Published in IEEE transactions on image processing, April 2017, vol. 26, issue 4, pp. 1626-1636, which should be cited to refer to this work. DOI : 10.1109/TIP.2017.2655438

Steerable Wavelet Machines (SWM): Learning Moving Frames for Texture Classification

Adrien Depeursinge, Zsuzsanna Püspöki, John Paul Ward, and Michael Unser

***Abstract*—We present texture operators encoding class-specific local organizations of image directions (LOID) in a rotation-invariant fashion. The LOIDs are key for visual understanding, and are at the origin of the success of the popular approaches such as local binary patterns (LBP) and the scale-invariant feature transform (SIFT). Whereas LBPs and SIFT yield handcrafted image representations, we propose to learn data-specific representations of the LOIDs in a rotation-invariant fashion. The image operators are based on steerable circular harmonic wavelets (CHW), offering a rich and yet compact initial representation for characterizing natural textures. The joint location and orientation required to encode the LOIDs is preserved by using moving frames (MF) texture representations built from locally-steered image gradients that are invariant to rigid motions. In a second step, we use support vector machines to learn a multi-class shaping matrix for the initial CHW representation, yielding data-driven MFs called steerable wavelet machines (SWM). The SWM forward function is composed of linear operations (*i.e.*, convolution and weighted combinations) interleaved with non-linear *steermax* operations. We experimentally demonstrate the effectiveness of the proposed operators for classifying natural textures. Our scheme outperforms recent approaches on several test suites of the Outex and CUReT databases.**

***Index Terms*—Texture classification, feature learning, moving frames, support vector machines, steerability, rotation-invariance, illumination-invariance, wavelet analysis.**

I. Introduction

One major difference between texture and object recognition in natural images relates to the ability of vision systems to characterize local versus global scene layouts. Most natural textures do not follow global image layouts and can only be described in terms of arrangements and repetitions of local pattern ensembles or primitives [1]. These primitives can be characterized in terms of the local organization of image directions (LOIDs). The latter are key for visual understanding [2] and texture segregation [3] (see Figure 1). LOIDs have been leveraged in the literature to define [4] and discriminate texture classes [5–10]. They capture the joint information between positions and orientations in images. It is the difference in this coupling that makes images f_1 and f_2 in Figure 4 visually distinct while both images have the same global density of small horizontal and vertical bars.

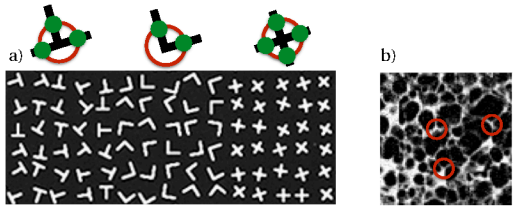

Fig. 1. Importance of the LOIDs in preattentive texture segregation [3]. a) X-shaped micropatterns (right) are easily separated from L-shaped ones (center), whereas T-shaped micropatterns (left) are found to be closer to L-shaped ones. The LOIDs can be distinguished by counting the number of endpoints of the primitives. b) texture associated with lung fibrosis in a CT scan. The LOIDs are characterized by junctions of collagen filaments.

The wealth of local texture patterns (*i.e.*, the LOIDs) is tightly related to the size of the observation window when the texture function $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^2$ is digitized on a discrete lattice indexed by $\boldsymbol{k} \in \mathbb{Z}^2$. In an extreme case, an image region composed of one pixel cannot form geometrical structures. Families of local image operators $g_i(\boldsymbol{x})$ can be designed to characterize the LOID subtypes (*e.g.*, edge or learned filters). Obtaining scalar texture measures often involves aggregating (*e.g.*, averaging) the outputs of local image operators $g_i(f(\boldsymbol{x}-\boldsymbol{m}))$ applied to $f(\boldsymbol{x})$ at the position $\boldsymbol{m} \in \mathbb{R}^2$ over an observation window $\boldsymbol{M}$ [11]. The latter raises two major challenges. First, the responses of the integrated operators becomes diffuse over $\boldsymbol{M}$, which hinders the spatial precision of texture segmentation approaches. Second, the effect of integration becomes even more destructive when unidirectional operators are jointly used to characterize the local organization of image directions (LOID) [10, 12] (*e.g.*, curvelets [13], co-occurrences [14], directional filterbanks [15, 16]). When separately integrated, the responses of unidirectional individual operators are not local anymore and their joint responses become only sensitive to the global density of image directions in $\boldsymbol{M}$. For instance, the joint responses of image gradients $g_{1,2}(f(\boldsymbol{x})) = \left(\left|\frac{\partial f(\boldsymbol{x})}{\partial x_1}\right|, \left|\frac{\partial f(\boldsymbol{x})}{\partial x_2}\right|\right)$ are not able to discriminate between the two textures classes $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ shown in Figure 4 when integrated over the full image domain $\boldsymbol{M}$.

An even bigger challenge is to design texture operators that can characterize the LOIDs in a rotation-invariant fashion [5, 7]. The latter is required to recognize image structures independently from both their local orientations and the global orientation of the image (see Figure 1). Examples of such structures are collagen meshes, vascular, bronchial or dendritic trees in biomedical images, river

The authors are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland (e-mail: adrien.depeursinge@epfl.ch; michael.unser@epfl.ch). A. Depeursinge is also with the MedGIFT group, Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre 3960, Switzerland. J. P. Ward is with the Department of Mathematics, University of North Carolina A&T, Greensboro, North Carolina 27411, USA.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes a Matlab toolbox for reproducing the experiments of this paper. Contact adrien.depeursinge@hevs.ch for further questions about this work.

deltas or urban areas in satellite images, complex biomedical tissue structures or crystals in petrographic analysis.

The above-mentioned imaging modalities yield images with normalized pixel sizes defined in physical units. The setting is therefore fundamentally different from photographic imagery resulting from scene captures obtained with varying viewpoints [17–19]. Since the spatial units are fixed, it is not desirable to enforce any form of scale invariance which truly entails the risk of regrouping patterns of different nature. More importantly, the scale is itself a powerful discriminative property. In this context, it is required to design texture operators that are invariant to the family of Euclidean transforms (also called rigid motions).

More generally, the rigid-motion invariant characterization of the joint location and orientation structure of texture (*i.e.*, the LOIDs) can be efficiently carried out using moving frames (MF) representations [8]. The key idea of MFs is to locally adapt a coordinate frame directly to a curve (*e.g.*, using the tangent as the first unit vector of the frame), rather than using extrinsic coordinates (see Figure 4). Image representations obtained from MFs can therefore be designed to be invariant to Euclidean transformations [20]. Moreover, deriving the local orientation of the frame tends to preserve the joint information between positions and orientations even when the operators are integrated (*e.g.*, averaged) over an image domain $\boldsymbol{M}$.

MFs have been used in computer vision to characterize the differential geometry of curves in Faugeras [20], and more specifically, to describe the perceptual organization of texture flows in Zucker *et al.* [8]. They were also referred to as "gauge coordinates" in [21]. They have been implictely used to characterize the LOIDs by popular approaches such as local binary patterns (LBP), maximum response of oriented filterbanks, and the scale-invariant feature transform (SIFT). LBPs [5] and their extensions [22–29] are specifically encoding the LOIDs in a rotation-invariant fashion with uniform circular pixel sequences. Extensions were proposed to include richer pixel dependencies based on local differences [22] and medians [29]. The maximum-response filterbank 8 (MR8) used the largest response of filters over various orientation only to locally normalize image directions [16]. Local discrete histogram of gradients (HOG) are used to encode the LOIDs in SIFT with approximate rotation-invariance [9, 17]. More recently, local continuous rotation-invariant HOGs were proposed by Liu *et al.* based on circular harmonic representations [10]. However, all of the above-mentioned methods are yielding handcrafted image descriptors that are not tailored to the specific image recognition task in hand. On the other hand, classical deep learning and dictionary learning approaches do not enforce the characterization of the LOIDs. They require learning similar kernel profiles at multiple orientations using data augmentation [30]. The scattering transform (ScatNet [12, 31]) is based on deep convolutional networks that are specifically designed to preserve the structure of the roto-translation group, but it does not yield data driven image representations.

In this work, we propose to bridge the gap between handcrafted MF-based features and learned representations with steerable wavelet machines (SWM). The cornerstone of our approach is to learn MF representations from locally steered linear combinations of circular harmonic wavelets (CHW) using support vector machines (SVM). CHWs are naturally encoding the LOIDs in terms of circular harmonics [32]. They provide continuous rotation-invariant versions of both LBPs [33] and HOGs [10]. Moreover, CHWs are encoding the LOIDs in a multi-resolution hierarchy and stand out as the canonical basis of steerable wavelet frames [34], providing ideally-suited initial representations for learning signal-adapted steerable wavelets. Based on the latter property, data-driven steerable wavelets are constructed from learned linear combinations of CHWs. Optimally discriminant features are constructed from the responses of the set of locally-oriented learned wavelets, yielding data-driven MF representations encoding the LOIDs with invariance to rigid motions.

The remainder of the paper is organized as follows. The SWM architecture is detailed in Section II-B. The mathematical foundations, construction and properties of steerable CHWs are detailed in Sections II-A, II-C, II-D and II-E. The construction steps of steerable CHW frames are (i) define a bandlimited isotropic mother wavelet that forms a frame on $L_2\left(\mathbb{R}^2\right)$ and (ii) apply the multi-order complex Riesz transform on it. Step (i) fixes the spatial supports (frequency bands) on top of which class-specific steerable wavelets can be learned from linear combinations of CHWs. The fundamentals of MFs are recalled in Section II-G. The learning of class-specifc MFs from shaped original CHW frames using SVMs is described in Section II-H. The behavior of SWMs and their ability to classify natural textures is evaluated and discussed in Sections III and IV, respectively.

II. Material and Methods

A. Notation

A point in the spatial domain $\mathbb{R}^2$ is represented by the vector variable $\boldsymbol{x}$, and by $\boldsymbol{\omega}$ in the Fourier domain. A 2-D function f is represented by $f(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{R}^2$, and by $f_{\text{pol}}(r,\theta)$ with $r \in \mathbb{R}^+$, $\theta \in [0, 2\pi)$, in the Cartesian and polar coordinate systems, respectively. In the Fourier domain, we use the notations $\hat{f}(\boldsymbol{\omega})$, with $\boldsymbol{\omega} \in \mathbb{R}^2$ and $\hat{f}_{\text{pol}}(\rho,\varphi)$ with $\rho \in \mathbb{R}^+$, $\varphi \in [0, 2\pi)$. The Fourier transform of an $L_1\left(\mathbb{R}^2\right)$ function f is computed according to

$$\widehat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{x}, \boldsymbol{\omega} \rangle} \mathrm{d}\boldsymbol{x}. \tag{1}$$

The average of $f(\boldsymbol{x})$ over the image domain $\boldsymbol{M}$ is noted $\overline{f(\boldsymbol{x})} = \frac{1}{m(\boldsymbol{M})} \int_{\boldsymbol{M}} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$, where $m(\boldsymbol{M})$ is the measure of $\boldsymbol{M}$.

B. Steerable Wavelet Machines

The architecture of SWMs is detailed in Figure 2. An image f_i is mapped to feature maps $\boldsymbol{t}_{i,\boldsymbol{x}}$ with a forward pass through the SWM layers. f_i is first convolved with the family of CHWs $\phi^{(n)}$. The resulting coefficients are mapped

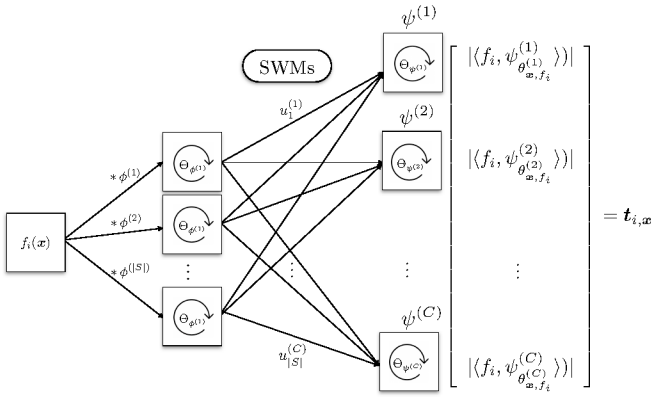

Fig. 2. Global architecture of SWMs. An input image f_i is mapped to output feature maps $\boldsymbol{t}_{i,\boldsymbol{x}}$ with a forward pass through the SWM layers. The forward function is composed of linear operations (*i.e.*, convolution and weighted combinations) interleaved by non-linear *steermax* operations denoted with circular arrows and angle maps $\Theta_{\phi,\psi}$.

to an initial gradient-based MF representation with a non-linear *steermax* operation based on the angle map $\Theta_{\phi^{(1)}}$ (denoted with circular arrows in Figure 2, see Eq. (18)). Class-wise templates $\psi^{(c)}$ are constructed from learned linear combinations $\boldsymbol{u}^{(c)}$ of CHWs in the gradient-based MF representation. A final *steermax* operation based on the learned angle maps $\Theta_{\psi^{(c)}}$ (following Eq. (23)) yields the final feature representation $\boldsymbol{t}_{i,\boldsymbol{x}}$. These feature maps can be further used by either a segmentation model, or aggregated over a region $\boldsymbol{M}$ and used by a classifier (*e.g.*, SVMs, *k*-nearest neighbors).

C. Isotropic Wavelet Frames

The construction of our steerable wavelet frames is initialized with a tight wavelet frame of $\mathbb{R}^2$, described by a mother wavelet ϕ whose translations and dilations generate the basis functions. The collection of isotropic bandpass filters ϕ controls the spatial support of the texture operators. In particular, at location (*i.e.*, grid point) $\boldsymbol{x}_{\boldsymbol{k}} = 2^s\boldsymbol{k}$, $\boldsymbol{k} \in \mathbb{Z}^2$, and scale s:

$$\phi_{s,\boldsymbol{k}}(\boldsymbol{x}) = \phi_s(\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{k}}) = \frac{1}{2^s}\phi\left(\tfrac{\boldsymbol{x}-\boldsymbol{x}_{\boldsymbol{k}}}{2^s}\right) = \frac{1}{2^s}\phi\left(\tfrac{\boldsymbol{x}}{2^s}-\boldsymbol{k}\right). \tag{2}$$

In the Fourier domain, (2) corresponds to

$$\begin{aligned}\hat{\phi}_{s,\boldsymbol{k}}(\rho,\varphi) = \widehat{\phi_s(\cdot-\boldsymbol{x}_{\boldsymbol{k}})}(\boldsymbol{\omega}) &= 2^s\hat{\phi}(2^s\boldsymbol{\omega})\mathrm{e}^{-\mathrm{j}\langle \boldsymbol{x}_{\boldsymbol{k}},\boldsymbol{\omega}\rangle} \\ &= 2^s\hat{\phi}(2^s\rho)\mathrm{e}^{-\mathrm{j}\rho_{\boldsymbol{k}}\rho\cos(\varphi-\varphi_{\boldsymbol{k}})}.\end{aligned} \tag{3}$$

Proposition 1 determines sufficient conditions for such a wavelet system.

Proposition 1 (*c.f.* [34, Proposition 4.1.])**.** *Let* $\hat{h}:[0,\infty)\to\mathbb{R}$ *be a smooth function satisfying:*

1) $\hat{h}(\rho) = 0$ *for* $\rho > \pi$ *(bandlimited),*
2) $\sum_{s\in\mathbb{Z}} \left|\hat{h}\left(2^s\rho\right)\right|^2 = 1,$
3) $\left.\frac{\mathrm{d}^n\hat{h}}{\mathrm{d}\rho^n}\right|_{\rho=0} = 0$ *for* $n = 0,\ldots,N$ *(vanishing moments).*

Using any norm p *as* $1 \le p \le \infty$*, the mother wavelet* ϕ *whose Fourier transform is given by*

$$\hat{\phi}(\boldsymbol{\omega}) = \hat{h}\left(\|\boldsymbol{\omega}\|_{\ell_p}\right) \tag{4}$$

generates a normalized tight wavelet frame of $L_2(\mathbb{R}^2)$ *whose basis functions*

$$\phi_{s,\boldsymbol{k}}(\boldsymbol{x}) = \phi\left(\boldsymbol{x}-2^s\boldsymbol{k}\right) \tag{5}$$

have vanishing moments up to order N*. In particular, any* $f \in L_2\left(\mathbb{R}^2\right)$ *can be represented as*

$$f = \sum_{s\in\mathbb{Z}}\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \langle f, \phi_{s,\boldsymbol{k}}\rangle\phi_{s,\boldsymbol{k}}. \tag{6}$$

As a particular example of such wavelets, in this work, we use Simoncelli's isotropic wavelet [35] defined by its radial frequency profile

$$\hat{h}(\rho) = \begin{cases} \cos\left(\frac{\pi}{2}\log_2\left(\frac{2\rho}{\pi}\right)\right), & \frac{\pi}{4} < \rho \le \pi \\ 0, & \text{otherwise.}\end{cases} \tag{7}$$

From the primal isotropic wavelet defined in this section we generate polar separable ones by the application of the multi-order complex Riesz transform.

D. The Multi-Order Complex Riesz Transform

The multi-order complex Riesz transform is used to obtain systematic representations of local circular frequencies, which are required to characterize the LOIDs. The first-order complex Riesz transform corresponds to the multi-dimensional extension of the Hilbert transform and was introduced in the literature by Larkin [36, 37]. The latter is defined in the Fourier domain as

$$\mathscr{R}f(\boldsymbol{x}) \leftrightarrow \frac{(\omega_x + \mathrm{j}\omega_y)}{\|\boldsymbol{\omega}\|}\hat{f}(\boldsymbol{\omega}) = \mathrm{e}^{\mathrm{j}\varphi}\hat{f}_{\mathrm{pol}}(\rho,\varphi). \tag{8}$$

Similarly to the Hilbert transform, it corresponds to a convolution-type operator that acts as an allpass filter. Its phase response is completely encoded in the orientation.

The Riesz transform is translation- and scale-invariant. More precisely,

$$\forall\,\boldsymbol{y}\in\mathbb{R}^2, \quad \mathscr{R}f(\cdot-\boldsymbol{y})(\boldsymbol{x}) = \mathscr{R}f(\cdot)(\boldsymbol{x}-\boldsymbol{y}) \tag{9}$$

$$\forall\,a\in\mathbb{R}^+, \quad \mathscr{R}f\left(\frac{\cdot}{a}\right)(\boldsymbol{x}) = \mathscr{R}f(\cdot)\left(\frac{\boldsymbol{x}}{a}\right). \tag{10}$$

The nth-order complex Riesz transform $\mathscr{R}^n$ is defined as the n-fold iterate of the complex Riesz transform $\mathscr{R}$. In the Fourier domain,

$$\mathscr{R}^n f(\boldsymbol{x}) \leftrightarrow \mathrm{e}^{\mathrm{j}n\varphi}\hat{f}_{\mathrm{pol}}(\rho,\varphi). \tag{11}$$

Isolated transform orders are orthogonal to each other. The higher order Riesz transform inherits the invariance properties of the complex Riesz transform, since they are preserved through iteration. Thus, it is scale- and translation-invariant, and provides a unitary mapping from an $L_2\left(\mathbb{R}^2\right)$ tight wavelet frame to another one.

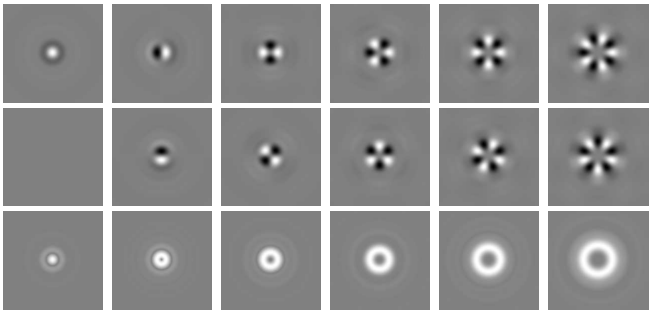

Fig. 3. Profiles of CHWs $\phi_{s,\boldsymbol{k}}^{(n)}$ for $n=0,\ldots,5$. Top, middle and bottom rows correspond the real, imaginary parts and absolute values, respectively.

E. Circular Harmonic Wavelet Frames

We apply the multi-order complex Riesz transform to a primal isotropic function that satisfies Proposition 1. The generated wavelet frames are called circular harmonic wavelets (CHW) and allow systematic characterizations of image scales and directions. We note that our CHWs are similar to ones of Jacovitti [32], with the difference that the latter ones are non-tight. The new wavelet functions are defined as $\phi^{(n)} := \mathscr{R}^n\phi$. More precisely, in Fourier, we have

$$\mathscr{F}\left\{\mathscr{R}^n\{\phi_s(\cdot-\boldsymbol{y})\}\right\}(\rho,\phi) = 2^s\hat{h}(2^s\rho)\mathrm{e}^{\mathrm{j}n\phi-\mathrm{j}\rho_0\rho\cos(\phi-\phi_0)}. \quad (12)$$

The n-channel tight wavelet frame is generated as $\{\phi_{s,\boldsymbol{k}}^{(n)} = \mathscr{F}^{-1}\{\hat{\phi}_{s,\boldsymbol{k}}^{(n)}\}\}_{n\in S}$. In this case, the elements of the distinct set S are called harmonics (corresponding to the exponentials). The nth-order CHW $\phi_{s,\boldsymbol{k}}^{(n)}$ has a rotational symmetry of order n around its center that corresponds to the nth-order rotational symmetry of $\mathrm{e}^{\mathrm{j}n\phi}$. CHWs are depicted in Figure 3 for $n=0,\ldots,5$.

The wavelets $\phi_{s,\boldsymbol{k}}^{(n)}$ form a tight wavelet frame, thus any finite-energy function f can be decomposed as

$$f = \sum_{n,s,\boldsymbol{k}} \left\langle f, \phi_{s,\boldsymbol{k}}^{(n)}\right\rangle \phi_{s,\boldsymbol{k}}^{(n)}. \quad (13)$$

A remarkable property of the CHWs is that of being self-steerable, where any rotation of $\phi_{s,\boldsymbol{k}}^{(n)}$ can be expressed as a linear combination of their own real and imaginary parts. More precisely,

$$\phi_{s,\boldsymbol{0},\theta_0}^{(n)}(\boldsymbol{x}) = \phi_{s,\boldsymbol{0}}^{(n)}(\mathbf{R}_{-\theta_0}\boldsymbol{x}) = e^{jn\theta_0}\phi_{s,\boldsymbol{0}}^{(n)}(\boldsymbol{x}), \quad (14)$$

where $\mathbf{R}_{-\theta_0} = \begin{bmatrix}\cos(\theta_0) & -\sin(\theta_0)\\ \sin(\theta_0) & \cos(\theta_0)\end{bmatrix}$. Therefore, any rotation of a multi-order CHW representation can be obtained with the block-diagonal steering matrix $\mathbf{A}_{\theta_0}$ as

$$\begin{bmatrix} \mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0},\theta_0}^{(1)}\right\rangle\right)\\ \mathrm{Im}\left(\left\langle f,\phi_{s,\boldsymbol{0},\theta_0}^{(1)}\right\rangle\right)\\ \vdots\\ \mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0},\theta_0}^{(n)}\right\rangle\right)\\ \mathrm{Im}\left(\left\langle f,\phi_{s,\boldsymbol{0},\theta_0}^{(n)}\right\rangle\right)\\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\theta_0) & -\sin(\theta_0) & & & & \\ \sin(\theta_0) & \cos(\theta_0) & & & & \\ & & \ddots & & & \\ & & & \cos(n\theta_0) & -\sin(n\theta_0) & \\ & & & \sin(n\theta_0) & \cos(n\theta_0) & \\ & & & & & \ddots \end{bmatrix}}_{\mathbf{A}_{\theta_0}} \begin{bmatrix} \mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(1)}\right\rangle\right)\\ \mathrm{Im}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(1)}\right\rangle\right)\\ \vdots\\ \mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(n)}\right\rangle\right)\\ \mathrm{Im}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(n)}\right\rangle\right)\\ \vdots \end{bmatrix}.$$

It can be noticed that $\mathbf{A}_{\theta_0}$ is sparse and the steering of multi-order representations requires much less computation when compared to other steerable wavelet representations with full steering matrices [34] (*e.g.*, real Riesz wavelets, Simoncelli's pyramid).

F. Texture Representations from CHWs

The absolute values of the collection of subbands provided by (13) yields a rich and compact representation for characterizing natural textures because it allows encoding the LOIDs for each position $\boldsymbol{x}$ and for a fixed scale s. The use of multi-order harmonics $n=0,\ldots,|S|$ provides a rich characterization of the local angular spectrum. The representation based on the complex modulus $\left|\left\langle f,\phi_{s,\boldsymbol{k}}^{(n)}\right\rangle\right|$ is rotation-invariant, but it discards the phase shifts between the harmonics. This is undesirable since two texture functions with different inter-harmonics phase shifts will be mixed. As an alternative, the representation based on real parts $\left|\mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{k}}^{(n)}\right\rangle\right)\right|$ preserves the phases between the harmonics. However, this representation has two major drawbacks for texture recognition. First it is not invariant to rotations, *i.e.*,

$$\forall\theta_0\neq 0,2\pi,\quad \begin{bmatrix} \left|\mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(0)}\right\rangle\right)\right|\\ \left|\mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(1)}\right\rangle\right)\right|\\ \vdots\\ \left|\mathrm{Re}\left(\left\langle f,\phi_{s,\boldsymbol{0}}^{(|S|)}\right\rangle\right)\right| \end{bmatrix} - \begin{bmatrix} \left|\mathrm{Re}\left(\left\langle f_{\theta_0},\phi_{s,\boldsymbol{0}}^{(0)}\right\rangle\right)\right|\\ \left|\mathrm{Re}\left(\left\langle f_{\theta_0},\phi_{s,\boldsymbol{0}}^{(1)}\right\rangle\right)\right|\\ \vdots\\ \left|\mathrm{Re}\left(\left\langle f_{\theta_0},\phi_{s,\boldsymbol{0}}^{(|S|)}\right\rangle\right)\right| \end{bmatrix} \neq \mathbf{0}, \quad (15)$$

where $f_{\theta_0} = f(\mathbf{R}_{-\theta_0}\boldsymbol{x})$. It will therefore not provide the same representation for two identical textures that are rotated versions of each other. This issue is addressed in Proposition 2 (see Section II-G). Second, it can hardly distinguish between texture classes that differ in terms of their LOIDs only when integrated over an image domain $\boldsymbol{M}$, since each element $\int_{\boldsymbol{M}}\left|\left\langle f,\phi_{s,\boldsymbol{k}}^{(n)}\right\rangle\right|\mathrm{d}\boldsymbol{x}$ is not local anymore. Both issues will be discussed in the next section, where solutions are proposed and exemplified for $n=1$ (*i.e.*, the gradient).

G. MF Representations from Locally Steered Gradients

In this section, we show how to analytically derive rotation- and translation-invariant texture representations from locally steered gradients (the gradient vector is equivalent to CHWs with $n=1$.) using moving frames. We also provide some evidence that the MFs tend to preserve the joint location and orientation structure of texture, which enables better characterization of the LOIDs when compared to using unaligned unidirectional texture operators.

Let $\{\mathbf{e}_1,\mathbf{e}_2\}$ be the canonical basis for $\mathbb{R}^2$, and let $\boldsymbol{x}$ denote the coordinates with respect to this basis; *i.e.*, $\boldsymbol{x}=(x_1,x_2)$ represents the point $x_1\mathbf{e}_1+x_2\mathbf{e}_2$. Let P be a rotation (by an angle θ_0) and translation (by a vector $\boldsymbol{y}$) of the plane $\mathbb{R}^2$. We consider a gray scale image $F:P\to\mathbb{R}$. We suppose that F can be evaluated as $f(\boldsymbol{x})$ when $\theta_0=0$ and $\boldsymbol{y}=\mathbf{0}$. When P is rotated and translated, F is evaluated in global coordinates as $f(\mathbf{R}_{-\theta}(\boldsymbol{x}-\boldsymbol{y}))$. Our goal is to define a moving frame for P in global coordinates using basis vectors $\{\mathbf{e}_{1,\boldsymbol{x}},\ \mathbf{e}_{2,\boldsymbol{x}}\}$, where

$$\mathbf{e}_{1,\boldsymbol{x}} = \cos(\theta_{\boldsymbol{x}})\mathbf{e}_1 + \sin(\theta_{\boldsymbol{x}})\mathbf{e}_2, \quad (16)$$

$$\mathbf{e}_{2,\boldsymbol{x}} = \cos(\theta_{\boldsymbol{x}}+\pi/2)\mathbf{e}_1 + \sin(\theta_{\boldsymbol{x}}+\pi/2)\mathbf{e}_2. \quad (17)$$

This frame will be defined by the local geometry of F so that it will be invariant to translations and rotations of P. For now, we assume that the wavelet scale s and the harmonic index $n=1$ are fixed; however, the same computation will be valid for any value.

Definition 1 (Optimal angle $\theta_{\boldsymbol{x}}$ and moving frames)**.** *We consider a manifold P of the form described above. Any point on the manifold can be written in global coordinates as* $\boldsymbol{x} = (x_1, x_2)$*. For this point, we compute the optimal angle, with respect to F, as*

$$\begin{aligned}\theta_{\boldsymbol{x},F} :&= \underset{\theta\in[0,2\pi)}{\arg\max}\left(\mathrm{Re}\left(\left\langle F, \phi^{(1)}_{s,\mathbf{0},\theta}(\cdot-\boldsymbol{x})\right\rangle\right)\right)\\ &= \underset{\theta\in[0,2\pi)}{\arg\max}\left(\mathrm{Re}\left(\left\langle F, \phi^{(1)}_{s,\mathbf{0}}(\mathbf{R}_{-\theta}(\cdot-\boldsymbol{x}))\right\rangle\right)\right).\end{aligned} \tag{18}$$

We also define the moving frame representation with respect to F to be the decomposition of an image using the locally steered multi-order CHWs

$$\phi^{(n)}_{s,\boldsymbol{x},\theta_{\boldsymbol{x},F}} = \phi^{(n)}_{s,\mathbf{0}}(\mathbf{R}_{-\theta_{\boldsymbol{x},F}}(\cdot-\boldsymbol{x})). \tag{19}$$

Note that inner products are taken with respect to the global coordinates.

Proposition 2. *The moving frame is invariant to rotation and translation. We have*

$$\theta_{\boldsymbol{x},f(\mathbf{R}_{-\theta_0}\cdot-\boldsymbol{y})} - \theta_0 = \theta_{\mathbf{R}_{-\theta_0}\boldsymbol{x}-\boldsymbol{y},f}. \tag{20}$$

The proof of Proposition 2 is detailed in Appendix A.

A discrete moving frame representation $\{\mathbf{e}_{1,\boldsymbol{k}},\ \mathbf{e}_{2,\boldsymbol{k}}\}$ is obtained from the discretization of $\{\mathbf{e}_{1,\boldsymbol{x}},\ \mathbf{e}_{2,\boldsymbol{x}}\}$ with $k_1 = x_1/\Delta x_1$, $k_2 = x_2/\Delta x_2$. A remarkable property following Definition 1 is that the effect of integration on the MF representation over an image domain $\boldsymbol{M}$ does not dissociate the joint responses of directional operators because the orientation $\theta_{\boldsymbol{x},F}$ of all wavelets $\phi^{(n=1,\ldots,|S|)}_{s,\mathbf{0},\theta_{\boldsymbol{x},F}}$ varies for each global coordinate $\boldsymbol{x}$. Therefore, the MF representation $\left|\sum_{n=0}^{|S|}\mathrm{Re}\left(\left\langle f, \phi^{(n)}_{s,\mathbf{0},\theta_{\boldsymbol{x}}}(\boldsymbol{x})\right\rangle\right)\right|$ tends to preserve the joint location and orientation structure of texture, yielding a precise characterization of the LOIDs.

H. Learning Moving Frames from Multi-Order CHWs

Equation (18) defines MFs optimal angles $\theta_{\boldsymbol{x},F}$ based on the gradient. However, the latter is handcrafted and does not allow finding local orientations that are useful to discriminate the texture classes of a considered set C. Following our previous work [7], we use linear SVMs in a feature space spanned by the absolute values of the multi-order subbands $\left|\mathrm{Re}\left(\left\langle f, \phi^{(n)}_{s,\boldsymbol{k}}\right\rangle\right)\right|$ to learn optimal linear combinations (*i.e.*, in the sense of structural risk minimization [38]) of consecutive harmonics for a class c in a one-versus-all (OVA) classification configuration. For a set of classes $c = 1,\ldots,C$, the latter will generate a shaping matrix $\mathbf{U}$ of the canonical CHW representation of steerability. This will add directionality to resulting wavelet profiles, and yield class-specific local orientations $\theta^{(c)}_{\boldsymbol{x},F}$ to construct MFs.

We formulate the transform similarly to Unser et al. [34], with the difference that $\mathbf{U}$ is not necessarily orthogonal. The transformation is described as

$$\begin{bmatrix}\psi^{(1)}_{s,\boldsymbol{k}}\\ \vdots\\ \psi^{(C)}_{s,\boldsymbol{k}}\end{bmatrix} = \mathbf{U}\begin{bmatrix}\phi^{(0)}_{s,\boldsymbol{k}}\\ \vdots\\ \phi^{(|S|)}_{s,\boldsymbol{k}}\end{bmatrix}. \tag{21}$$

$\{\psi^{(c)}_{s,\boldsymbol{k}}\}$ are the new wavelet channels at scale s and location $\boldsymbol{k}$. The new wavelets are also steerable and span the same space as the wavelet frame $\phi^{(n)}_{s,\boldsymbol{k}}$. L_2-SVMs are used to find the optimal linear combination of harmonic channels $\boldsymbol{u}^{(c)}$ (the lines of $\mathbf{U}$) for the texture class c. Considering a training set of I texture instances $\boldsymbol{v}_{i=1,\ldots,I}$, the SVMs find the separating hyperplane $\boldsymbol{u}^{(c)}$ with the maximum margin $\frac{1}{\|\boldsymbol{u}^{(c)}\|}$ between the instances with positive versus negative labels y_i^+ and y_j^-, respectively [38]. More precisely, $\boldsymbol{u}^{(c)}$ is a solution of the primal formulation

$$\begin{aligned}&\min_{\boldsymbol{u}_c,\xi,b}\left\{\frac{\|\boldsymbol{u}^{(c)}\|^2}{2} + Q\sum_{i=1}^{I}\xi_i^2\right\} \quad \text{subject to}\\ &\qquad y_i\left(\langle\boldsymbol{u}^{(c)}, \boldsymbol{v}_i\rangle - b^{(c)}\right) \geq 1-\xi_i, \quad \forall i.\end{aligned} \tag{22}$$

ξ_i is called a slack variable and loosens the margin constraints when the classification configuration is not linearly separable ($\xi_i > 1$). $b^{(c)}$ is the offset of $\boldsymbol{u}^{(c)}$. The regularization variable Q is used to control the cost of errors. The instances $\boldsymbol{v}_i$ that are located within the margin ($0 \leq \xi_i \leq 1$) are called the support vectors. The primal formulation in (22) can be solved with a dual formulation where a Lagrangian based on the primal variables is minimized [38].

By creating different training sets for each class where the labels y_i^+ are set for all instances $\boldsymbol{v}_i$ of the class c and y_j^- are set for the instances $\boldsymbol{v}_j$ of all other classes, the shaping matrix $\mathbf{U}$ can be built and a collection of class-specific texture signatures $\psi^{(c=1,\ldots,C)}_{s,\boldsymbol{k}}$ are obtained. This allows creating a new collection of class-specific MFs from the optimal angles $\theta^{(c=1,\ldots,C)}_{\boldsymbol{x},F}$, defined with respect to F, as

$$\begin{aligned}\theta^{(c)}_{\boldsymbol{x},F} :&= \underset{\theta\in[0,2\pi)}{\arg\max}\left(\left\langle F, \psi^{(c)}_{s,\mathbf{0},\theta}(\cdot-\boldsymbol{x})\right\rangle\right)\\ &= \underset{\theta\in[0,2\pi)}{\arg\max}\left(\left\langle F, \psi^{(c)}_{s,\mathbf{0}}(\mathbf{R}_{-\theta}(\cdot-\boldsymbol{x}))\right\rangle\right).\end{aligned} \tag{23}$$

Since $\psi^{(c)}_{s,\boldsymbol{k}}$ inherits the rotation- and translation-invariance properties of $\phi^{(n)}_{s,\boldsymbol{k}}$ (see [34]), the learned MF representation is also invariant to rotation and translation, which can be demonstrated following the proof of Propositon 2. Equation (23) allows defining the basis vectors $\left\{\mathbf{e}^{(c)}_{1,\boldsymbol{x}},\ \mathbf{e}^{(c)}_{2,\boldsymbol{x}}\right\}$ of the class-specific MF representation, where

$$\mathbf{e}^{(c)}_{1,\boldsymbol{x}} = \cos(\theta^{(c)}_{\boldsymbol{x}})\mathbf{e}_1 + \sin(\theta^{(c)}_{\boldsymbol{x}})\mathbf{e}_2, \tag{24}$$

$$\mathbf{e}^{(c)}_{2,\boldsymbol{x}} = \cos(\theta^{(c)}_{\boldsymbol{x}} + \pi/2)\mathbf{e}_1 + \sin(\theta^{(c)}_{\boldsymbol{x}} + \pi/2)\mathbf{e}_2. \tag{25}$$

The learned MF representation $\left|\left\langle f, \psi^{(c)}_{s,\mathbf{0},\theta_{\boldsymbol{x}}}(\boldsymbol{x})\right\rangle\right|$ encodes the LOIDs that are now specific to the class c. Intuitively, the learned MFs can be seen as class-specific detectors that are applied and rotated at each point of the image to evaluate the magnitude of their responses, *i.e.*, probing the presence of the texture class c in a rotation-invariant fashion.

In summary, the SWM forward function maps an input image f_i to feature maps $\boldsymbol{t}_{i,\boldsymbol{x}}$ trough linear operations (*i.e.*, convolution and linear combinations) interleaved by non-linear *steermax* operations (see Figure 2). The final feature representation $\boldsymbol{t}_{i,\boldsymbol{x}}$ can be further used by either a segmentation model, or aggregated over a region $\boldsymbol{M}$ and used by a classifier (*e.g.*, SVMs, k-nearest neighbors).

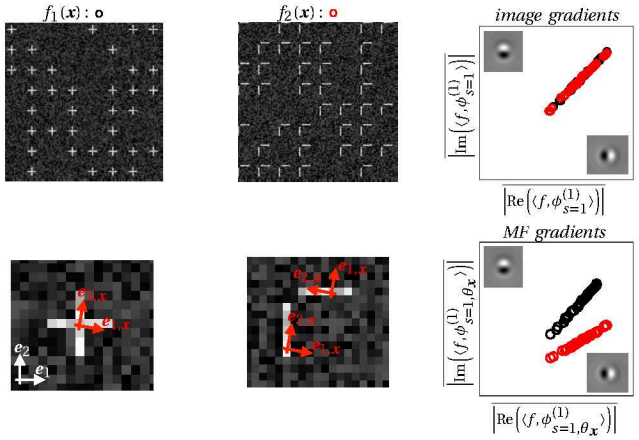

Fig. 4. $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ only differs in terms of the LOIDs. The joint responses of image gradients expressed in terms of global coordinates $\{\mathbf{e}_1, \mathbf{e}_2\}$ can hardly discriminate between f_1 and f_2 when averaged over the full image (top right). However, the gradient vector expressed in terms of the MFs $\{\mathbf{e}_{1,\boldsymbol{x}}, \mathbf{e}_{2,\boldsymbol{x}}\}$ perfectly separates between the two textures (bottom right). The imaginary part of the MF gradient gets higher responses on crosses in f_1 than on bars in f_2. One circle in the gradient representation corresponds to one realization (*i.e.*, full image) of $f_{1,2}$.

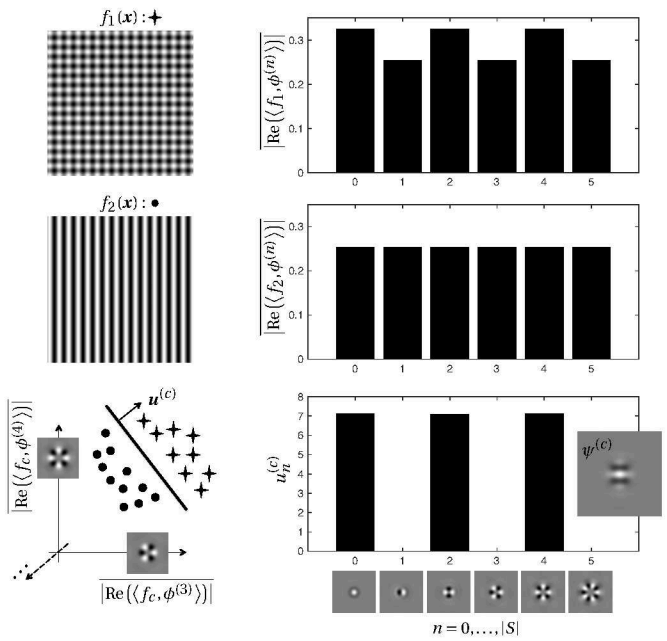

Fig. 5. Illustration of the template learning process for $|S| = 5$. An optimally discriminant template $\psi^{(c)}$ for textures f_1 (top left) versus f_2 (middle left) was learned with linear SVMs (bottom left). f_1 and f_2 have identical average responses $\left|\mathrm{Re}\left(\langle f_c, \phi^{(n)}\rangle\right)\right|$ for the odd harmonics $n = 1, 3, 5$ (see top and middle right). Therefore, the SVMs assigned non-null weights $u_n^{(c)}$ to even harmonics $n = 0, 2, 4$ only (see bottom right). The new representation $\left\langle f_i, \psi^{(c)}_{s,\mathbf{0},\theta_0} \right\rangle$ can be further used to derive learned MFs representations based on local optimal angles $\theta^{(c)}_{\boldsymbol{x},f_i}$ with Eq. (23).

III. Experimental Results

The behavior and performance of the proposed texture operators are evaluated in this section. The ability of MFs to characterize the LOIDs is first demonstrated in Section III-A. A toy problem is presented in Section III-B to illustrate the moving frame learning process. A full evaluation of the classification performance of MFs with three test suites of the Outex database and the CUReT database is described in Section III-C.

A. Gradient-Based MF Representations of the LOIDs

The ability of gradient-based MFs (see Section II-G) to discriminate textures that differ in terms of the LOIDs only is illustrated in Figure 4. The gradient vector $\left(\left|\mathrm{Re}\left(\langle f, \phi^{(1)}_{s=1}\rangle\right)\right|, \left|\mathrm{Im}\left(\langle f, \phi^{(1)}_{s=1}\rangle\right)\right|\right)$ expressed in terms of global coordinates $\{\mathbf{e}_1, \mathbf{e}_2\}$ cannot accurately discriminate between the textures f_1 and f_2 when averaged over the image domain $\boldsymbol{M}$ (see Figure 4 top right). However, the gradient vector $\left(\left|\mathrm{Re}\left(\langle f, \phi^{(1)}_{s=1,\theta_{\boldsymbol{x}}}\rangle\right)\right|, \left|\mathrm{Im}\left(\langle f, \phi^{(1)}_{s=1,\theta_{\boldsymbol{x}}}\rangle\right)\right|\right)$ expressed in the MFs $\{\mathbf{e}_{1,\boldsymbol{x}}, \mathbf{e}_{2,\boldsymbol{x}}\}$ perfectly separates between f_1 and f_2 (see Figure 4 bottom right). The optimal angle MF angle $\theta_{\boldsymbol{x},F}$ was defined based on $s = 2$ (*i.e.*, the second wavelet scale), which is why the imaginary part of the gradient of scale 1 is not null.

B. Moving Frame Learning with Synthetic Textures

The moving frame learning process is illustrated in Figure 5 for two synthetic textures f_1 (sum of vertical and horizontal sines) versus f_2 (vertical sine only), and $|S| = 5$. The SVMs assigned non-null weights $u_{n,c}$ to even harmonics only (*i.e.*, $n = 0, 2, 4$) since $\mathrm{Re}\left(\phi^{(n=1,3,5)}\right)$ are not sensitive to horizontal directions. The corresponding profile $\psi^{(c)}$ corresponds qualitatively to a detector of horizontal sines, the latter being required to discriminate f_1 and f_2.

C. Texture Classification with SWMs

We evaluated the performance of SWMs for texture classification using the Outex [39], CUReT [40] and UIUC [17] databases. Both require using texture operators that are invariant to Euclidean transforms and illumination changes. Test suites designed for extensively testing the rotation-invariant properties of the algorithms exist and come with pre-defined training and testing sets, which allows for direct performance comparisons between approaches (*i.e.*, identical validation methods). The cardinalities of the classes are balanced both in the training and test sets for all problems. The test suites are Outex_TC_10, Outex_TC_12, CUReT and UIUC, which were widely used to compare texture classification approaches [5, 7, 15, 16, 18, 22–29, 41–45].

Outex is a set of real textures photographed with controlled illumination conditions and consists of 24 texture classes with pronounced directional structures. Three different color spectra were used for image capture to evaluate illumination invariance of approaches: 2300 Kelvin (K) horizon sunlight denoted as "horizon", 2856 K incandescent denoted as "inca", and 4000 K fluorescent tl84 denoted as "tl84". Each texture sample was captured using nine rotation angles (0°, 5°, 10°, 15°, 30°, 45°, 60°, 75°, and 90°) to focus on the rotation-invariant properties of the approaches. There are 20 128×128 texture instances per class (see Figure 6). The Outex_TC_10 test suite has a total

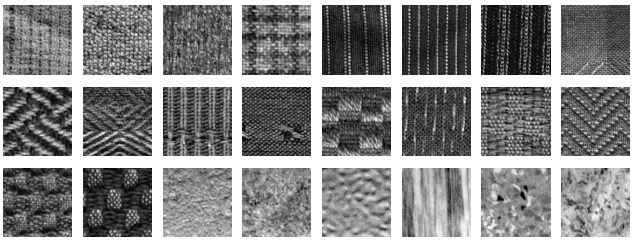

Fig. 6. 128×128 unrotated blocks from the 24 texture classes of the Outex database.

of 4320 ($24 \times 20 \times 9$) image instances of illuminant "inca". The training set consists of the 480 (24×20) non-rotated images and the remaining 3840 ($24 \times 20 \times 8$) images from 8 orientations are constituting the test set. Outex_TC_12 includes two subproblems: P0 and P1. Both problems use the same training set as in Outex_TC_10 (*i.e.*, 24×20 non-rotated images of illuminant "inca"). The test sets consist of all samples captured using illuminant "tl84" for P0 and "horizon" for P1 and contain 4320 images each.

The CUReT [40] database contains 61 texture classes with 92 200×200 images each under varying illumination direction but at a constant scale. For each class, training and test sets are obtained from even random splits of the 92 images. The reported accuracies were obtained after averaging over 10 Monte-Carlo (MC) repetitions.

The UIUC [17] dataset contains 25 classes with 40 640×480 images each, captured under varying viewpoints. It therefore includes strong intra-class variations in texture scale in addition to image orientation. For each class, training and test sets are obtained from even random splits of the 40 images. The reported accuracies were obtained after averaging over 10 MC repetitions.

6 dyadic CHW scales were used to cover the spatial spectrum of the images with an undecimated wavelet transform. The templates $\psi_s^{(c)}$ were learned using images from the training set. Each of them was learned and steered for each scale separately. The cost of errors Q of the internal SVM in Eq. (22) was set to 10^2 for all experiments. The absolute values of the feature maps $\boldsymbol{t}_{i,\boldsymbol{x}}$ in Figure 2 were averaged over the 128×128 images and used for classification. The latter were concatenated from each scale. From this final feature space, L_2-SVMs with Gaussian kernels (hereinafter referred to as K-SVMs) were constructed using the training set. The cost of errors Q in (22) and σ_K of the Gaussian kernel were optimized in the intervals $[10^0, 10^8]$ and $[10^{-9}, 10^2]$, respectively.

The classification performance is shown in Figure 7 for the three classification subproblems of Outex and for different numbers of combined harmonics $|S|$. The performance for the CUReT database is shown in Figure 8. Two representations are compared:

- *CHW*, *i.e.*, the complex modulus of the collections of CHW subbands provided by (13): $\left|\left\langle f_i, \phi_{s,\boldsymbol{x}}^{(n)} \right\rangle\right|$. The feature dimensionality is $6 \cdot (n+1)$, *i.e.*, from 6 to 66.
- *SWMs*, *i.e.*, the final feature representation $\boldsymbol{t}_{i,\boldsymbol{x}}$ based on moving frames provided by (23) with learned $\mathbf{U}$: $\left|\left\langle f_i, \psi_{s,\boldsymbol{x},\theta_{\boldsymbol{x},f_i}^{(c)}}^{(c)} \right\rangle\right|$ (see Figure 2). The feature dimensionality is $6 \cdot C$, *i.e.*, 144 for Outex, 366 for CUReT and 150 for UIUC.

The influence of the final classifier is studied for Outex_TC_10, where linear SVMs (L-SVMs) and k-nearest neighbors (kNN) are compared to K-SVMs (see Figure 7). The cost of errors Q was optimized in $[10^0, 10^8]$ for L-SVMs. The number of neighbors k was optimized in $[0, 10]$ for kNNs.

The performance of nineteen other approaches for rotation-invariant texture classification based on Outex TC_10, TC_12 P0, TC_12 P1, CUReT and UIUC are reported in Table I and compared to the proposed approach.

IV. Discussions and Conclusions

We developed novel texture operators that can encode the multi-scale class-specific LOIDs in a translation- and rotation-invariant fashion. Whereas current approaches encoding the LOIDs (*e.g.*, LBPs, MR8, SIFT, ScatNet) yield handcrafted image features, the proposed approach learns class-specific encoding of the LOIDs that is relevant to the specific image recognition task in hand. The cornerstone of the proposed method is to generate MFs from locally steered linear combinations of CHWs. Class-specific MFs were obtained by using SVMs to learn optimal transformations (*i.e.*, in the sense of structural risk minimization [38]) of the initial CHW representation, the latter corresponding to the canonical representation of wavelet steerability [34]. The full SWM forward function is composed of linear operations (*i.e.*, convolution and weighted combinations) interleaved by non-linear *steermax* operations (see Figure 2). The application scope of SWMs is restricted to image modalities with pixel sizes defined in physical units (*e.g.*, medical and satellite imaging, material analysis), where the image scale is an important discriminative property.

The discriminatory power of gradient-based MFs was first qualitatively demonstrated in Figure 4, which yielded feature representations that were linearly separable between texture classes that only differed in terms of their LOIDs. This verified that the joint location and orientation structure of textures are preserved when the proposed texture operators are integrated (*i.e.*, averaged) over an image domain $\boldsymbol{M}$. The invariance of MFs to Euclidean transformations was demonstrated in Appendix A.

A proof of concept of the MF learning process was illustrated in Figure 5 with the construction of a discriminant template $\psi^{(c)}$ between a texture f_1 (sum of vertical and horizontal sines) versus f_2 (vertical sine only). Discriminating between f_1 and f_2 only requires detecting the presence of horizontal image directions: the SVMs transformed the initial CHW representation of texture $\langle f, \phi^{(n)} \rangle$ into the angular-selective representation $\langle f, \psi^{(c)} \rangle$, where only channels that are sensitive to horizontal directions received non-null weights. The local orientation maximization of this particular angular-selective representation yielded MFs that are optimally discriminant between f_1 and f_2.

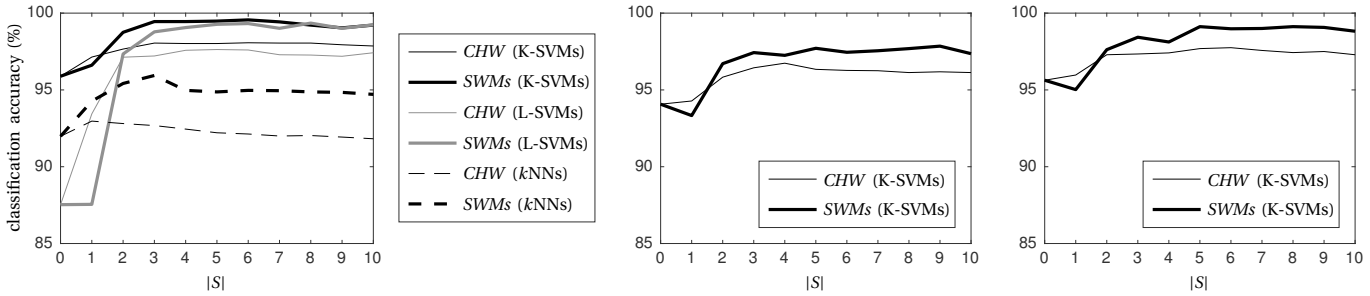

Fig. 7. Texture classification accuracies for Outex_TC_10 (left), Outex_TC_12 P0 (middle) and Outex_TC_12 P1 (right) and for different number of combined harmonics $|S| = 0, \ldots, 10$. Various classifiers are compared for Outex_TC_10 (*i.e.*, K-SVMs, L-SVMs, *k*NNs).

TABLE I
PERFORMANCE COMPARISON WITH OTHER APPROACHES FOR ROTATION-INVARIANT TEXTURE CLASSIFICATION BASED FOR OUTEX, CURET AND UIUC. THE STUDIES ARE ORDERED BY DECREASING CLASSIFICATION ACCURACY FOR OUTEX TC_10.

Study	Outex_TC_10	Outex_TC_12 P0	Outex_TC_12 P1	CUReT	UIUC	description
Liu *et al.* 2016 [29]	99.87	99.49	99.7	99.02	–	Median robust extended LBP
Liu *et al.* 2012 [22]	99.7	98.7	98.1	97.29	–	Extended LBPs
Proposed (*SWMs*)	99.56	97.85	99.12	96.86	89.12	Steerable wavelet machines
Guo *et al.* 2010 [24]	99.32	95.32	94.53	95.86	–	Completed LBPs
Khellah 2011 [41]	99.27	94.4	92.85	95	–	Dominant neighborhood structure combined with LBPs
Shrivastava *et al.* 2015 [23]	99.19	96.97	96.93	95.81	92.84	Noise invariant structure patterns (based on LBPs)
He *et al.* 2011 [25]	99.18	96.2	96.2	93.04	–	LBP textons
Sifre *et al.* 2012 [31]	98.75	–	–	–	–	ScatNet: Scattering transform (based on wavelets and deep convolutional networks)
Guo *et al.* 2012 [26]	98.64	95.99	94.16	94.49	–	LBPs based on high-order directional derivatives
Depeursinge *et al.* 2014 [7]	98.4	97.8	98.4	–	–	Steerable Riesz wavelets
Zand *et al.* 2015 [42]	98.38	–	–	–	–	Combined Gabor wavelets and curvelets
Guo *et al.* 2010 [27]	98.15	95.39	95.57	94.15	–	LBP variance
Ojala *et al.* 2002 [5]	97.9	90.2	87.2	–	–	Original LBP implementation
Hadizadeh 2015 [28]	97.3	–	–	94.51	–	LBPs on top of Gabor wavelet coefficients
Varma *et al.* 2009 [18] (perf. reported in [25])	94.11	92.64	92.64	97.47	97.83	Patch statistics (intensity-based)
Varma *et al.* 2005 [16] (perf. reported in [18, 25, 27, 44])	92.5 (best) 72.57 (worst)	90.9 (best) 87.49 (worst)	91.1 (best) 87.49 (worst)	98.4	92.94	Maximum response filterbank (MR8)
Zhang *et al.* 2012 [45] (perf. reported in [42])	79.22	–	–	–	–	Rotation-invariant curvelets
Lazebnik *et al.* 2005 [17] (CUReT perf. reported in [46])	75.26	60.44	57.43	72	92.61	Rotation-invariant feature transform (RIFT) with dense sampling
Leung *et al.* 2001 [15] (perf. reported in [44])	51.87	–	–	–	–	Leung-Malik filterbank
Xu *et al.* 2010 [19]	–	–	–	–	98.6	Multi-orientation wavelet leaders

The classification performance of the proposed operators was evaluated in Section III-C (see Figures 7, 8 and Table I). It can be observed that $|S| = 1$ provided poor accuracies, which can be explained by the fact the templates were learned on top of the gradient-based MF representations. Starting from orders as low as $|S| = 2$, *SWMs* provided equal or superior performance when compared to *CHW*, highlighting the superiority of learned representations when compared to handcrafted ones. It also underlines the importance of the inter-harmonic phase information, which is discarded by *CHW*. The performance gain observed between $|S| = 2$ and $|S| = 4$ suggests that the number of harmonics of the initial CHW representation needs to be rich enough to learn relevant operators and shape significant directional wavelet profiles. Tuning the number of harmonics $|S|$ acted as regularization optimization of the wealth of the operators. This is particularly symptomatic when analyzing the performance drop in Figure 8 for $|S| > 6$, where high-order SWMs are not generalizing well. The influence of the final classifier was studied for Outex_TC_10 in Figure 7 (left). L-SVMs, K-SVMs and *k*NNs showed all a large classification improvement when using *SWMs*. The

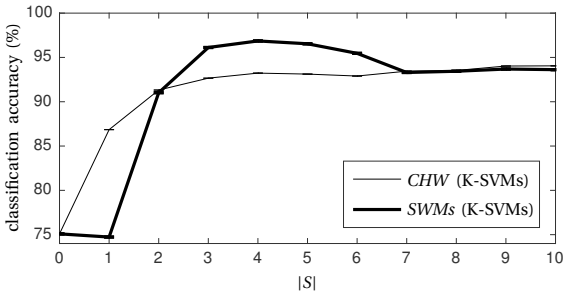

Fig. 8. Texture classification accuracies for CUReT for a varying number of combined harmonics $|S| = 0, \ldots, 10$.

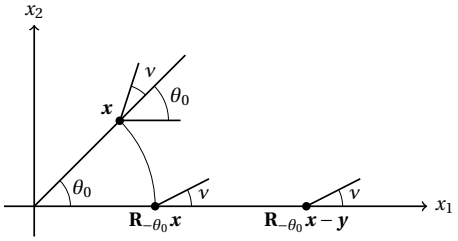

Fig. 9. Illustrating the proof of Proposition 2. The moving frame is invariant to any rotation parameterized by θ_0 and to any translation parameterized by $\boldsymbol{y}$. Note that neither $\mathbf{R}_{-\theta_0}\boldsymbol{x}$ nor $\mathbf{R}_{-\theta_0}\boldsymbol{x}-\boldsymbol{y}$ need lie on the x_1-axis.

top accuracies were obtained by SVMs, where L-SVMs and K-SVMs yielded very close performance for $|S| > 3$. The computing time for the SWM forward function of a 128×128 image of the Outex dataset was of 0.83 second for $|S| = 5$ with MATLAB R2015b, The MathWorks Inc., Natick, Massachusetts, USA on a 2.5 GHz Intel Core i7 CPU.

Overall, the performance obtained with SWMs were very competitive when compared to the state-of-the-art (see Table I) on Outex and CUReT. The top performances on the Outex test suites were very close to the LBP-based methods of Liu *et al.* [22, 29]. When compared to the latter, SWMs have the advantage of a small number of free-parameters (essentially $|S|$), as well as compact feature dimensions. Feature dimensionality as large as 800 are reported in [29]. Such a large number of dimensions should be avoided to limit the risk of overfitting when the number of training instances are as low as 480 in the Outex database. For all subsets, a number of harmonics $|S| \in [2,8]$ was found to provide stable performances, which suggests that this free-parameter is not difficult to optimize for a new application. The multi-order CHWs yielded an excellent initial representation for building and learning MFs. The combinations of harmonics allowed encoding both symmetric and anti-symmetric profiles, providing an excellent characterization of the local circular phase and frequencies. CHW relate to rotation-invariant LBP [33] by modeling local circular harmonics and come with a more complete theoretical framework for encoding the LOIDs at multiple scales. Moreover, CHWs are linear operators and do not require the binarization step carried out with LBPs, the latter entailing the risk of discarding important information concerning the dynamic and differential range of local pixel values. The top performance was already obtained with a relatively small number of harmonics $|S|$ of two to six. When circular harmonics are coupled with isotropic wavelet frames, the coverage of the spatial spectrum can be fully controlled, which is not the case for the family of classical LBP operators. The proposed approach also achieved top performance with the two Outex_TC_12 subproblems and the CUReT. This highlighted the robustness of the operators to changes in illumination. The latter is naturally achieved by using zero-mean (*i.e.*, bandpass) operators. The performance obtained on the UIUC dataset is relatively low because our method is not regrouping patterns that are similar at different scales (the steerable wavelets are learned for each scale independently). As expected, the methods achieving high performance on UIUC are invariant to image scale (e.g., [17–19]). However, the latter (e.g., [17, 18]) are providing lower performance on the Outex and CUReT because they discard scale as a discriminative property (see Table I).

We are currently extending the framework to 3-D based on [47, 48]. Future work will also include revealing and exploiting the visual diversity of texture patterns in order to account for texture classes composed of multiple distinct visual events (*e.g.*, see Figure 6) [49]. We are also working on the learning of the radial profile. The authors will make the implementation available to the community.

APPENDIX A
PROOF OF PROPOSITION 2

Proof. Suppose there is an image F on the manifold P, and F is given by $g = f(\cdot - \boldsymbol{y})$. We then have $\theta_{\boldsymbol{x},f} = \theta_{\boldsymbol{x}+\boldsymbol{y},f(\cdot-\boldsymbol{y})} = \theta_{\boldsymbol{x}+\boldsymbol{y},g}$, *i.e.*, shifting the manifold does not change the computed angle. Hence the frame remains the same at each point of P.

Now suppose the manifold P is oriented so that F is computed as $g = f(\mathbf{R}_{-\theta_0}\cdot)$. We then have $\theta_{\mathbf{0},g} - \theta_0 = \theta_{\mathbf{0},f(\mathbf{R}_{-\theta_0}\cdot)} - \theta_0 = \theta_{\mathbf{0},f}$. We interpret this to mean that the frame of the rotated manifold is equivalent to the rotation of the frame of the original manifold, which is the invariance that we sought to show. In general, we will have

$$\theta_{\boldsymbol{x},f(\mathbf{R}_{-\theta_0}\cdot)} - \theta_0 = \theta_{\mathbf{R}_{-\theta_0}\boldsymbol{x},f}. \tag{26}$$

We can combine these two invariance properties to see that a similar result holds when the manifold is both translated and rotated. □

The proof of Proposition 2 is illustrated in Figure 9. Consider a point $\boldsymbol{x}$ in the plane, an angle θ_0 and a shift $\boldsymbol{y}$. Suppose that the optimal angle at the point $\mathbf{R}_{-\theta_0}\boldsymbol{x}-\boldsymbol{y}$ for the unrotated and unshifted function f is ν, *i.e.*, $\nu = \theta_{\mathbf{R}_{-\theta_0}\boldsymbol{x}-\boldsymbol{y},f}$. Then the optimal angle for the rotated and shifted function $f(\mathbf{R}_{-\theta_0}\cdot - \boldsymbol{y})$ at the point $\boldsymbol{x}$ is $\theta_0 + \nu$, *i.e.*, $\theta_0 + \nu = \theta_{\boldsymbol{x},f(\mathbf{R}_{-\theta_0}\cdot-\boldsymbol{y})}$. Combining these equations, we have

$$\theta_{\boldsymbol{x},f(\mathbf{R}_{-\theta_0}\cdot-\boldsymbol{y})} - \theta_0 = \theta_{\mathbf{R}_{-\theta_0}\boldsymbol{x}-\boldsymbol{y},f}. \tag{27}$$

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation (under grant PZ00P2_154891), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement n° 267439, and the Hasler Foundation.

REFERENCES

[1] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep convolutional filter banks for texture recognition and segmentation," *CoRR*, vol. abs/1411.6836, 2015.

[2] C. Blakemore and F. W. Campbell, "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images," *The Journal of Physiology*, vol. 203, no. 1, pp. 237–260, 1969.

[3] J. R. Bergen and M. S. Landy, "Computational modeling of visual texture segregation," in *in Computational Models of Visual Processing*, pp. 253–271, MIT Press, 1991.

[4] T. Watanabe and P. Cavanagh, "Texture laciness: the texture equivalent of transparency?," *Perception*, vol. 25, no. 3, pp. 293–303, 1996.

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray–scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[6] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[7] A. Depeursinge, A. Foncubierta, D. Van De Ville, and H. Müller, "Rotation–covariant texture learning using steerable Riesz wavelets," *IEEE Transactions on Image Processing*, vol. 23, pp. 898–908, 2014.

[8] O. Ben-Shahar and S. Zucker, "The perceptual organization of texture flow: a contextual inference approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 4, pp. 401–417, 2003.

[9] D. G. Lowe, "Object recognition from local scale–invariant features," in *Proceedings of the International Conference of Computer Vision*, ICCV 1999, (Corfu, Greece), 1999.

[10] K. Liu, H. Skibbe, T. Schmidt, T. Blein, K. Palme, T. Brox, and O. Ronneberger, "Rotation–invariant HOG descriptors using Fourier analysis in polar and spherical coordinates," *International Journal of Computer Vision*, vol. 106, no. 3, pp. 342–364, 2014.

[11] M. Papadakis, G. Gogoshin, I. A. Kakadiaris, D. J. Kouri, and D. K. Hoffman, "Nonseparable radial frame multiresolution analysis in multidimensions and isotropic fast wavelet algorithms," in *Proc. SPIE Wavelets: Applications in Signal and Image Processing X*, vol. 5207, pp. 631–642, 2003.

[12] E. Oyallon and S. Mallat, "Deep roto–translation scattering for object classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2865–2873, 2015.

[13] E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surface Fitting*, pp. 105–120, Vanderbilt University Press, 2000.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[15] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three–dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.

[16] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1265–1278, 2005.

[18] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2032–2047, 2009.

[19] Y. Xu, X. Yang, H. Ling, and H. Ji, "A new texture descriptor using multifractal analysis in multi–orientation wavelet pyramid," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 161–168, 2010.

[20] O. D. Faugeras, "Cartan's moving frame method and its application to the geometry and evolution of curves in the Euclidean, affine and projective planes," tech. rep., Institut National de Recherche en Informatique et en Automatique (INRIA), 1993.

[21] L. Florack, B. Ter Haar Romeny, J. Koenderink, and M. Viergever, "Cartesian differential invariants in scale–space," *Journal of Mathematical Imaging and Vision*, vol. 3, no. 4, pp. 327–348, 1993.

[22] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol. 30, no. 2, pp. 86–99, 2012.

[23] N. Shrivastava and V. Tyagi, "Noise–invariant structure pattern for image texture classification and retrieval," *Multimedia Tools and Applications*, pp. 1–20, 2015.

[24] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[25] Y. He, N. Sang, and R. Huang, "Local binary pattern histogram based texton learning for texture classification," in *IEEE International Conference on Image Processing*, ICIP, pp. 841–844, 2011.

[26] Z. Guo, Q. Li, J. You, D. Zhang, and W. Liu, "Local directional derivative pattern for rotation invariant texture classification," *Neural Computing and Applications*, vol. 21, no. 8, pp. 1893–1904, 2012.

[27] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.

[28] H. Hadizadeh, "Noise-resistant and rotation-invariant texture description and representation using local Gabor wavelets binary patterns," in *International Symposium on Artificial Intelligence and Signal Processing (AISP)*, AISP, pp. 30–34, 2015.

[29] L. Liu, S. Lao, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1368–1381, 2016.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[31] L. Sifre and S. Mallat, "Combined scattering for rotation invariant texture analysis," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 24–27, 2012.

[32] G. Jacovitti and A. Neri, "Multiresolution circular harmonic decomposition," *IEEE Trans. on Signal Proc.*, vol. 48, pp. 3242–3247, 2000.

[33] J. Fehr, "Rotational invariant uniform local binary patterns for full 3D volume texture analysis," in *Finnish Signal Processing Symposium (FINSIG), 2007*, (Oulu, Finland), 2007.

[34] M. Unser and N. Chenouard, "A unifying parametric framework for 2D steerable wavelet transforms," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 102–135, 2013.

[35] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, 2000.

[36] K. G. Larkin, D. J. Bone, and M. A. Oldfield, "Natural demodulation of two-dimensional fringe patterns. I. General background of the spiral phase quadrature transform," *Journal of the Optical Society of America A*, vol. 18, pp. 1862–1870, 2001.

[37] K. G. Larkin, "Natural demodulation of two-dimensional fringe patterns. II. Stationary phase analysis of the spiral phase quadrature transform," *Journal of the Optical Society of America A*, vol. 18, pp. 1871–1881, 2001.

[38] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.

[39] T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen, "Outex – new framework for empirical evaluation of texture analysis algorithms," in *16th International Conference on Pattern Recognition*, pp. 701–706, IEEE Computer Society, 2002.

[40] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real–world surfaces," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 1–34, 1999.

[41] F. M. Khellah, "Texture classification using dominant neighborhood structure," *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3270–3279, 2011.

[42] M. Zand, S. Doraisamy, A. A. Halin, and M. R. Mustaffa, "Texture classification and discrimination for region–based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 305–316, 2015.

[43] N. Doshi and G. Schaefer, "A comparative analysis of local binary pattern texture classification," in *Visual Communications and Image Processing (VCIP)*, pp. 1–6, 2012.

[44] O. Ghita, D. Ilea, A. Fernandez, and P. Whelan, "Local binary patterns versus signal processing texture analysis: a study from a performance evaluation perspective," *Sensor Review*, vol. 32, pp. 149–162, 2012.

[45] D. Zhang, M. Islam, G. Lu, and I. Sumana, "Rotation invariant curvelet features for region based image retrieval," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 187–201, 2012.
[46] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.
[47] J. P. Ward and M. Unser, "Harmonic singular integrals and steerable wavelets in $l_2(\mathbb{R}^d)$," *Applied and Computational Harmonic Analysis*, vol. 36, no. 2, pp. 183–197, 2014.
[48] H. Skibbe, M. Reisert, T. Schmidt, T. Brox, O. Ronneberger, and H. Burkhardt, "Fast rotation invariant 3D feature computation utilizing efficient local neighborhood operators," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1563–1575, 2012.
[49] A. Depeursinge, A. Foncubierta, H. Müller, and D. Van De Ville, "Rotation–covariant visual concept detection using steerable Riesz wavelets and bags of visual words," in *SPIE Wavelets and Sparsity XV*, vol. 8858, pp. 885816–885816–11, SPIE, 2013.

Adrien Depeursinge received the B.Sc. and M.Sc. degrees in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2003 and 2005, respectively, with a specialization in signal and image processing. From 2006 to 2010, he performed his Ph.D. thesis on medical image analysis with a focus on texture analysis and content–based image retrieval at the University Hospitals of Geneva (HUG). He then spent two years as a Postdoctoral Fellow at the Department of Radiology of the School of Medicine at Stanford University. He has currently a joint position as a Professor of Computer Science at the Institute of Information Systems, University of Applied Sciences Western Switzerland (HES–SO), and as a Senior Research Scientist in the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne (EPFL).

Zsuzsanna Püspöki received her PhD Diploma in Electrical Engineering in 2016 from the École polytechnique fédérale de Lausanne (EPFL), Switzerland. There, she developed methodologies and frameworks for the efficient analysis of biomedical images with the focus on local transformable representations and their applications for feature extraction. Currently, she is a research assistant at the Lausanne University Hospital in the Laboratory for Research in Neuroimaging (LREN), directed by Prof. Bogdan Draganski. She is currently working on problems related to the understanding of neurodegenerative diseases and MRI imaging.

John Paul Ward received a B.S. degree in mathematics from the University of Georgia, Athens, and a Ph.D. in mathematics from Texas A&M University, College Station, in 2005 and 2010, respectively. He did postdoctoral work at Texas A&M University, College Station in the math department; the Swiss Federal Institute of Technology, Lausanne, Switzerland in the Biomedical Imaging Group; and the University of Central Florida, Orlando in the mathematics department. Since 2016, he is an assistant professor in the mathematics department at North Carolina Agricultural and Technical State University..

Michael Unser (M'89–SM'94–F'99) is professor and director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His primary area of investigation is biomedical image processing. 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 published over 250 journal papers on those topics. He is the author with P. Tafti of the book "An introduction to sparse stochastic processes", Cambridge University Press 2014.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda USA, conducting research on bioimaging.

Dr. Unser has held the position of associate Editor-in-Chief (2003-2005) for the IEEE Transactions on Medical Imaging. He is currently member of the editorial boards of SIAM J. Imaging Sciences, IEEE J. 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 (BISP) of the IEEE Signal Processing Society. Prof. Unser is a fellow of the IEEE (1999), an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is the 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).