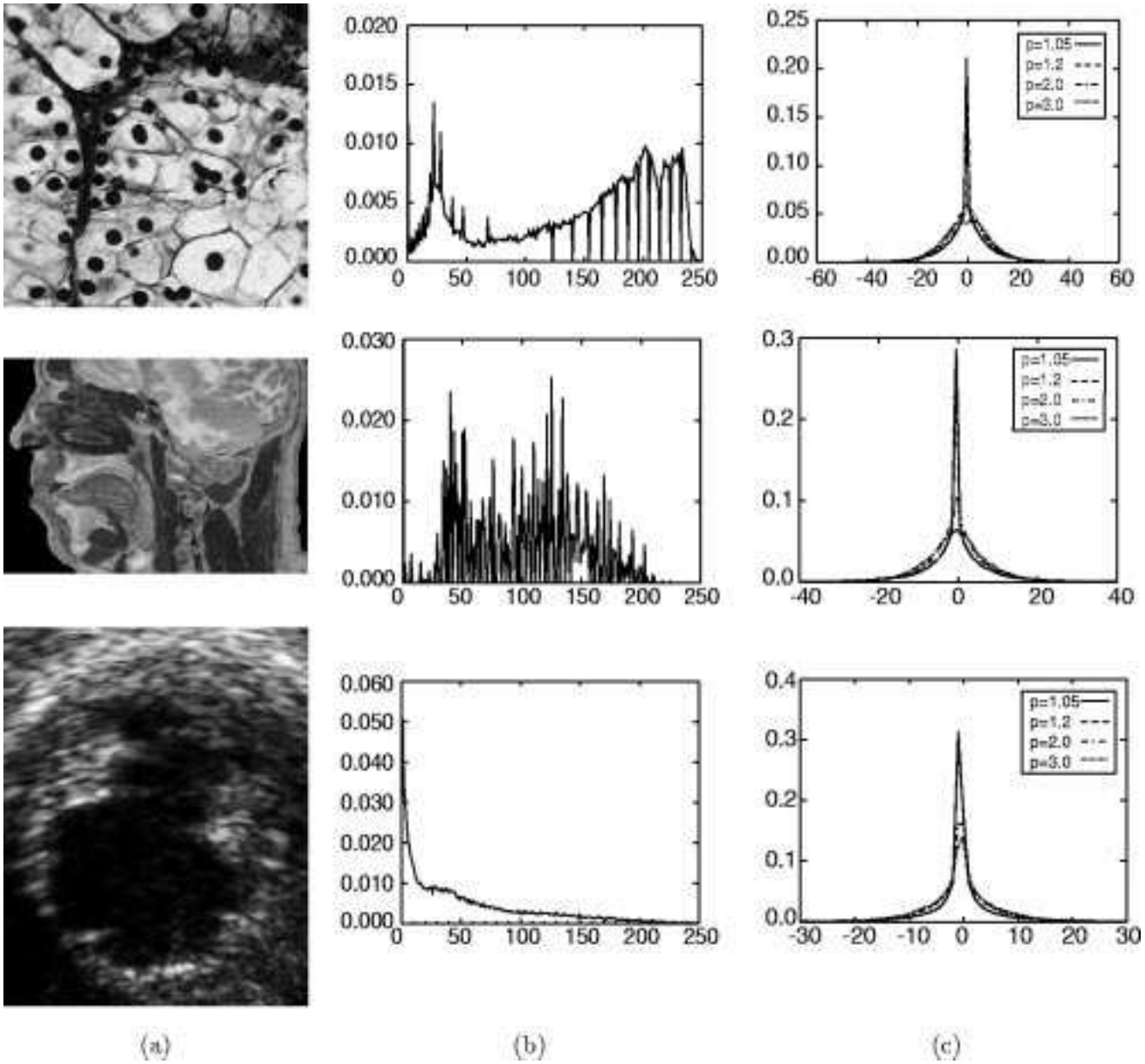

Fig. 8. (a) Biomedical images, (b) corresponding histogram, and (c) histogram of the detail images (original minus the approximated version at scale 2 using cubic splines as basis functions) for different values of p. Note the high peak at zero for p close to 1.

n increases, the spline approximation converges to Shannon's solution which minimizes the L_2-error when the function is bandlimited or very lowpass (usual case for an image). The main drawback of the least-squares distance measure is that it does not penalize enough oscillations and ringing artifacts. Interestingly, if one looks at the ℓ_1-approximation, the optimal model turns out to be the cubic spline ($n = 3$). The fact that too high an order splines are not good with respect to ℓ_1-approximations is not surprising because the basis functions tend to $\mathrm{sinc}(x)$ whose samples are not in ℓ_1. This is also consistent with the fact that the ℓ_1-distance is the one that penalizes ringing most.

The ℓ_1-cubic splines are also best in terms of data compression (entropy minimization), combining a good order of approximation with a reduction of artifacts.

VI. Discussion

A. ℓ_p-Pyramid Versus Median Pyramid

As mentioned in the introduction, median pyramids have been widely used in the literature because of their desirable properties of edge and detail preservation [15], [27].

The reducer operator of the median pyramids computes the decimated version of a median filter output. We can, as with our method, either start from the original image for all the resolution levels, or apply the successive refinement scheme. As we have already pointed out before, the results will be different.

Our model is equivalent to a median pyramid in one particular case: while minimizing the error in the ℓ_1-sense and using as the B-spline of degree zero as the interpolation function. If the quality of the low-order interpolation is not satisfactory (e.g.,

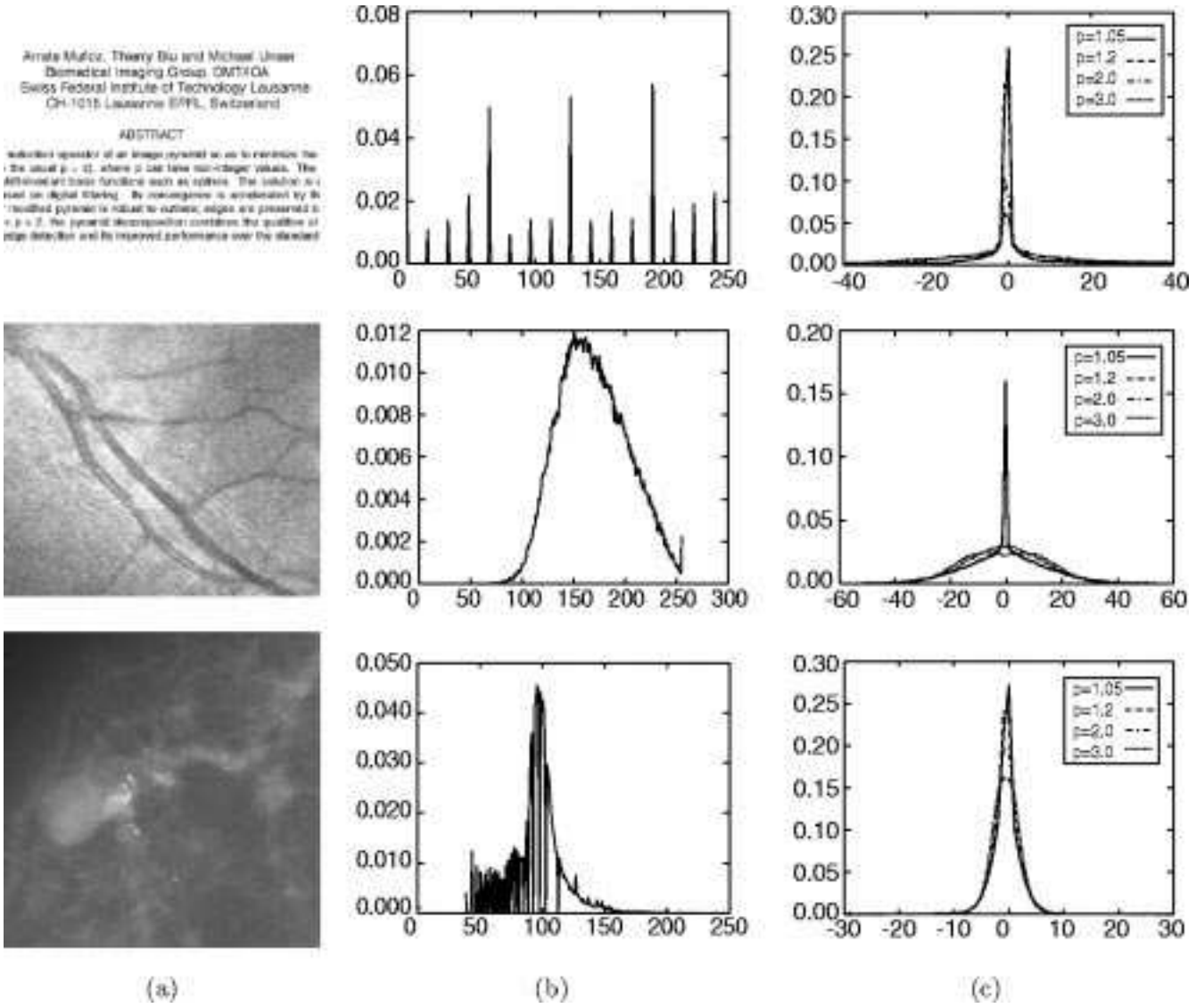

Fig. 9. (a) Test images, (b) corresponding histogram, and (c) histogram of the detail images (original minus the approximated version at scale 2 using cubic splines as basis functions) for different values of p. Note the high peak at zero for p close to 1.

TABLE III
SNR$_{\ell_p}$ MEASURES CORRESPONDING TO THE IMAGES IN FIG. 7

energy/distance	ℓ_3-optimal	ℓ_2-optimal	$\ell_{1.2}$-optimal	$\ell_{1.05}$-optimal
SNR$_{\ell_3}$ (dB)	**16.22**	16.08	15.43	15.11
SNR$_{\ell_2}$ (dB)	17.34	**17.47**	17.21	16.99
SNR$_{\ell_{1.2}}$ (dB)	18.38	18.79	**18.99**	18.90
SNR$_{\ell_{1.05}}$(dB)	18.58	19.05	19.37	**19.31**
Kullback-Leibler	0.189	0.163	0.138	**0.130**
Entropy	4.485	4.442	4.397	**4.388**

because of blocking artifacts), we can simply increase the approximation order by increasing the spline degree. We will still be optimal in the ℓ_1-sense but our reduction operator will no longer correspond to a median filter. Naturally, the approximation error decreases as our interpolation model improves.

The advantages of our ℓ_p-approximation model over a classical median pyramid can be summarized as follows.

- The reduction operator is consistent with our approximation model.
- The error is minimized in a well-defined sense.
- The existence of an underlying continuous model allows for the evaluation of continuously-defined operators.
- The model is flexible as it is possible to tune the parameter p and the degree of the B-spline which determines the space in which the original image is projected.

B. Perceptual Relevance of the ℓ_1 Metric

There are two related aspects when computing and evaluating image approximations that should be considered:

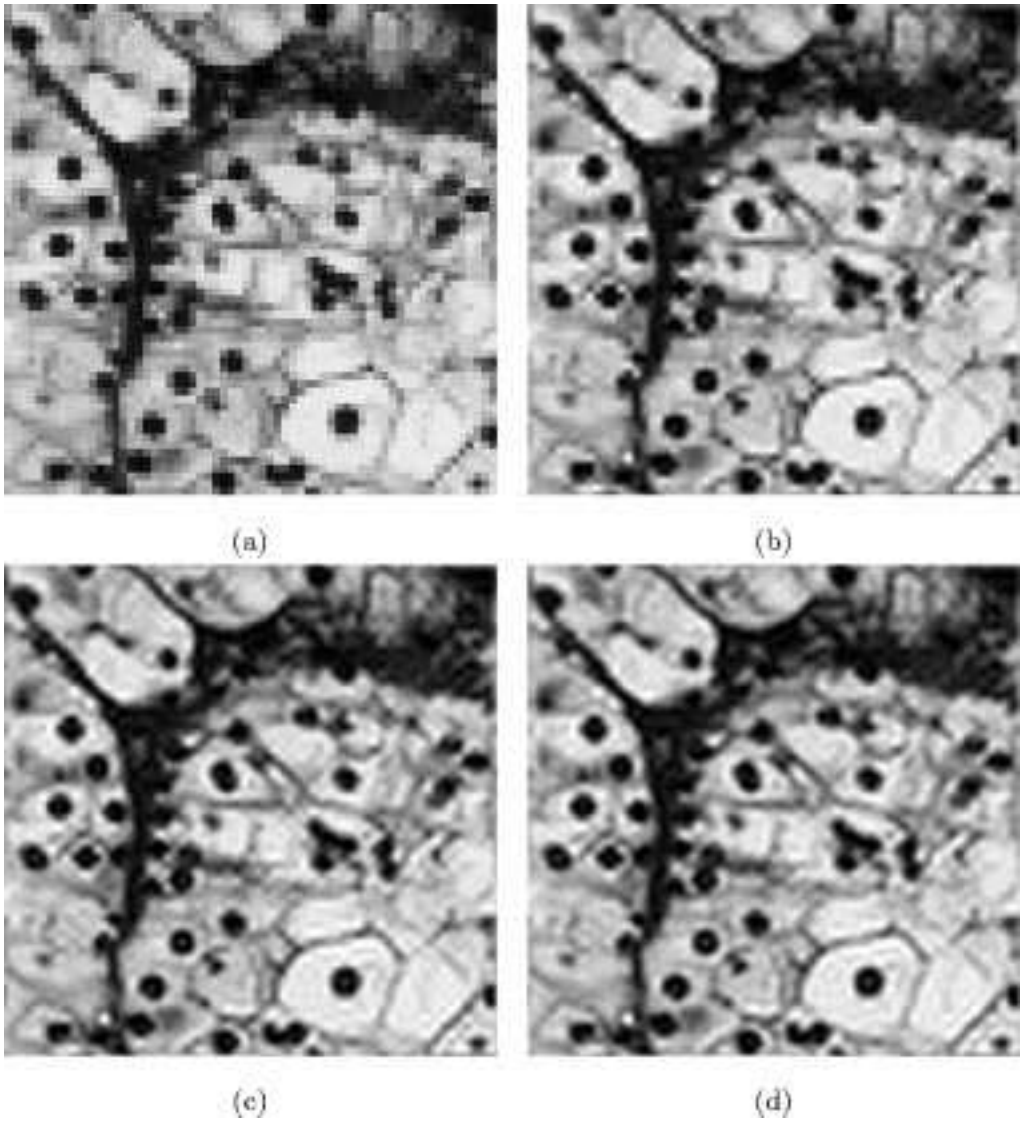

Fig. 10. Comparison of the least-squares ($p = 2.0$) approximations calculated at scale 4 for different degrees of the splines. (a) $n = 0$, (b) $n = 1$, (c) $n = 3$, and (d) $n = 5$. See Table IV for quantitative error information.

- the optimization criterion and the algorithm (ℓ_p-projection) used to approximate the input signal;
- the ℓ_p-metric used to measure the approximation error.

Obviously, if we know the metric that best matches our visual perception of image quality, it makes good sense to use the corresponding approximation algorithm.

From a perceptual point of view, what we consider to be a good result depends on the sensitivity of the human observer to details at different frequencies and contrasts [28]. From the examples collected in this paper and our experimentation with the algorithms, we are tempted to conclude that the ℓ_1-projections look perceptually better than the ones obtained with larger values of p. On the other hand, the error images for ℓ_1-projections also contain details and features that are more noticeable visually. This is consistent with the observation that ℓ_1-approximation has a stronger tendency than others to simplify images.

In [29], DeVore *et al.* performed experiments to determine the L_p-norm that best matched the response of the visual system. They concluded that the L_1-norm was the most appropriate for measuring image compression errors. They used these results to justify their nonlinear wavelet-based compression algorithm. Their findings correlate well with our results. It is clear from the images that we have presented that ringing is disturbing visually. The ℓ_1-norm comes out best because it is the one that penalizes the oscillations most (esp., ringing due to the sinc is bounded if we measure it with the L_2 norm whereas it is not with L_1).

VII. Conclusions

We have presented a theoretical framework for obtaining multiresolution image approximations with non-Euclidean norms. In addition, we have proposed an efficient iterative algorithm based on digital filtering to calculate these approximations. In the experimental part, we found ℓ_1-pyramids to be the most promising ones. Overall, they led to better feature preservation and resulted in less ringing artifacts. They also produced the sparsest error images which is relevant for coding applications. These are all properties that should make them useful for multiscale processing.

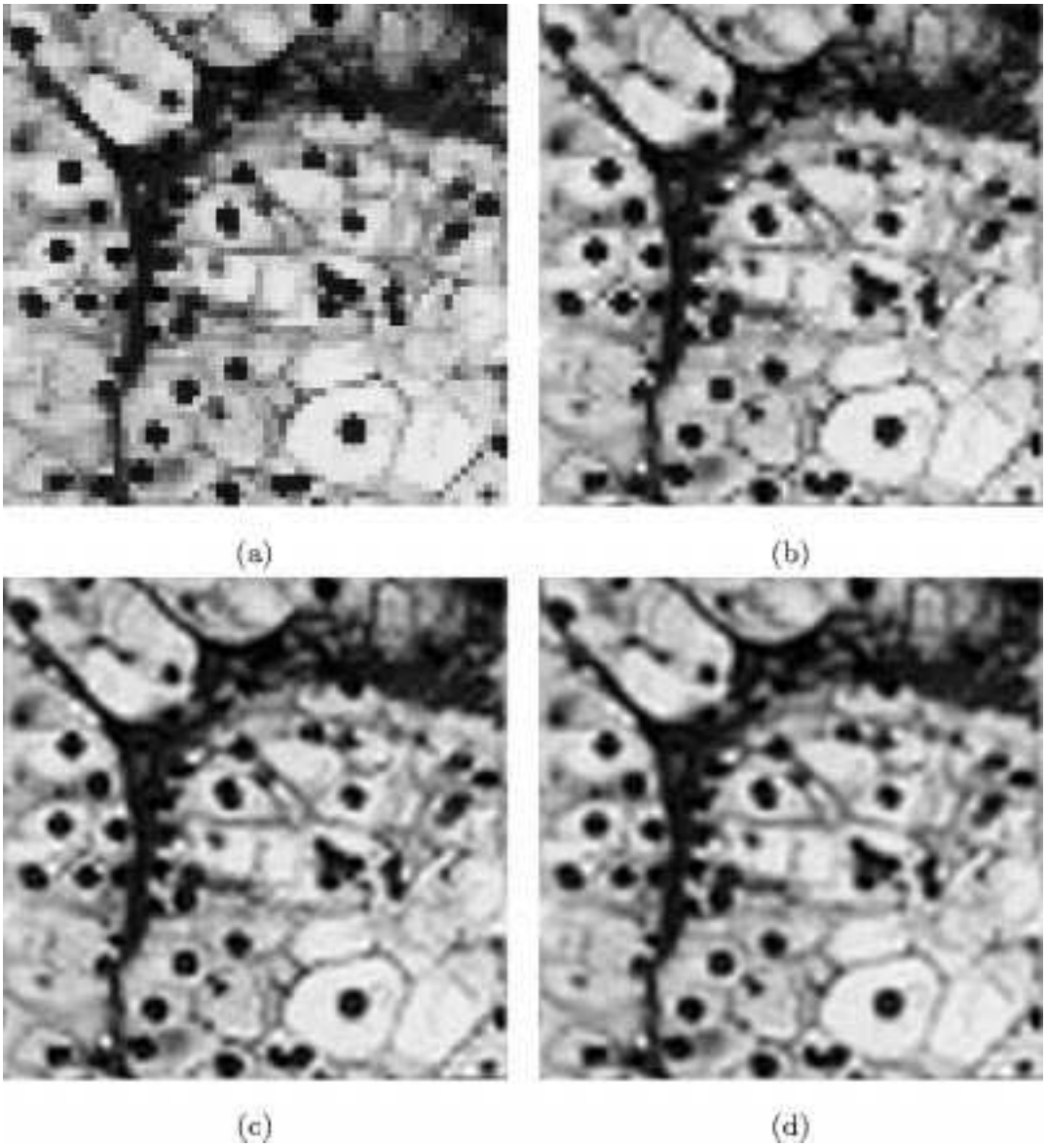

Fig. 11. Comparison of the ℓ_1-approximations calculated at scale 4 for different degrees of the splines. (a) $n = 0$, (b) $n = 1$, (c) $n = 3$, and (d) $n = 5$. See Table IV for quantitative error information.

TABLE IV
SNR$_{\ell_p}$ MEASURES AND ENTROPY CORRESPONDING TO THE ℓ_1 AND ℓ_2-APPROXIMATIONS FOR DIFFERENT SCALES AND DEGREES OF THE SPLINE BASIS FUNCTIONS

	Degree	$n=0$		$n=1$		$n=3$		$n=5$	
N		ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2	ℓ_1	ℓ_2
2	SNR$_{\ell_p}$ (dB)	22.60	19.77	25.74	22.78	**26.23**	24.15	25.20	**24.23**
	Entropy	3.995	4.043	3.495	3.757	**3.494**	**3.645**	3.648	3.650
3	SNR$_{\ell_p}$ (dB)	19.79	16.73	21.39	19.09	**21.76**	19.87	21.29	**19.92**
	Entropy	4.185	4.383	4.096	4.224	**4.082**	**4.157**	4.149	4.159
4	SNR$_{\ell_p}$ (dB)	17.78	15.00	19.01	16.96	**19.31**	17.47	19.02	**17.50**
	Entropy	4.526	4.610	4.410	4.487	**4.386**	**4.442**	4.426	4.443
5	SNR$_{\ell_p}$ (dB)	16.51	13.87	17.62	15.70	**17.90**	16.10	17.68	**16.13**
	Entropy	4.656	4.767	4.586	4.642	**4.556**	**4.603**	4.583	4.603

Another interesting finding is that cubic splines gave the best results among all other splines when the approximation was done in the ℓ_1-norm. Unlike the ℓ_2-ranking which always gives the advantage to higher-order approximations, this result correlates well with the fact that the cubic spline model is often the preferred one in applications [20]. Again, this supports the general perception that cubic B-splines offer the best compromise in terms of approximation power versus the support of the basis functions.

APPENDIX

Our proof of Theorem 1 was inspired by the work of Aldroubi *et al.* [17] who considered L_p-Riesz basis in a continuous framework. It makes use of two classical results

Proposition 1 (Young's Inequality): If $b \in \ell_1$ and $a \in \ell_p$ then $||a * b||_{\ell_p} \leq ||b||_{\ell_1} \cdot ||a||_{\ell_p}$.

Lemma 1 (Wiener's Lemma): Let $a \in \ell_1$ with $A(e^{j2\pi f}) \neq 0 \; \forall f$, then $(a)^{-1} \leftrightarrow 1/A(e^{j2\pi f})$ is in ℓ_1 as well.

These are used to establish the following.

Lemma 2: Let $h \in \ell_1$ generate a N-shift-invariant ℓ_2-Riesz basis. Then, its ℓ_2-dual $\overset{\circ}{h}$ defined by $\overset{\circ}{h} = (a_h^{-1})_{\uparrow N} * h$ with $a_h(l) = \langle h(k), h(k - lN) \rangle = (h * h^T)_{\downarrow N}$, is in ℓ_1 as well.

Proof: Thanks to Young's inequality, we have that $a_h \in \ell_1$ because

$$||a_h||_{\ell_1} = ||(h * h^T)_{\downarrow N}||_{\ell_1} \leq ||h * h^T||_{\ell_1} \leq ||h||_{\ell_1}^2 < +\infty.$$

Since $\{h(k - lN)\}_{l \in \mathbb{Z}}$ is a ℓ_2-Riesz basis, the autocorrelation function

$$\alpha \leq \underbrace{\sum_{i=0}^{N-1} \left| H\left(e^{j2\pi(f-i)/N}\right) \right|^2}_{A_h(e^{j2\pi f})} \leq \beta$$

is positive definite (cf. Theorem 2). The conditions of Wiener's lemma are met, thus, $a_h \in \ell_1$ implies that a_h^{-1} belongs to ℓ_1 as well. We show that $\overset{\circ}{h} \in \ell_1$ by using Young's inequality and the fact that upsampling does not change the value of the norm

$$\left\| \overset{\circ}{h} \right\|_{\ell_1} = ||(a_h^{-1})_{\uparrow N} * h||_{\ell_1} \leq ||a_h^{-1}||_{\ell_1} \cdot ||h||_{\ell_1}.$$

■

We can now proceed with the proof of Theorem 1.

Proof: The goal is to establish upper and lower bounds in the norm equivalence

$$\forall c \in \ell_p, \quad A \cdot ||c||_{\ell_p} \leq ||s||_{\ell_p} = ||c_{\uparrow N} * h||_{\ell_p} \leq B \cdot ||c||_{\ell_p}.$$

The upper bound is easily localized using Young's Inequality and the fact that $||c_{\uparrow N}||_{\ell_p} = ||c||_{\ell_p}$

$$||c_{\uparrow N} * h||_{\ell_p} \leq ||c||_{\ell_p} \underbrace{||h||_{\ell_1}}_{B}.$$

To determine a lower bound we will work with the dual filter $\overset{\circ}{h}$ which is in ℓ_1 as well, as a consequence of Lemma 2. Since $\overset{\circ}{h}$ and h are biorthogonal, we have that $\forall s \in V_N$

$$s(k) = \sum_{l \in \mathbb{Z}} \underbrace{\left\langle s, \overset{\circ}{h}(k - lN) \right\rangle}_{c_l} h(k - lN)$$

$$||c||_{\ell_p} = \left\| \left[s * \overset{\circ}{h}^T \right]_{\downarrow N} \right\|_{\ell_p} \leq \left\| s * \overset{\circ}{h}^T \right\|_{\ell_p} \leq \underbrace{\left\| \overset{\circ}{h} \right\|_{\ell_1}}_{A^{-1}} ||s||_{\ell_p}$$

with $s = c_{\uparrow N} * h$. So, we have found A such as $A||c||_{\ell_p} \leq ||c_{\uparrow N} * h||_{\ell_p}$. ■

ACKNOWLEDGMENT

The authors would like to thank A. Aldroubi for fruitful discussions. The clear cell carcinoma of the kidney image is courtesy of the Cornell University Medical Center.

REFERENCES

[1] A. Rosenfeld, *Multiresolution Image Processing and Analysis*. Berlin, Germany: Springer, 1984.
[2] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact code," *IEEE Trans. Commun.*, vol. 15, no. 1–2, pp. 1–21, 1986.
[3] S. Mallat, *A Wavelet Tour of Signal Processing*, 2 ed. New York: Academic, 1999.
[4] ——, "Wavelets for a vision," *Proc. IEEE*, vol. 4, pp. 604–614, Apr. 1996.
[5] O. Egger, P. Fleury, T. Ebrahimi, and M. Kunt, "High-performance compression of visual information—A tutorial review: I. Still pictures," *Proc. IEEE*, vol. 87, pp. 976–1013, June 1999.
[6] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding," in *Proc. SPIE*, 1987, pp. 50–58.
[7] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
[8] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing: Part I—Theory, Part II—Efficient design and applications," *IEEE Trans. Signal Processing*, vol. 41, pp. 821–832, Feb. 1993.
[9] ——, "The L_2 polynomial spline pyramid," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 364–379, Apr. 1993.
[10] P. P. Vaidyanathan and B. Vrcelj, "Biorthogonal partners and applications," *IEEE Trans. Signal Processing*, vol. 49, pp. 1013–1027, May 2001.
[11] A. Aldroubi, M. Unser, and M. Eden, "Cardinal B-spline filters: Stability and convergence to the ideal lowpass filter," *Signal Process.*, vol. 28, pp. 1–12, 1992.
[12] I. Defee and Y. Neuvo, "Nonlinear filters in image pyramid generation," in *Proc. IEEE Int. Conf. Systems Engineering*, Fairborn, OH, 1991, pp. 269–272.
[13] A. Toet, "A morphological pyramidal image decomposition," *Pattern Recognit. Lett.*, vol. 9, no. 1, pp. 255–261, 1989.
[14] J. Goutsias and A. M. Heijmans, "Nonlinear multiresolution signal decomposition schemes—Part I: Morphological pyramids; Part II: Morphological wavelets," *IEEE Trans. Image Processing*, vol. 9, pp. 1862–1913, Nov. 2000.
[15] J. L. Starck, F. Murtagh, and M. Louys, "Astronomical image compression using the pyramidal median transform," in *Astronomical Data Analysis Software and Systems IV*: Astron. Soc. Pacific, 1995, vol. 77, pp. 268–271.
[16] V. Melnik, I. Shmulevich, K. Egiazarian, and J. Astola, "Block-median pyramidal transform: Analysis and denoising applications," *IEEE Trans. Signal Processing*, vol. 49, pp. 264–372, Feb. 2001.
[17] A. Aldroubi and K. Gröchenig, "Nonuniform sampling and reconstruction in shift-invariant spaces," *SIAM Rev.*, vol. 43, no. 4, pp. 585–620, 2001.
[18] A. Aldroubi and M. Unser, "Oblique projections in discrete signal subspaces of ℓ_2 and the wavelet transform," *Proc. SPIE*, vol. 2303, pp. 36–46, 1994.
[19] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-order interpolation of minimal support," *IEEE Trans. Image Processing*, vol. 10, pp. 1069–1080, July 2001.
[20] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Mag.*, vol. 16, pp. 22–38, Nov. 1999.
[21] A. Aldroubi, M. Unser, and M. Eden, "Discrete spline filters for multiresolutions and wavelets of ℓ_2," *SIAM J. Math. Anal.*, vol. 25, pp. 1412–1432, 1994.
[22] R. Kress, *Numerical Analysis*. New York: Springer, 1998.
[23] D. Pierre, *Optimization Theory With Applications*. New York: Dover Publications, 1986.
[24] O. Rioul, "A discrete-time multiresolution theory," *IEEE Trans. Signal Processing*, vol. 41, pp. 2591–2606, Aug. 1993.

[25] T. Blu and M. Unser, "Quantitative Fourier analysis of approximation techniques: Part I—Interpolators and projectors; Part II—Wavelets," *IEEE Trans. Signal Processing*, vol. 47, pp. 2783–2806, Oct. 1999.

[26] M. Unser, A. Aldroubi, and M. Eden, "Polynomial spline signal approximations: Filter design and asymptotic equivalence with Shannon's sampling theorem," *IEEE Trans. Inform. Theory*, vol. 38, pp. 95–103, Feb. 1992.

[27] V. Melnik, I. Shmulevich, K. Egiazarian, and J. Astola, "Iterative block-median pyramid transform-based denoising," in *Proc. SPIE*, 1999, pp. 502–511.

[28] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer.*, vol. 7, pp. 2030–2040, 1990.

[29] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, Mar. 1992.

Arrate Muñoz (S'00) was born in Pamplona, Spain, in 1973. She did her studies in telecommunication engineering in the Public University of Navarra (UPNA), Navarra, Spain, from which she graduated in July 1997. She received the Ph.D. degree from the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland, for her thesis "Nondyadic and nonlinear multiresolution image approximations."

She is now a Research Assistant with the Biomedical Imaging Group, EPFL. Her main research topics are geometrical transformations of images, nonuniform splines, and non-Euclidean norms applied to multiresolution.

Thierry Blu (M'96) was born in Orléans, France, in 1964. He received the "Diplome d'ingnieur" degrees from École Polytechnique, France, in 1986 and from Télécom Paris (ENST), France, in 1988. In 1996, he received the Ph.D degree in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding.

He is currently with the Biomedical Imaging Group, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, on leave from France Télécom National Center for Telecommunications Studies (CNET), Issy-les-Moulineaux, France. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, approximation and sampling theory, and psychoacoustics.

Michael Unser (M'88–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering in 1981 and 1984, respectively, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD. He is now Professor and Head of the Biomedical Imaging Group, EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He is the author of more than 90 published journal papers in these areas. He is on the editorial boards of including *Signal Processing*. He serves as regular chair for the SPIE Conference on Wavelets, held annually since 1993.

Dr. Unser is an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING. He is on the editorial boards of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992-1995) and IEEE SIGNAL PROCESSING LETTERS (1994-1998). He was general co-chair for the First IEEE International Symposium on Biomedical Imaging (ISBI'2002), Washington, DC, July 7-10, 2002. He received the 1995 Best Paper Award and the 2000 Magazine Award from the IEEE Signal Processing Society.