Maximally Localized Radial Profiles for Tight Steerable Wavelet Frames

Pedram Pad, *Student Member, IEEE*, Virginie Uhlmann, and Michael Unser, *Fellow, IEEE*

***Abstract*— A crucial component of steerable wavelets is the radial profile of the generating function in the frequency domain. In this paper, we present an infinite-dimensional optimization scheme that helps us find the optimal profile for a given criterion over the space of tight frames. We consider two classes of criteria that measure the localization of the wavelet. The first class specifies the spatial localization of the wavelet profile, and the second that of the resulting wavelet coefficients. From these metrics and the proposed algorithm, we construct tight wavelet frames that are optimally localized and provide their analytical expression. In particular, one of the considered criterion helps us finding back the popular Simoncelli wavelet profile. Finally, the investigation of local orientation estimation, image reconstruction from detected contours in the wavelet domain, and denoising indicate that optimizing wavelet localization improves the performance of steerable wavelets, since our new wavelets outperform the traditional ones.**

***Index Terms*— Isotropic wavelets, filter design, steerable wavelets, localization, denoising, image reconstruction.**

I. Introduction

ISOTROPIC wavelets are purely radial functions that allow for an orientation-free decomposition of images, while retaining all other popular features of classical wavelet bases such as multiresolution analysis. The steerable pyramid [1]–[5] is a well-known construction that relies on such wavelets. In this setting, a purely angular element is included in order to rotate derivatives of the wavelets and impose a preferred directionality. Some well-known algorithms for denoising ([6], including the widely-used Bayesian least-squares Gaussian-scale-mixture (BLS-GSM) algorithm [7]), texture analysis (or synthesis) [8], [9], and regularization with sparsity constraints for inverse problems [10], [11] rely on the steerable pyramid, although methods that do not exploit steerability are also available for these tasks. Steerability is a crucial aspect in many other image-processing applications such as finding the dominant orientation at each image location, detecting contours [12], or identifying features in a rotation-invariant fashion [13]. More recently, algorithms for image reconstruction from the small subset of wavelet coefficients called the "primal sketch" have been proposed relying on the steerable pyramid [14], [15]. In this work, we study the design of wavelet profiles for use in applications relying on steerable tight frames. The specification of steerable wavelet frames includes two components: a radial profile and a directional components can be optimized separately. The angular component is represented using circular harmonics [4]. Here, we concentrate on the radial profile, which determines the localization.

In order to generate an isotropic wavelet transform, the underlying basis functions must satisfy several properties, the main ones being isotropy and perfect reconstruction of the image. Another desirable feature is that the basis functions form a tight frame. In this way, the wavelet transform is self-reversible, enabling simpler and faster algorithms. The isotropy and perfect-reconstruction conditions are ensured by choosing a radially bandlimited mother wavelet that satisfies some partition of unity in the frequency domain [16], [17]. Many such bases have been proposed, which include the Meyer [18], Papadakis [19], and Simoncelli [20] wavelets. Inspired by the biological visual system, the Simoncelli wavelet is the one implemented in the original version of the steerable pyramid and the BLS-GSM denoising algorithm. Due to its good performance in a wide range of practical applications, it remains a commonly used profile.

Localization, either in the spatial or frequency domain, is an important consideration for specifying such wavelets. On the practical side, steerable wavelets are bandlimited with infinitely many vanishing moments, which tends to induce oscillations that can be visually displeasing. It is observed that more-localized wavelets result in fewer oscillations and are less subject to truncation artifacts. Moreover, it has been theoretically shown that wavelets with better localization are more efficient for decoupling and sparsifying signals [21]. It is worth mentioning that the Simoncelli wavelet performs well in a wide range of applications and is shown to be the most-localized wavelet in a specific sense [22].

Here, we introduce a method to design radial profiles for steerable tight frames. Since the frequency response of steerable wavelets is polar-separable, we can concentrate on the task of optimizing the radial frequency profile. We focus on moment-based measures of localization and propose two different classes of criteria depending on whether we consider the localization in the spatial domain or in the wavelet domain.

Manuscript received August 5, 2015; revised January 22, 2016 and March 7, 2016; accepted March 14, 2016. Date of publication March 22, 2016; date of current version April 7, 2016. This work was supported in part by the Swiss National Science Foundation under Grant 200020-121763 and Grant 200020-144355 and in part by the European Research Council under Grant ERC-2010-AdG 267439-FUN-SP. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alin M. Achim. *(Pedram Pad and Virginie Uhlmann contributed equally to this work.)*

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: pedram.pad@epfl.ch; virginie.uhlmann@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2545301

1057-7149 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Two criteria can be derived within each class, depending on whether one wants to consider localization over the whole space or in each radial direction. Ultimately, the choice of a particular criterion among these is guided by the application. We describe an algorithm using the calculus of variations to optimize the wavelet corresponding to each measure through gradient descent. We then obtain analytical expressions for the optimally localized profiles.

Our preliminary investigations have revealed the practical benefits of such an optimization [22]–[24]. In this work, we extend the framework and present a unified treatment that relies on more sophisticated tools and improves on our previous findings. As a result, we derive new wavelet profiles that exhibit optimal localization properties. We then show the benefit of our optimized design in three practical applications, namely, local orientation estimation, image reconstruction from edges, and denoising. These experiments highlight different use-cases in which each of the proposed type of localization (spatial versus wavelet domain) is desirable. In particular, we provide additional results on the image-reconstruction problem compared to [23], as well as further study of the performance of our wavelets for local orientation estimation and for the BLS-GSM denoising algorithm.

The organization of the paper is as follows: We review in Section II some important aspects of the design of isotropic wavelet frames, both on the theoretical and practical side. In Section III, we specify our measures of localization and propose a step-by-step algorithm to design optimally localized profiles corresponding to each measure. We then provide the closed-form expression of the resulting optimal wavelets. Finally, we focus on three practical applications in Section IV, namely, local orientation estimation in filamentous structures, image reconstruction from a primal sketch and image denoising using BLS-GSM. We use our novel optimally localized wavelet profiles and compare them against well-known wavelet profiles such as the Papadakis, Meyer, and Simoncelli wavelets.

II. Preliminaries on Isotropic Wavelets

In this section, we specify the class of wavelets that we consider and optimize in this paper. Let the mother wavelet ψ be a function from $\mathbb{R}^2$ to $\mathbb{R}$. The complete wavelet frame is given by

$$\psi_{i,\mathbf{k}}(\mathbf{x}) = 2^{-\frac{i}{2}}\psi\left(\frac{\mathbf{x}}{2^i} - \mathbf{k}\right), \tag{1}$$

in which $i \in \mathbb{Z}$ and $\mathbf{k} \in \mathbb{Z}^2$. It has a redundancy factor of $4/3$.

Our first assumption is that ψ is isotropic. Thus, we write

$$\psi(\mathbf{x}) = \psi(r), \tag{2}$$

where $r = \|\mathbf{x}\|$. Here, for simplicity, we use the notation ψ both for the wavelet and for the radial profile of the wavelet. Throughout the paper, we put the arguments of operators in curly brackets and the arguments of functionals in parentheses. When there is no ambiguity, we may choose to do away with brackets and parentheses, a typical case being the gradient ∇. Now, let h be the 2-dimensional Fourier transform of ψ, defined as

$$h(\boldsymbol{\omega}) = \mathcal{F}\{\psi\}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} \psi(\mathbf{x}) \exp(-\mathrm{j}\boldsymbol{\omega}^\top \mathbf{x}) \mathrm{d}\mathbf{x}. \tag{3}$$

According to the properties of the Fourier transform, h is also isotropic. Hence, like (2), we write

$$h(\boldsymbol{\omega}) = h(\rho), \tag{4}$$

where $\rho = \|\boldsymbol{\omega}\|$. According to [26], ψ and h are related, for $r \geq 0$, through the Hankel transform as

$$\psi(r) = \mathcal{H}\{h\}(r) = \int_0^\infty h(t)\ J_0(rt)\ t \mathrm{d}t. \tag{5}$$

The second assumption is that ψ creates a tight-frame family. This implies the condition [16], [17]

$$\sum_{i\in\mathbb{Z}} |h(2^i\rho)|^2 = 1 \quad \text{for} \quad \forall \rho \in \mathbb{R}^+ \backslash \{0\}. \tag{6}$$

There is an equivalence between stating (6) for all positive ρ and stating (6) for $\rho \in [\frac{\pi}{2}, \pi]$ because $2^i\rho$, for $i \in \mathbb{Z}$, can reach any arbitrary positive value.

With these preliminaries, we are ready to give a framework that allows us to find the most-localized wavelet among the ones that satisfy the conditions of Section II. The proposed method can be used for a broad class of measures of localization.

Algorithm 1 Most Localized Wavelet

initialize: $h \in L_2([0, \infty])$
initialize: $\eta > 0$
repeat
  $\tilde{h} \leftarrow h - \eta\ \nabla V\{h\}$
  $h \leftarrow \mathcal{P}\{\tilde{h}\}$
until h converges
return h

III. Localized Isotropic Wavelets

In this section, we present a general framework that relies on the calculus of variations to find the optimal wavelet with respect to a given localization measure. We restrict ourselves to wavelets specified in Section II and focus on two natural classes of moment-based measures.

Assume that the functional V is a given measure of localization. We shall consider that this measure operates in the Fourier domain. When V is Gâteaux differentiable, the natural method of minimizing it would be to use a variation of the steepest-descent algorithm [27]. To do this, we first need the gradient of V. As V is a functional on an infinite-dimensional space, we have to rely on the calculus of variations to obtain its gradient. The second issue which should be taken care of is that, during the optimization steps, we have to be careful not to leave the set of tight frames. Thus, we need to characterize the projector onto the space of tight wavelet frames. Having these two major components, the outline of the optimization algorithm is given in Algorithm 1, in which $\nabla V(h)(\rho)$ denotes the gradient of V at ρ, and $\mathcal{P}$ denotes the orthogonal projector onto the space of tight wavelet frames. This corresponds to the

standard projected gradient descent algorithm in an infinite dimensional space [28].

A. Measures of Localization

We now propose four measures of localization split between two natural classes. The first class consists of measures of the variance. Variance is the most well-known measure of localization as less variance implies more concentration around the center. In addition, we know from the uncertainty principle that the best achievable localization of a function is inversely proportional to the localization of its Fourier transform, and vice versa. More precisely, for a function $\psi : \mathbb{R}^2 \to \mathbb{R}$, we have [29]

$$\frac{\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^2} \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}} \cdot \frac{\int_{\mathbb{R}^2} \|\boldsymbol{\omega}\|^2 |\mathcal{F}\{\psi\}(\boldsymbol{\omega})|^2 \mathrm{d}\boldsymbol{\omega}}{\int_{\mathbb{R}^2} |\mathcal{F}\{\psi\}(\boldsymbol{\omega})|^2 \mathrm{d}\boldsymbol{\omega}} \geq \frac{1}{16\pi^4}. \quad (7)$$

The first term of the left-hand side is the variance of the wavelet itself, and the second term is the variance of its Fourier transform. Thus, setting an upper bound on the variance in one domain imposes a lower bound on the variance in the other domain. In practice, we are interested in bandlimited wavelets, which implies that the variance of the wavelet in the Fourier domain is bounded from above. Thus, the variance of the wavelet in the space domain is bounded from below. Since we are interested in finding the wavelet profile that attains the minimum value for the variance, our first measure of localization is given by

$$V_{2\mathrm{D}}(\psi) = \frac{\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^2} \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}} = \frac{\int_0^\infty r^2 \psi(r)^2 \; r \mathrm{d}r}{\int_0^\infty \psi(r)^2 \; r \mathrm{d}r}. \quad (8)$$

Isotropic wavelets are often used in a directional framework, for instance by applying the Riesz transform or by applying an angular mask. This suggests that the variance of the 1D radial profile of the isotropic wavelet can also be a good candidate for the measure of localization. Hence, we propose the second variance-based measure of localization

$$V_{1\mathrm{D}}(\psi) = \frac{\int_0^\infty r^2 \psi(r)^2 \mathrm{d}r}{\int_0^\infty \psi(r)^2 \mathrm{d}r}. \quad (9)$$

It measures the spread of the wavelet along each radial line.

The second class of measures focuses on the localization of the wavelet coefficients rather than that of the wavelet profile. More precisely, the energy of a function computed over some spatial neighbourhood should be well represented by the wavelet coefficients associated to that neighbourhood and its vicinity. According to [22], if $f = \sum_{\boldsymbol{m} \in \mathbb{Z}^2} f_{\boldsymbol{m}}$ is an L_2-function from $\mathbb{R}^2$ to $\mathbb{R}$ and $f_{\boldsymbol{m}}$ is the restriction of f to the unit square centered at $\boldsymbol{m}$, then

$$|\langle f_{\boldsymbol{m}}, \psi_{i,\mathbf{k}} \rangle| \leq C_i \left(\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x} \right)^{1/2} \|f_{\boldsymbol{m}}\|_2 |\mathbf{k} - \boldsymbol{m}|^{-1}, \quad (10)$$

where C_i is a constant that only depends on the scale i and is independent of ψ and f. The ℓ_1-norm of a vector is denoted as $|\cdot|$ and the standard L_2-norm of a function as $\|\cdot\|_2$. Thus, as the wavelet $\psi_{i,\mathbf{k}}$ gets further from position $\boldsymbol{m}$, the contribution of $f_{\boldsymbol{m}}$ in the corresponding wavelet coefficient decays. Moreover, the rate of decay is controlled by a constant that corresponds to the unnormalized variance of the wavelet profile ψ. We hence propose this value as a third measure of localization, this time for the wavelet coefficients. We define

$$U_{2\mathrm{D}}(\psi) = \int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x} = \int_0^\infty r^2 \psi(r)^2 \; r \mathrm{d}r. \quad (11)$$

Accounting for the fact that isotropic wavelets are often used in a directional setting, we define the unnormalized variance of the 1D radial profile of the wavelet in analogy to (9) as our last measure of localization

$$U_{1\mathrm{D}}(\psi) = \int_0^\infty r^2 \psi(r)^2 \mathrm{d}r. \quad (12)$$

To summarize, the first class of measures focuses on the localization of the shape of the wavelet in the space domain while the second class of measures describes the localization of the wavelet coefficients. We thus expect the first class to match applications that benefit from a local wavelet analysis. The second class should, on the contrary, be more appropriate in the context of applications that involve some form of wavelet-domain N-term approximation. For both classes of measures, the index 2D indicates that the spread of the wavelet is a measure over the whole space. The index 1D, conversely, measures the spread in each radial direction. In applications where a directionality component is imposed on top of the isotropic profile (*e.g.*, in detection tasks using steerable filters), we predict that profiles with the best radial localization should exhibit the best performances.

B. Gradient of the Functionals and Projector Onto the Space of Tight Frames

In this subsection, we give the explicit expressions of the two major components that are required for implementing Algorithm 1: the gradient of the localization criterion and the operator $\mathcal{P}$ that projects a solution onto the space of tight wavelet frames.

For formalization purpose, we consider h as a function in the weighted L_2-space of $[0, \infty]$, $\mathcal{L}_w$, whose inner product is defined as

$$\langle f, g \rangle_w = \int_0^\infty f(\rho) g(\rho) w(\rho) \mathrm{d}\rho, \quad (13)$$

where w is a strictly positive weighting function.

We shall see that the inclusion of such a weight will provides us with some degrees of freedom to design the projection operator, which can then be used advantageously to simplify the implementation. Since we are operating in a Hilbert space, the choice of w specifies the metric and hence the form of the orthogonal projection operator $\mathcal{P}_w : \mathcal{L}_w \to \mathcal{T}$, where $\mathcal{T}$ is the set of functions satisfying the tight frame property (6). Now, if we switch to another weighting function v, we can define another "orthogonal" projection operator $\mathcal{P}_v : \mathcal{L}_v \to \mathcal{T}$ which is such that $\mathcal{P}_v h = h$ for all $h \in \mathcal{T}$ as well as $\mathcal{P}_v \mathcal{P}_v f = \mathcal{P}_v f$ for all $f \in \mathcal{L}_w$. In other words, $\mathcal{P}_v$ also constitutes a valid projector for the space $\mathcal{L}_w$ (including $L_2[0, \infty]$ with $w = 1$), albeit not necessarily the one that minimizes the corresponding approximation error. The important point here

is that using $\mathcal{P}_v$ rather than $\mathcal{P}_w$ will not modify the outcome of the optimization process. Another way to put it is that the underlying Karush-Kuhn-Tucker conditions of optimality in the Lagrange multiplier method of optimization (see [29]) are independent of the actual choice of the Hilbert space $\mathcal{L}_w$.

Now, defining

$$A_m(h) = \int_0^\infty r^m \mathcal{H}\{h\}(r)^2 \mathrm{d}r, \tag{14}$$

we see that $V_{2\mathrm{D}}$, $V_{1\mathrm{D}}$, $U_{2\mathrm{D}}$, and $U_{1\mathrm{D}}$ can be easily written in terms of A_m for different m. Thus, we only need to compute the gradient of A_m. Using the basic rules of differentiation, we have

$$\nabla V_{2\mathrm{D}}\{h\} = \frac{1}{A_1(h)}\nabla A_3\{h\} - \frac{A_3(h)}{A_1(h)^2}\nabla A_1\{h\}, \tag{15}$$

$$\nabla V_{1\mathrm{D}}\{h\} = \frac{1}{A_0(h)}\nabla A_2\{h\} - \frac{A_2(h)}{A_0(h)^2}\nabla A_0\{h\}, \tag{16}$$

$$\nabla U_{2\mathrm{D}}\{h\} = \nabla A_3\{h\} \tag{17}$$

and

$$\nabla U_{1\mathrm{D}}\{h\} = \nabla A_2\{h\}. \tag{18}$$

The main point for our purpose is that the functional $A_m : \mathcal{L}_w \to \mathbb{R}$ is Gâteaux differentiable and that its infinite-dimensional gradient can be obtained explicitly as shown in Appendix A. The ultimate outcome is

$$\nabla A_m\{h\}(\rho) = \frac{2\rho}{w(\rho)} \int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(\rho r)\mathrm{d}r, \tag{19}$$

where $\mathcal{H}\{h\}$ is the Fourier-Hankel transform of h. Note that this functional gradient depends on the weighting function of the space, w.

The final ingredient for our algorithm is the operator that projects a function onto the set of tight-frames. Here, unlike in the case of the gradient, the computational complexity of the orthogonal projector is strongly dependent on the choice of w. In fact, in the following theorem, we will see that there is a very specific weighting function $w = v$ for which we can have a closed-form formula for the required projector.

Theorem 1: Let $\mathcal{L}_w$ be the Hilbert space whose inner product is specified by (13) and let $\mathcal{T}$ be the set of functions in $\mathcal{L}_w$ satisfiying the tight frame property (6). Then, the operator

$$\mathcal{P}\{\tilde{h}\}(\rho) = \frac{\tilde{h}(\rho)}{\sqrt{\sum_{i\in\mathbb{Z}} \tilde{h}^2(2^i\rho)}} \tag{20}$$

is a projector from $\mathcal{L}_w \to \mathcal{T}$. In particular, it is the orthogonal projector $\mathcal{L}_v \to \mathcal{T}$ for the weighting function

$$v(t) = 2^i \quad \text{for } \frac{\pi}{2^{i+1}} \le t \le \frac{\pi}{2^i}. \tag{21}$$

The proof is given in Appendix B.

The form (20) of the projector is intuitively very reasonnable. The simplification results from the choice of the appropriate metric in the proof of the theorem. The result is non-obvious a priori because this is the only instance of w for which we are able to carry out the computation to the end. Constraint (6) is the equation of the unit infinite-dimensional sphere for each value of ρ. The projector (20) is therefore projecting $\tilde{h}$ on the unit sphere for each value of ρ. The theorem ensures that we have an equivalence between the ℓ_2-norm projection in the space of sequences corresponding to each value of ρ and a projection in the weighted L_2-space of functions on $\mathbb{R}$ for the weighting function v specified by (21).

C. Numerical Optimization

In practice, we are interested in a fast implementation of wavelet transforms relying filter-banks. Focusing on wavelets that are supported on $[\frac{\pi}{4}, \pi]$ [2], the highpass h_H and lowpass h_L filters of the filter-bank are simply given by

$$h_H(\boldsymbol{\omega}) = \begin{cases} 0 & \|\boldsymbol{\omega}\| \le \dfrac{\pi}{4} \\ h(\|\boldsymbol{\omega}\|) & \dfrac{\pi}{4} < \|\boldsymbol{\omega}\| < \dfrac{\pi}{2} \\ 1 & \dfrac{\pi}{2} \le \|\boldsymbol{\omega}\|, \end{cases} \tag{22}$$

$$h_L(\boldsymbol{\omega}) = \begin{cases} 1 & \|\boldsymbol{\omega}\| \le \dfrac{\pi}{4} \\ h(2\|\boldsymbol{\omega}\|) & \dfrac{\pi}{4} < \|\boldsymbol{\omega}\| < \dfrac{\pi}{2} \\ 0 & \pi < \|\boldsymbol{\omega}\|, \end{cases} \tag{23}$$

respectively. According to the tight-frame constraint (6), we automatically have that

$$h_L^2(\boldsymbol{\omega}) + h_H^2(\boldsymbol{\omega}) = 1. \tag{24}$$

Hence, when looking for the minimizer of a given criterion, we optimize the values of the function $h(\rho)$ on the interval $[\frac{\pi}{4}, \pi]$. It is worth noting that a filter-bank construction for 1D tight wavelet frames with arbitrary dilation has been proposed in [30]. The resulting wavelets have more frequency contents than the ones we consider in this work and are thus more localized in time domain. However, due to the lack of a simple relation between the filters and the wavelet profiles, investigation and optimization of the localization of such profiles is beyond the scope of our paper.

We apply our numerical optimization algorithm for each of the proposed measures of localization. First of all, it has been shown analytically [22] that the Simoncelli wavelet minimizes the criterion $U_{2\mathrm{D}}$. Thus, we already know the optimal profile with respect to the measure $U_{2\mathrm{D}}$. Regarding $V_{2\mathrm{D}}$, $V_{1\mathrm{D}}$, and $U_{1\mathrm{D}}$, we run Algorithm 1 by uniformly taking 512 samples of $h(\rho)$ for ρ in $\pi/4$ to π on a logarithmic scale, hence simplifying the computation of the projection map (20). To calculate $\mathcal{H}\{h\}(r)$, we compute the integral of (5) from 0 to 300 relying on the trapezoid method with 3,000 intervals. The algorithm is left to run until absolute variations of $V(h)$ fall under 10^{-3}.

In this optimization settings, we obtain the minimum values 1.73, 0.39, and 1.64 for $V_{2\mathrm{D}}(h)$, $V_{1\mathrm{D}}(h)$, and $U_{1\mathrm{D}}(h)$, respectively. However, due to the Gibbs phenomenon that results from the truncation of the Fourier transform, the resulting wavelets exhibit ringing artifacts. In order to remove these effects and obtain a smooth profile for practical applications, we fitted a closed-form formula to the numerically obtained wavelets. We thus propose four wavelets named $h_{V_{2\mathrm{D}}}$, $h_{U_{2\mathrm{D}}}$, $h_{V_{1\mathrm{D}}}$, and $h_{U_{1\mathrm{D}}}$, which correspond to each of the considered measures of localization. From (22), (23), and (24), it is

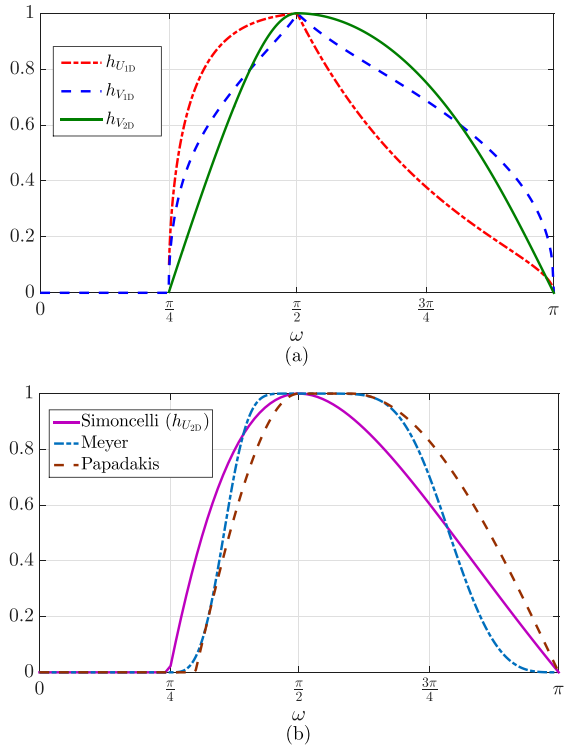


Fig. 1. Wavelet profiles in Fourier domain for (a) the proposed optimal profiles, and (b) existing ones (Simoncelli, Meyer, and Papadakis).

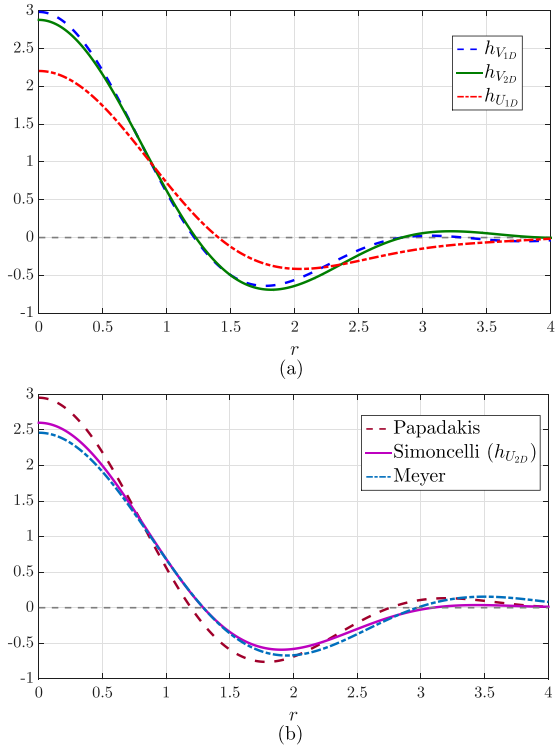


Fig. 2. Wavelet profiles in space domain for (a) the proposed optimal profiles, and (b) existing ones (Simoncelli, Meyer, and Papadakis).

sufficient to specify the wavelet profile either on $[\frac{\pi}{4}, \frac{\pi}{2}]$ or on $[\frac{\pi}{2}, \pi]$ to describe it entirely. The expressions of $h_{V_{2D}}$ and $h_{U_{2D}}$ are more easily given on the interval $[\frac{\pi}{4}, \frac{\pi}{2}]$ as

$$h_{V_{2D}}(\rho)\big|_{\rho\in[\frac{\pi}{4},\frac{\pi}{2}]} = \frac{\sqrt{6} - \sqrt{1 + 20(\frac{2\rho}{\pi} - 1)^2}}{\sqrt{6} - 1}, \quad (25)$$

$$h_{U_{2D}}(\rho)\big|_{\rho\in[\frac{\pi}{4},\frac{\pi}{2}]} = \cos\left(\frac{\pi}{2}\log_2\frac{2\rho}{\pi}\right). \quad (26)$$

We recall that $h_{U_{2D}}$ corresponds to the Simoncelli wavelet. The profiles $h_{V_{1D}}$ and $h_{U_{1D}}$ are better expressed on the interval $[\frac{\pi}{2}, \pi]$ as

$$h_{V_{1D}}(\rho)\big|_{\rho\in[\frac{\pi}{2},\pi]} = \left(\log_2\frac{\pi}{\rho} - 0.005\sqrt{\frac{\pi}{\rho}}\sin\left(\pi\log_2\frac{\pi}{\rho}\right)\right)^{\frac{2}{5}}, \quad (27)$$

$$h_{U_{1D}}(\rho)\big|_{\rho\in[\frac{\pi}{2},\pi]} = \sqrt{\frac{(\log_2\frac{\pi}{\rho} + 0.6)^4 - 0.6^4}{1.6^4 - 0.6^4}}. \quad (28)$$

The radial profiles of these wavelets are shown in Figures 1 and 2 in Fourier and space domains, respectively. We have that $V_{2D}(h_{V_{2D}}) = 1.74$, $V_{1D}(h_{V_{1D}}) = 0.40$, and $U_{1D}(h_{U_{1D}}) = 1.65$. These values are only marginally suboptimal. The values of the different measures of localization for each of these wavelets as well as for more traditional ones are given in Table I. Moreover, we note that Figure 2 confirms our expectation that $h_{V_{1D}}$ has the most localized profile shape.

The measures V_{2D} and U_{2D} can also be interpreted as the normalized and unnormalized third-order moment of the radial profile of the wavelet, respectively, while V_{1D} and U_{1D} correspond to its normalized and unnormalized second-order moment. Furthermore, having finite values for higher-order moments in the space domain implies being smoother in the Fourier domain. Thus, the minimizers of V_{2D} and U_{2D} necessarily have finite V_{1D} and U_{1D} values. However, the converse is not always true.

TABLE I
LOCALIZATION OF DIFFERENT WAVELETS MEASURED BY V_{2D}, V_{1D}, U_{2D}, AND U_{1D} (EQUATIONS (8)-(9) AND (11)-(12))

Wavelet type	Localization			
	V_{2D}	V_{1D}	U_{2D}	U_{1D}
$h_{V_{2D}}$ (25)	**1.74**	0.44	3.88	2.19
$h_{V_{1D}}$ (27)	$+\infty$	**0.40**	$+\infty$	2.03
$h_{U_{2D}}$ (Simonceli) (26)[8]	1.84	0.46	**3.55**	1.93
$h_{U_{1D}}$ (28)	$+\infty$	0.52	$+\infty$	**1.65**
Papadakis [19]	2.06	0.49	4.93	2.52
Meyer [18]	2.88	0.66	6.04	2.61
Shannon [3]	$+\infty$	$+\infty$	$+\infty$	$+\infty$

As we see in Table I, the minimizers of V_{1D} and U_{1D} have infinite values for V_{2D} and U_{2D}. This is in accordance with the roughness of the profiles at points $\frac{\pi}{4}$, $\frac{\pi}{2}$, and π (see Figure 1). In the case of the Shannon wavelet, the discontinuities of the frequency-domain profile results in a slow decay in the spatial domain that brings V_{1D} and U_{1D} to infinity. From Table I, we observe that the Simoncelli wavelet is not only optimal for the criterion U_{2D}, but also exhibits reasonable values for other measures of localization.

TABLE II

ESTIMATION OF LOCAL ORIENTATIONS FOR THE REGIONS OF INTEREST DEPICTED IN FIGURE 3B

Wavelet type	Angle [°]					
	1	2	3	4	5	6
Ground Truth	*95.36*	*129.09*	*83.88*	*166.50*	*33.02*	*13.24*
$h_{V_{2D}}$ (25)	95.88	133.19	80.48	164.66	36.74	15.95
$h_{V_{1D}}$ (27)	96.08	133.59	79.14	163.13	37.44	15.90
$h_{U_{2D}}$ (Simonceli) (26)[8]	96.42	133.75	78.37	164.55	37.72	16.67
$h_{U_{1D}}$ (28)	**95.57**	**132.88**	**83.19**	**165.07**	**36.12**	**15.53**
Papadakis	97.31	134.91	79.46	163.39	41.65	18.04
Meyer	97.15	134.67	79.65	164.25	41.61	18.75
Shannon	96.78	136.61	76.97	136.10	70.73	22.16

TABLE III

ERROR IN LOCAL ORIENTATION ESTIMATES FOR THE REGIONS OF INTEREST DEPICTED IN FIGURE 3B

Wavelet type	Absolute error [°]					
	1	2	3	4	5	6
$h_{V_{2D}}$ (25)	0.52	4.10	3.41	1.85	3.72	2.71
$h_{V_{1D}}$ (27)	0.73	4.49	4.75	3.38	4.41	2.66
$h_{U_{2D}}$ (Simonceli) (26)[8]	1.07	4.66	5.52	1.95	4.69	3.43
$h_{U_{1D}}$ (28)	**0.22**	**3.79**	**0.69**	**1.43**	**3.09**	**2.29**
Papadakis	1.95	5.82	4.42	3.11	8.62	4.80
Meyer	1.79	5.57	4.24	2.26	8.59	5.51
Shannon	1.43	7.51	6.91	30.41	37.71	8.92

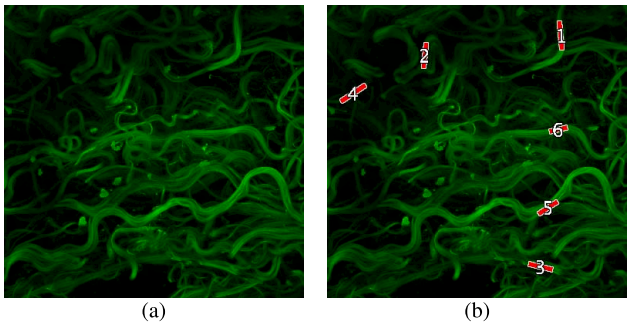
IV. APPLICATIONS

We now demonstrate the benefit of well-localized wavelet profiles for practical applications. In particular, we study the performance of our wavelets and compare them against other existing popular profiles for the problems of local orientation estimation, image reconstruction from edges and denoising.

A. Estimation of Local Orientation

In this section, we focus on the task of estimating the local orientation of ridge-like objects (*e.g.*, filaments) using a steerable ridge detector. We construct Hessian-like wavelets, the design of which is made easy by selecting an isotropic kernel and applying the generalized Riesz-wavelet transform using the appropriate shaping matrix, as described in [4, Sec. 5.1.3]. We refer the reader to [31] for a detailed description on how to steer Hessian filters to retrieve the orientation corresponding to the best response of the ridge detector at every point of the image. To perform a multiscale ridge detection at every location using the Hessian filter, we go through every scale of the wavelet pyramid and select the one where the strongest filter response can be found. The final output of our experiment is therefore an angle map with the same dimensions as the input image, and which contains at each pixel the estimated local orientation yielding the best ridge filter response. The Riesz-wavelet transform [4] and the extraction of local orientation estimation for each point of the input image have been implemented as a Java-based plug-in for the open-source image-analysis software ImageJ [32].[1]

In our experiment, we rely on a 512×512 pixels 8-bits image (Figure 3a) in which several regions of interest (ROIs) made of short line segments have been manually selected by an expert, and where local orientation should be estimated. The angle that each of the manually placed ROI form with the horizontal direction is considered as ground truth and corresponds to the orientation that shall automatically be retrieved. We estimate the local orientation of each ROI with the trivial isotropic profile (Shannon [3]), several popular isotropic wavelet profiles (Simoncelli [8], which corresponds to $h_{U_{2D}}$ (26), Papadakis [19], and Meyer [18]), as well as with the wavelets we propose ($h_{U_{1D}}$ (28), $h_{V_{2D}}$ (25) and $h_{V_{1D}}$ (27)). To obtain an estimation of the local orientation of each ROI, we average the orientation estimates provided by the steerable filter under the ROI (*i.e.*, we average the values of the pixels that belong to the line segment composing the ROI). We investigate the quality of each of the local orientation estimate by comparing the ground truth orientation with the automatically retrieved one (Table II). We also report the absolute error between the ground truth and each of the estimates in Table III. The experiment is conducted using 4 scales of wavelet decomposition.



(a) (b)

Fig. 3. Estimation of local orientation. (a) Original filaments image from [33], and (b) overlaid local regions of interest with their label.

[1]Source code available by request to the authors.

In this experiment, we observe that best results are obtained with $h_{U_{1D}}$. This can be explained by the fact that the two classes of wavelets h_U and h_V are optimized for different applications. The profiles of $h_{V_{2D}}$ and $h_{V_{1D}}$ are most localized in the spatial domain as they optimize V_{2D} (8) and V_{1D} (9), while $h_{U_{2D}}$ and $h_{U_{1D}}$ optimally localizes wavelet coefficients following U_{2D} (11) and U_{1D} (12). The estimation of local orientation is better when the wavelet response is strong and well localized, as ridges (here, the filaments) are then more accurately detected. In the present experiment, a profile maximizing wavelet coefficients localization, and hence a criterion of the class U, is therefore desirable. This is confirmed by the results and the good performance obtained by $h_{U_{1D}}$ and $h_{U_{2D}}$. Also, among the class of h_U wavelets, one observes that $h_{U_{1D}}$ outperforms $h_{U_{2D}}$. An argument for this is that a 1-dimensional design is more suitable for steerable wavelets since they are inherently directional. In the current application, the wavelets align themselves with ridges so that the U_{1D} criterion, which measures the spread in each radial direction, is the most appropriate one.

B. Image Reconstruction From Edges

The experiment we study here is image approximation from a reduced set of wavelet-based edges. First, a multiscale *primal sketch* [14], or *edge map* [15], is extracted from the set of wavelet coefficients of the image. An approximation of the original image is then recovered from this small subset of coefficients relying on constrained optimization.

To extract a multiscale edge map from the input image, we rely on a gradient-like wavelet framework. It is implemented with the help of the generalized Riesz-wavelet transform [4], and of an appropriate shaping matrix [4, Sec. 5.1.1] that yields a pair of x- and y-derivative wavelets. Edges in the multiscale gradient signal are then detected based on a wavelet-domain version of the Canny edge detector, which includes non-maximum suppression and hysteresis thesholding [14]. Note that the Canny edge detector requires an estimation of the strength and orientation of the gradient for each point of the image, which is obtained by steering the gradient-like wavelets at every scale and location. The final edge map is composed of the wavelet coefficients retained by the multiscale edge detector. To preserve the graylevel information of the image, all coefficients of the lowpass residual of the wavelet decomposition are saved.

Reconstruction is then formulated as the constrained optimization problem

$$\text{minimize } \|z\|_1 \tag{29}$$

$$\text{subject to } z = \mathbf{W}^H \boldsymbol{f} \tag{30}$$

$$z_{\boldsymbol{i}} = q_{\boldsymbol{i}}, \quad \forall \boldsymbol{i} \in S, \tag{31}$$

with variables z and $\boldsymbol{f}$, where S is the set of indices for the wavelet coefficients that are part of the edge map, $\mathbf{W}^H$ the wavelet-analysis operator, and $\boldsymbol{f}$ an image. Finally, $q_{\boldsymbol{i}}$ denotes the wavelet coefficient of the original image at location $\boldsymbol{i}$, where $\boldsymbol{i}$ is an index of 2-D position and scale. This formulation is motivated by two main principles. First, we aim at conserving the elements of the edge-map in order to reconstruct the image. We refer to them as wavelet-based edges, as they are the output of a Canny edge detector applied on the wavelet transform of the image. This gives us constraint (31), which imposes the elements $q_{\boldsymbol{i}}$, $\boldsymbol{i} \in S$ to be fixed during the optimization process. Second, we want the estimated missing wavelet coefficients to project back onto an image. Knowing that images are sparse in the wavelet domain, we impose sparsity by minimizing the ℓ_1-norm of the wavelet transform z of the image, yielding (29). Our problem thus amounts to minimizing a convex functional under a finite set of linear constraints, which guarantees the existence of a feasible minimum that can be reached using appropriate optimization algorithms.

Here, we propose an improvement of the reconstruction algorithm in [15] that relies on a gradient descent of the augmented Lagrangian. Our new algorithm is based on the alternating-direction method of multipliers (ADMM), which motivates the introduction of the auxiliary variable z. ADMM is a method known to converge very fast to an acceptable solution and that guarantees the residual to be brought to zero. In practice, it is observed that the fast and moderately good estimate provided after 30 iterations of the algorithm is already visually satisfactory. We refer the reader to [34] for a complete description of the method.

To reconstruct the image with ADMM, we first form the augmented Lagrangian

$$\mathcal{L}(z, \boldsymbol{f}, \boldsymbol{\lambda}) = \|z\|_1 + \boldsymbol{\lambda}^T \left(z - \mathbf{W}^H \boldsymbol{f}\right) + \frac{\mu}{2} \left\| z - \mathbf{W}^H \boldsymbol{f} \right\|_2^2, \tag{32}$$

where μ is a step size that can be adapted to influence the speed of convergence. We rewrite (32) in terms of the scaled dual variable $\boldsymbol{u} = \frac{\boldsymbol{\lambda}}{\mu}$ in order to obtain simpler mathematical expressions, yielding

$$\mathcal{L}(z, \boldsymbol{f}, \boldsymbol{u}) = \|z\|_1 + \frac{\mu}{2} \left\| z - \mathbf{W}^H \boldsymbol{f} + \boldsymbol{u} \right\|_2^2 - \frac{\mu}{2} \|\boldsymbol{u}\|_2^2. \tag{33}$$

The ADMM algorithm for our problem thus consists of the three successive iterations

$$z^{(k+1)} = \arg\min_z \mathcal{L}(z, \boldsymbol{f}^{(k)}, \boldsymbol{u}^{(k)}), \tag{34}$$

$$\boldsymbol{f}^{(k+1)} = \arg\min_{\boldsymbol{f}} \mathcal{L}(z^{(k+1)}, \boldsymbol{f}, \boldsymbol{u}^{(k)}), \tag{35}$$

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \left(z^{(k+1)} - \mathbf{W}^H \boldsymbol{f}^{(k+1)}\right). \tag{36}$$

The update for z can be rewritten in an element-wise manner as

$$z^{(k+1)} = \arg\min_z \left(\|z\|_1 + \frac{\mu}{2} \left\| z - \mathbf{W}^H \boldsymbol{f}^{(k)} + \boldsymbol{u}^{(k)} \right\|_2^2 \right) \tag{37}$$

$$= \arg\min_z \left(\sum_{\boldsymbol{i} \in I} |z_{\boldsymbol{i}}| + \frac{\mu}{2} \sum_{\boldsymbol{i} \in I} \left| z_{\boldsymbol{i}} - v_{\boldsymbol{i}}^{(k)} \right|^2 \right), \tag{38}$$

where we denote by I the set of indices for all wavelet coefficients of the image. We introduced $v_{\boldsymbol{i}}^{(k)} = [\mathbf{W}^H \boldsymbol{f}^{(k)}]_{\boldsymbol{i}} - u_{\boldsymbol{i}}^{(k)}$ for convenience. For all elements $\boldsymbol{i} \in S$, (31) imposes that $z_{\boldsymbol{i}} = q_{\boldsymbol{i}}$, and no further computations are required. For $\boldsymbol{i} \notin S$, $z_{\boldsymbol{i}}$ should be colinear with $v_{\boldsymbol{i}}$ in order to annihilate the

second term in (38). Plugging $z_i = Cv_i$ into (38) and solving for the optimal constant C brings us to the component-wise expression of the z update

$$z_i^{(k+1)} = \begin{cases} q_i, & \text{if } i \in S \\ \left(1 - \dfrac{1}{\mu \left| v_i^{(k)} \right|}\right)_+ v_i^{(k)}, & \text{if } i \notin S, \end{cases} \tag{39}$$

where $(\cdot)_+ = \max(0, \cdot)$ corresponds to the shrinkage of $v_i^{(k)}$. Then, updating f boils down to an unconstrained quadratic optimization problem. It can hence be performed by taking the partial derivative of the augmented Lagrangian and solving for zero. This yields

$$f^{(k+1)} = \arg\min_f \left\| z - \mathbf{W}^H f + u \right\|_2^2 \tag{40}$$

$$= (\mathbf{W}\mathbf{W}^H)^{-1}\mathbf{W}\left(z^{(k+1)} + u^{(k)}\right). \tag{41}$$

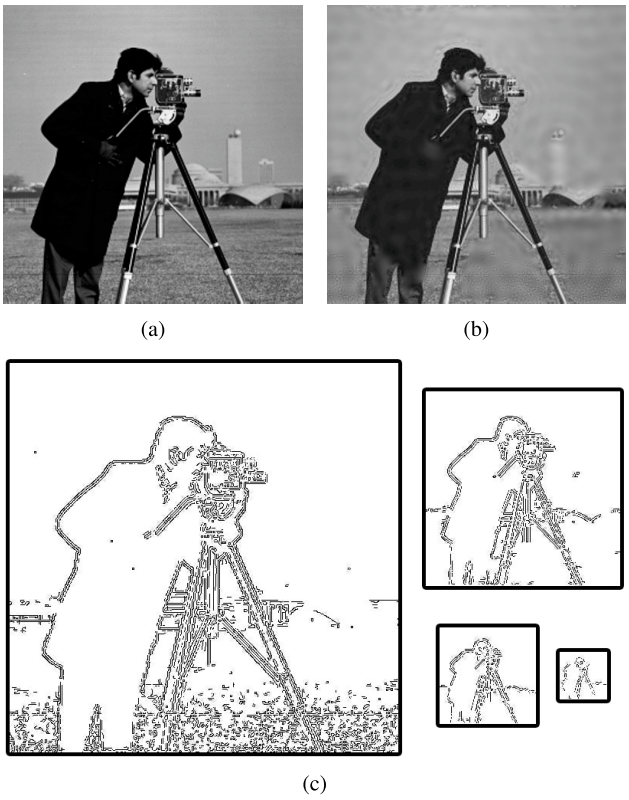
As $\mathbf{W}$ forms a tight frame, (41) can be further simplified by observing that $\mathbf{W}\mathbf{W}^H = \mathbf{I}$. Finally, u is simply modified through a linear update.

In practice, the algorithm is initialized with a $z^{(0)}$ composed of all the wavelet coefficients q_i, $i \in S$ retained in the edge map and the complete lowpass residual of the image to reconstruct. Then, $f^{(0)}$ is initialized as an image entirely composed of pixels with value zero and $\mathbf{W}^H f^{(0)}$ is obtained by taking its wavelet transform. Finally, $u^{(0)}$ is set as a pyramid of images composed only of zeros and having the same number of scales and dimensions as $\mathbf{W}^H f^{(0)}$. The parameter μ is empirically set to 10^6. With these settings, between thirty and fifty ADMM iterations were observed to be sufficient to reach "visual convergence", which corresponds to a situation where additional iterations bring unnoticeable visual improvements. Note that, through the whole optimization procedure, the left-multiplication by $\mathbf{W}^H$ or $\mathbf{W}$ corresponds to performing a wavelet analysis or synthesis, respectively. The algorithm can thus be executed in reasonable time as these two operations can be performed efficiently with the help of a filterbank implementation. To illustrate this experiment, we give in Figure 4 the original Cameraman image, its reconstruction from 7% of the wavelet coefficients, as well as the mask containing the coefficients retained by our multiscale wavelet-based edge detector.

We implemented[2] the Riesz-wavelet transform [4] as well as the edge-map extraction and subsequent image reconstruction using ADMM as a Java-based plug-in for the open-source image-analysis software ImageJ [32].

We gather results on a set of 5 standard test images, namely, House, Pirate, Peppers, Lena, Bridge, Cameraman, and Einstein, all being 512×512 pixels grayscale images. We run the same experiments with the trivial isotropic profile (Shannon [3]), several popular isotropic wavelet profiles (Simoncelli [8], which happens to be $h_{U_{2D}}$ (26), Papadakis [19], and Meyer [18]), and finally with our proposed wavelets ($h_{U_{1D}}$ (28), $h_{V_{2D}}$ (25) and $h_{V_{1D}}$ (27)). We investigate the reconstruction performance of the different wavelets in terms of the PSNR of the reconstructed image. We start by retaining 7% of the total number of wavelet coefficients in the image. These 7% are chosen among the set of wavelet-based edges retained by our multi-scale Canny edge detector.[3] Note that, as our test images all have the same size, this percentage corresponds to the same absolute number of coefficients in each case. All experiments are conducted using 4 scales of decomposition. Reconstruction results obtained after 30 iterations of the ADMM algorithm are shown in Table IV. In order to allow for a visual comparison of the performance, we also show in Figure 5 close-ups of the Lena image reconstructed using the different wavelet profiles. We here observe that $h_{U_{1D}}$ outperforms the other wavelets. Further experiments of reconstruction using a set of edge coefficients corresponding to 1 to 7% of the total number of wavelet coefficients in the image allows us to reach similar conclusions, as seen in Figure 6. Here, only results on Lena and Cameraman are displayed, as they are representative of the results observed in the remaining test images. This confirms that the proposed $h_{U_{1D}}$ profile is better for reconstruction than the other wavelets considered in this experiment.

Notice that, in this application, $h_{U_{1D}}$ followed by $h_{U_{2D}}$ outperforms in particular $h_{V_{2D}}$ and $h_{V_{1D}}$. As the construction of the edge map in the current experiment relies on the same kind



(a) (b) (c)

Fig. 4. Wavelet-based edge reconstruction. (a) Original Cameraman image, (b) final result after reconstruction using $h_{U_{1D}}$, and (c) binary masks featuring the wavelet coefficients saved for reconstruction at different scales. Here, 4 scales were used and 7% of the total number of coefficients were retained.

[2]Source code available by request to the authors.

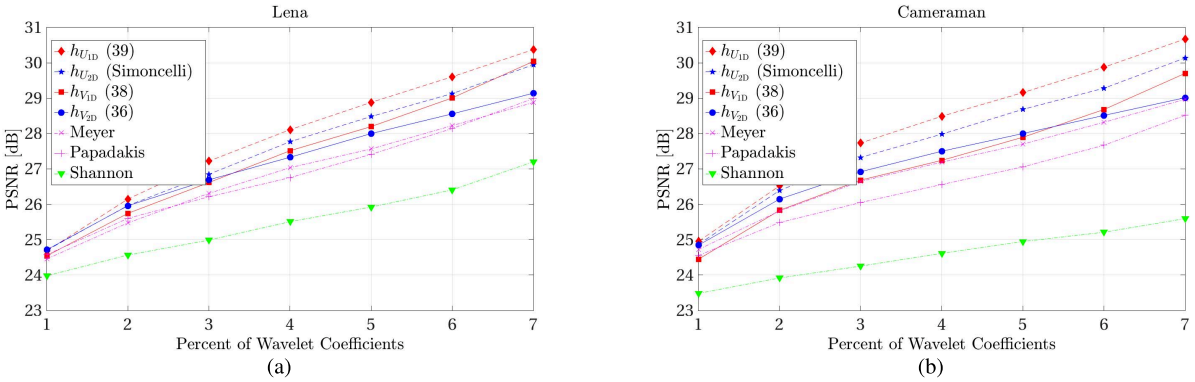
[3]In practice, we adapt the percentage of coefficients retained by the multi-scale Canny edge detector by changing the hysteresis thresholding parameters.

TABLE IV
RECONSTRUCTION FROM WAVELET-BASED EDGES (7% OF COEFFICIENTS)

Wavelet type	PSNR [dB]						
	House	Pirate	Peppers	Lena	Bridge	Cameraman	Einstein
$h_{V_{2D}}$ (25)	28.90	27.33	27.61	29.14	23.42	29.01	28.97
$h_{V_{1D}}$ (27)	29.72	27.32	27.96	30.04	23.99	29.70	29.28
$h_{U_{2D}}$ (26) (Simoncelli [8])	29.15	27.37	27.96	29.95	23.91	30.14	29.22
$h_{U_{1D}}$ (28)	**30.01**	**28.24**	**28.42**	**30.37**	**24.07**	**30.67**	**29.40**
Papadakis [19]	27.98	26.55	27.34	28.99	23.67	28.52	28.89
Meyer [18]	27.72	26.29	26.70	28.88	23.50	28.97	28.65
Shannon [3]	26.80	24.91	25.53	27.20	23.03	25.60	27.61

(a) (b) (c) (d) (e) (f) (g)

Fig. 5. Wavelet-based edge reconstruction. Close-up of reconstruction of Lena relying on (a) Shannon, (b) Meyer, (c) Papadakis, (d) $h_{U_{1D}}$, (e) $h_{U_{2D}}$ (Simoncelli), (f) $h_{V_{1D}}$, and (g) $h_{V_{2D}}$. The best PSNR is achieved by $h_{U_{1D}}$ shown in (e). Here, 4 scales were used and 7% of the total number of coefficients were retained.



(a) (b)

Fig. 6. Wavelet-based edge reconstruction. Evolution of the PSNR as a function of the percentage of retained wavelet coefficients using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.

of framework as the filaments detection for local orientation estimation in Section IV-A (namely, multiscale steerable filters), the same arguments hold for explaining these results. The reconstruction task obtains better results when the edge map contains sharper elements, which corresponds to better sets of edges. What matters most in this experiment is therefore again the optimal localization of the wavelet coefficients. This explains why the best performance is obtained with the profiles optimizing criterion of the class U ($h_{U_{1D}}$ and $h_{U_{2D}}$). Results can actually directly be interpreted from the values of U_{1D} provided in Table I. Starting from the optimal $h_{U_{1D}}$, the next most localized profiles sorted by distance to the optimum are Simoncelli ($h_{U_{2D}}$), $h_{V_{1D}}$, $h_{V_{2D}}$, Papadakis and Meyer. The quality of reconstruction obtained by the different profiles and shown in Figure 6 follow the same pattern, with Papadakis and Meyer being the worst and followed by $h_{V_{2D}}$ and $h_{V_{1D}}$, $h_{U_{2D}}$ (Simoncelli), and finally $h_{U_{1D}}$, which achieves the best results.

C. BLS-GSM Denoising

The BLS-GSM algorithm [7] is a famous and very efficient approach for recovering noise-corrupted images. The motivation behind this method is the observed strong correlation between wavelet coefficients located at similar positions at various orientations and scales. More specifically, the properties of the image (*i.e.*, the neighborhood of each wavelet coefficient) are modeled by a Gaussian scale mixture model (GSM) in the multiresolution wavelet transform space. The original, noise-free value of each coefficient is then estimated using Bayesian least squares (BLS) under the Gaussian scale mixture model, and making use of the correlation between coefficients in the pyramid. The algorithm therefore improves the denoising by taking advantage of local similarities. In practice, BLS-GSM is performed on subbands of an oriented multiresolution transformation of the noise-corrupted image, which corresponds in the original implementation to the steerable pyramid with Simoncelli's wavelet profile. We orient readers interested in a more detailed description of BLS-GSM to the very comprehensive paper of Rajaei [35].

A Matlab implementation of BLS-GSM has been released by the authors of the initial paper [7]. In order to perform the following experiments, we modify this original

TABLE V
BLS-GSM DENOISING AT $\sigma = 40$ WITH 4 ORIENTATIONS

Wavelet type	PSNR [dB]						
	House	Pirate	Peppers	Lena	Bridge	Cameraman	Einstein
$h_{V_{2D}}$ (25)	28.80	26.06	26.60	29.23	23.97	29.66	28.77
$h_{V_{1D}}$ (27)	**28.85**	**26.09**	**26.62**	**29.28**	**24.01**	**29.71**	**28.80**
$h_{U_{2D}}$ (26) (Simoncelli [8])	28.80	26.03	26.55	29.22	23.96	29.64	28.75
$h_{U_{1D}}$ (28)	28.70	25.98	26.38	29.13	23.94	29.60	28.69
Papadakis [19]	28.67	26.01	26.51	29.11	23.94	29.54	28.71
Meyer [18]	28.57	25.93	26.38	29.03	23.89	29.41	28.65
Shannon [3]	27.89	25.71	25.84	28.63	23.78	28.87	28.36

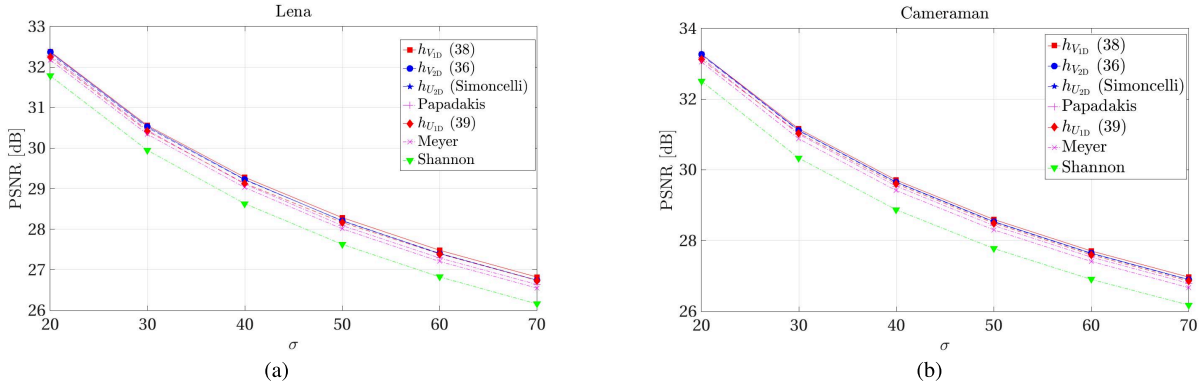


Fig. 7. BLS-GSM denoising. Evolution of the PSNR as a function of noise using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.

implementation[4] by replacing the Simoncelli wavelet by other wavelet profiles.

We perform several experiments in order to compare performance of different isotropic wavelet profiles when denoising with the BLS-GSM algorithm. We use the same image set as previously in IV-B. We also compare results between the same collection of wavelet profiles, namely Shannon, Meyer, Papadakis, Simoncelli ($h_{U_{2D}}$), $h_{U_{1D}}$, $h_{V_{2D}}$ and $h_{V_{1D}}$. In our first experiment, we corrupt the images with additive Gaussian noise of standard deviation $\sigma = 40$ and use the default number of orientations for the construction of the steerable pyramid, namely 4. We evaluate results in terms of the PSNR of the denoised image, as presented in Table V. From this, one observes that $h_{V_{1D}}$ (27) outperforms all other profiles. In order to further investigate this, we perform two additional experiments by varying the parameters of BLS-GSM. We first run the algorithm for each wavelet profile on images corrupted by noise levels with $\sigma \in [20, 70]$. Results are displayed in Figure 7. As expected, higher noise levels yield lower PSNR as the image becomes harder to retrieve. We also study the influence on the number of orientations chosen to built the steerable pyramid.[5] Increasing the number of orientations yield better results, but also significantly increases computation time. We show results for each of the studied wavelet profile in Figure 8. From these two experiments, we observe that $h_{V_{1D}}$, followed by $h_{V_{2D}}$, yields consistently better results than all the other popular profiles we tested, outperforming state-of-the-art results using the steerable pyramid. Although we only show here results on Lena and Cameraman due to space constraints, the same observation can be made using House, Pirate, Peppers, Bridge and Einstein.

Unlike $h_{V_{2D}}$ and $h_{V_{1D}}$, we observe that $h_{U_{1D}}$ yield less impressive results, and in particular does not compete with the Simoncelli profile ($h_{U_{2D}}$) initially used in BLS-GSM, which gives similar results than $h_{V_{2D}}$. This observation is consistent with the way each of the wavelets are constructed. As explained in Section IV-A, the h_U maximally localizes the wavelet coefficients, while the h_V have a profile which is optimally localized in spatial domain. The most desirable feature here, in comparison with local orientation estimation and image reconstruction from edges, is a spatially localized profile for the steerable pyramid. A transformation yielding very localized wavelet coefficients is actually even counter-productive as it concentrates the neighborhood of each coefficient, and thus reduces the amount of information that can be exploited by BLS-GSM. The profiles of choice for this application are hence of the class h_V. The values of V_{1D} that can be found in Figure 1 for the different profiles are consistent with these observations: $h_{V_{1D}}$ is optimal, then comes $h_{V_{2D}}$ and $h_{U_{2D}}$ (Simoncelli), which is the third closest to the optimal value. The Papadakis and $h_{U_{1D}}$ follow with V_{1D} values that are about equivalently far from the optimum, and the Meyer wavelet finally gets the worse V_{1D} localization value. This trend is conserved in our results, as seen in Figures 7 and 8. Sorting the profiles by result quality

[4]BLS-GSM Image Denoising Matlab Toolbox 1.0.3 (latest revision: February 23, 2005), available from http://www.io.csic.es/PagsPers/JPortilla/software/section/3-bayesian-denoising-in-the-wavelet-domain9/.

[5]The maximum number of orientations allowed by the Matlab implementation is 16.

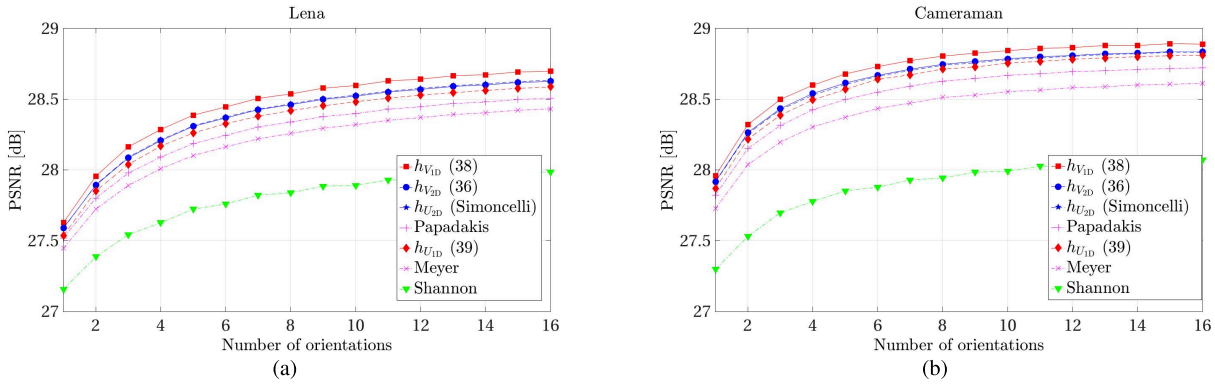


Fig. 8. BLS-GSM denoising. Evolution of the PSNR as a function of the number of angles using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.

(from best to worse) indeed yields $h_{V_{1D}}$, $h_{V_{2D}}$, $h_{U_{2D}}$ (Simoncelli), Papadakis, $h_{U_{1D}}$ and Meyer.

Note that although the difference in PSNR are marginal, the improvement comes at no cost as the only modification to the algorithm is a change of the radial wavelet profile. We also emphasize that these results do not imply that the proposed design should outperform denoising results based on other non-steerable wavelet frames. Rather, they indicate that the signal-domain localization of the wavelets is beneficial to the BLS-GSM algorithm.

V. Conclusion

In this paper, we have introduced a method for designing maximally localized isotropic tight-frame wavelets. The key ingredient is a measure of localization that can be optimized in order to derive the corresponding profile. We provide two classes of criterion for measuring localization either in the spatial or in the wavelet domain and express the resulting profiles optimizing each criterion either over the whole space or along each radial direction. We then consider three experimental settings in which we compare results obtained with our wavelets against state-of-the-art. First, we focus on the problem of estimating local orientation of filamentous structures, and then on the task of reconstructing an image from a small subset of edges in the wavelet domain. Both of these experiments rely on steerable filters, either Hessian- or gradient-based. In this setting, the wavelets obtained by optimizing the localization of wavelet coefficients outperforms existing isotropic wavelet profiles. Then, we demonstrate the efficiency of the most localized profiles in spatial domain in a denoising experiment using the popular BLS-GSM algorithm. These different use-cases show that both of our classes of localization criterion are relevant depending on the kind of application being considered, and that the proposed wavelet profiles are interesting candidates for image-processing tasks involving isotropic wavelets. Moreover, our results experimentally confirm the validity of the proposed localization criterion. The good or bad performance of the studied wavelet profiles can indeed be interpreted in the light of these two metrics. This further hints at the fact that our localization criterion are useful in order to study the localization of a given wavelet either in terms of its spatial profile, or of the coefficients it generates. In fact, the two proposed localization criteria can serve as quick estimates to assess the relative performance of any given isotropic tight-frame wavelet profile based on a simple calculation.

Appendix A
Computation of the Gradient

In order to obtain the gradient of the functional $A_m : \mathcal{L}_w \to \mathbb{R}$, we first calculate the Gâteaux derivative of A_m in the direction of a given function $g \in \mathcal{L}_w$ as

$$\begin{aligned} \mathrm{D}_g A_m(h) &= \frac{\partial}{\partial \epsilon} A_m(h+\epsilon g)\Big|_{\epsilon=0} \\ &= \int_0^\infty r^m \frac{\partial}{\partial \epsilon} \mathcal{H}\{h+\epsilon g\}(r)^2\Big|_{\epsilon=0} \mathrm{d}r. \end{aligned} \tag{42}$$

Due to the linearity of $\mathcal{H}\{\cdot\}$, we can write

$$\begin{aligned} &\mathrm{D}_g A_m(h) \\ &\quad= \int_0^\infty r^m \frac{\partial}{\partial \epsilon} \Big(\mathcal{H}\{h\}(r) + \epsilon \mathcal{H}\{g\}(r)\Big)^2\Big|_{\epsilon=0} \mathrm{d}r \\ &\quad= 2\int_0^\infty r^m\ \mathcal{H}\{h\}(r)\mathcal{H}\{g\}(r)\mathrm{d}r \\ &\quad= 2\int_0^\infty r^m \mathcal{H}\{h\}(r) \left(\int_0^\infty g(t) J_0(tr) t \mathrm{d}t\right) \mathrm{d}r \\ &\quad= 2\int_0^\infty g(t) t \left(\int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(tr) \mathrm{d}r\right) \mathrm{d}t. \end{aligned} \tag{43}$$

According to the definition of the gradient, we have that

$$\mathrm{D}_g A_m(h) = \int_0^\infty g(t)\ \nabla A_m\{h\}(t)\ w(t) \mathrm{d}t. \tag{44}$$

We therefore obtain

$$\nabla A_m\{h\}(\rho) = \frac{2\rho}{w(\rho)} \int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(\rho r) \mathrm{d}r. \tag{45}$$

Appendix B
Characterization of the Projector Onto the Space of Tight Frames

In order to characterize the orthogonal projector $\mathcal{P} : \mathcal{L}_w \to \mathcal{T}$, we have to solve a minimization problem.

For a given $\tilde{h} \in \mathcal{L}_w$, we are looking for a function $h \in \mathcal{T}$ that satisfies

$$\sum_{i\in\mathbb{Z}} |h(2^i\rho)|^2 = 1 \quad \text{for } \rho \in [\frac{\pi}{2}, \pi] \tag{46}$$

and that minimizes the functional

$$\|\tilde{h} - h\|_2^2 = \int_0^\infty (\tilde{h}(\rho) - h(\rho))^2 w(\rho) \mathrm{d}\rho. \tag{47}$$

We solve this problem using an infinite-dimensional Lagrange-multiplier method [36]. Similar to (42)-(45), we calculate the gradient of constraint (46) for each value of ρ as

$$\nabla \left\{ \sum_{i\in\mathbb{Z}} |h(2^i\rho)|^2 \right\} (t) = 2\frac{h(t)}{w(t)} \sum_{i\in\mathbb{Z}} \delta(t - 2^i\rho), \tag{48}$$

where δ is the Dirac delta distribution. The gradient of (47) is given by

$$\nabla \left\{ \|\tilde{h} - h\|_2^2 \right\} (t) = 2(h(t) - \tilde{h}(t)). \tag{49}$$

According to the Lagrange-multiplier method, there exists a function k supported on $[\frac{\pi}{2}, \pi]$ at the minimizer of (47) for which we have that [36]

$$\begin{aligned} h(t) - \tilde{h}(t) &= \int_0^\pi k(\rho) \frac{h(t)}{w(t)} \sum_{i\in\mathbb{Z}} \delta(t - 2^i\rho) \mathrm{d}\rho \\ &= \frac{h(t)}{w(t)} \sum_{i\in\mathbb{Z}} 2^i k(2^i t). \end{aligned} \tag{50}$$

Therefore,

$$h(t) = \frac{\tilde{h}(t)}{1 - \frac{1}{w(t)} \sum_{i\in\mathbb{Z}} 2^i k(2^i t)}. \tag{51}$$

Now, we fix the weight function w as

$$w(t) = 2^i \quad \text{for } \frac{\pi}{2^{i+1}} \le t \le \frac{\pi}{2^i}. \tag{52}$$

Equation (51) can now be simplified to

$$h(t) = \frac{\tilde{h}(t)}{1 - \sum_{i\in\mathbb{Z}} k(2^i t)}. \tag{53}$$

To determine k, we substitute (53) in (46) and obtain

$$\sum_{j\in\mathbb{Z}} \frac{\tilde{h}^2(2^j t)}{(1 - \sum_{i\in\mathbb{Z}} k(2^{i+j} t))^2} = 1. \tag{54}$$

Since the denominator of (53) is invariant to dilations by powers of 2, the denominator of (54) does not depend of j. It means that

$$1 - \sum_{i\in\mathbb{Z}} k(2^i t) = \Big(\sum_{j\in\mathbb{Z}} \tilde{h}^2(2^j t) \Big)^{\frac{1}{2}}. \tag{55}$$

By substituting (55) in (53), we get that

$$h(t) = \frac{\tilde{h}(t)}{\sqrt{\sum_{i\in\mathbb{Z}} \tilde{h}^2(2^i t)}}. \tag{56}$$

ACKNOWLEDGEMENT

The authors would like to thank Nicolas Chenouard for fruitful discussions regarding the problem of reconstructing an image from its wavelet-based edges. Also, we would like to thank John Paul Ward for helpful discussions about different mathematical aspects.

REFERENCES

[1] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[2] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 3. Washington, DC, USA, Oct. 1995, pp. 444–447.

[3] M. Unser, N. Chenouard, and D. Van De Ville, "Steerable pyramids and tight wavelet frames in $L_2(BBR^d)$," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2705–2721, Oct. 2011.

[4] M. Unser and N. Chenouard, "A unifying parametric framework for 2D steerable wavelet transforms," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 102–135, 2013.

[5] S. Fischer, F. Šroubek, L. Perrinet, R. Redondo, and G. Cristóbal, "Self-invertible 2D log-Gabor wavelets," *Int. J. Comput. Vis.*, vol. 75, no. 2, pp. 231–246, Nov. 2007.

[6] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 948–959, Jul. 2005.

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

[8] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–70, Oct. 2000.

[9] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with Gaussianized steerable pyramids," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2702–2718, Sep. 2006.

[10] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[11] H. Rabbani, "Image denoising in steerable pyramid domain based on a local Laplace prior," *Pattern Recognit.*, vol. 42, no. 9, pp. 2181–2193, Sep. 2009.

[12] F. Denis and A. Baskurt, "Multidirectional curvilinear structures detection using steerable pyramid," *J. Electron. Imaging*, vol. 13, no. 4, pp. 756–765, Oct. 2004.

[13] Z. Püspöki, V. Uhlmann, C. Vonesch, and M. Unser, "Design of steerable wavelets to detect multifold junctions," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 643–657, Feb. 2016.

[14] D. Van De Ville and M. Unser, "Complex wavelet bases, steerability, and the Marr-like pyramid," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2063–2080, Nov. 2008.

[15] N. Chenouard and M. Unser, "3D steerable wavelets in practice," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4522–4533, Nov. 2012.

[16] M. Papadakis, G. Gogoshin, I. A. Kakadiaris, D. J. Kouri, and D. K. Hoffman, "Nonseparable radial frame multiresolution analysis in multidimensions," *Numer. Funct. Anal. Optim.*, vol. 24, nos. 7–8, pp. 907–928, Jan. 2003.

[17] S. Held, M. Storath, P. Massopust, and B. Forster, "Steerable wavelet frames based on the Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 653–667, Mar. 2010.

[18] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[19] J. R. Romero, S. K. Alexander, S. Baid, S. Jain, and M. Papadakis, "The geometry and the analytic properties of isotropic multiresolution analysis," *Adv. Comput. Math.*, vol. 31, nos. 1–3, pp. 283–328, Oct. 2009.

[20] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992.

[21] P. Pad and M. Unser, "Optimality of operator-like wavelets for representing sparse AR(1) processes," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4827–4837, Sep. 2015.

[22] J. P. Ward, P. Pad, and M. Unser, "Optimal isotropic wavelets for localized tight frame representations," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1918–1921, Nov. 2015.
[23] P. Pad, V. Uhlmann, and M. Unser, "VOW: Variance-optimal wavelets for the steerable pyramid," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 2973–2977.
[24] A. Depeursinge *et al.*, "Optimized steerable wavelets for texture analysis of lung tissue in 3-D CT: Classification of usual interstitial pneumonia," in *Proc. IEEE 12th Int. Symp. Biomed. Imag., (ISBI)*, Brooklyn, NY, USA, Apr. 2015, pp. 403–406.
[25] M. D. Buhmann, *Radial Basis Functions*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[26] H. Sagan, *Introduction to the Calculus of Variations*. New York, NY, USA: Dover, 2012.
[27] Y. I. Alber, A. N. Iusem, and M. V. Solodov, "On the projected subgradient method for nonsmooth convex optimization in a Hilbert space," *Math. Programming*, vol. 81, no. 1, pp. 23–35, Mar. 1998.
[28] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," *J. Fourier Anal. Appl.*, vol. 3, no. 3, pp. 207–238, May 1997.
[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[30] İ. Bayram and I. W. Selesnick, "Frequency-domain design of overcomplete rational-dilation wavelet transforms," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2957–2972, Aug. 2009.
[31] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1007–1019, Aug. 2004.
[32] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with ImageJ," *Biophoton. Int.*, vol. 11, no. 7, pp. 36–42, 2004.
[33] R. Rezakhaniha *et al.*, "Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy," *Biomech. Model. Mechanobiol.*, vol. 11, nos. 3–4, pp. 461–473, Mar. 2012.
[34] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[35] B. Rajaei, "An analysis and improvement of the BLS-GSM denoising method," *Image Process. Line*, vol. 4, pp. 44–70, Apr. 2014.
[36] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY, USA: Wiley, 1969.

Pedram Pad (S'08) received the B.Sc. degrees in electrical engineering (communications and signal processing) and mathematical sciences (pure mathematics) and the M.Sc. degree in electrical engineering (communications and signal processing) from the Sharif University of Technology, Tehran, Iran, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Switzerland. His research interests include different aspects of information theory and signal processing, in particular applications of the former in the latter.

Virginie Uhlmann received the M.Sc. degree in bioengineering from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2012, where she is currently pursuing the Ph.D. degree with the Biomedical Imaging Group under the direction of M. Unser. She completed her master's thesis in the Imaging Platform with the Broad Institute, Cambridge, MA, under the supervision of A. Carpenter. She is working on applied problem related to image segmentation and tracking, and on approximation and spline theory. Her research interests include image processing, computer vision, machine learning, and life sciences. She received the competitive Excellence Fellowship at the master's level from EPFL in 2011 and 2012, and the Best Student Paper Award from the 2014 IEEE International Conference on Image Processing. She also received a Best Student Paper Award Nomination from the IEEE International Symposium on Biomedical Imaging in 2015.

Michael Unser (M'89–SM'94–F'99) was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, USA, from 1985 to 1997, conducting research on bioimaging. He is currently a Professor and the Director of the Biomedical Imaging Group with the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has authored over 250 journal papers on those topics. He has authored the book entitled *An Introduction to Sparse Stochastic Processes*, (Cambridge University Press, 2014) with P. Tafti. His current research interests include biomedical image processing. He is an EURASIP Fellow in 2009, and a member of the Swiss Academy of Engineering Sciences. He is a recipient of several international prizes, including three IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010). He was the Associate Editor-in-Chief of the IEEE TRANSACTIONS ON MEDICAL IMAGING from 2003 to 2005. He is currently a member of the Editorial Boards of *SIAM Journal on Imaging Sciences*, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and *Foundations and Trends in Signal Processing*. He is the Founding Chair of the Technical Committee on Bio Imaging and Signal Processing of the IEEE Signal Processing Society.