Discriminative Features for Texture Retrieval Using Wavelet Packets

ANDREA VIDAL[1], **(Student Member, IEEE), JORGE F. SILVA**[2], **(Senior Member, IEEE), AND CARLOS BUSSO**[1], **(Senior Member, IEEE)**

[1]Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA
[2]Information and Decision System Group (IDS), Universidad de Chile, Santiago 8370451, Chile

Corresponding author: Andrea Vidal (axv170003@utdallas.edu)

ABSTRACT Wavelet Packets (WPs) bases are explored seeking new discriminative features for texture indexing. The task of WP feature design is formulated as a learning decision problem by selecting the filter-bank structure of a basis (within a WPs family) that offers an optimal balance between estimation and approximation errors. To address this problem, a computationally efficient algorithm is adopted that uses the tree-structure of the WPs collection and the Kullback-Leibler divergence as a discrimination criterion. The adaptive nature of the proposed solution is demonstrated in synthetic and real data scenarios. With synthetic data, we demonstrate that the proposed features can identify discriminative bands, which is not possible with standard wavelet decomposition. With data with real textures, we show performance improvements with respect to the conventional Wavelet-based decomposition used under the same conditions and model assumptions.

INDEX TERMS Texture indexing, wavelet packets, minimum probability of error, complexity regularization, minimum cost tree pruning.

I. INTRODUCTION

In the current information age, we have access to unprecedented sources of digital image content. Consequently, being able to index and organize these documents based solely on the content extracted from the signals without relying on metadata or expensive human annotations has become a central problem [1]–[21]. In this context, an important task in image processing is texture retrieval. This problem has been richly studied over the last two decades with different frameworks and approaches [3]–[21], including, more recently, deep learning approaches [22], [22]–[26], [26]–[29].

In a nutshell, the texture retrieval problem can be formulated in two stages. The first stage, feature extraction (FE), implies the creation of low-dimensional descriptions of the image (i.e., the dimensionality reduction phase) with the objective of capturing the semantic high-level information that discriminates relevant texture classes. The second stage proposes a similarity measure (SM) on the feature space to

compare and organize the images in terms of their signal content.

For the FE stage, the Wavelet transform (WT) has been widely adopted as a tool to decompose and organize the signal content in sub-spaces associated with different levels of resolution (or scale) information [30], [31]. Based on this sub-space decomposition, energy features have been used as a signature that represents the salient texture attributes for texture indexing [32]. For the SM stage, numerous approaches have been proposed to compare images based on their feature representations using perceptual studies, heuristics and statistical principles [3]–[6], [9], [11], [14], [19], [33]–[35]. For statistical principles, we highlight the seminal work of Do and Vetterli [4] and its numerous extensions [12], [16], [19], [33]. They proposed an hypothesis testing (HT) formulation, where the optimal (joint) process for FE and SM is derived in a closed-form.

Texture images are considered as realizations of a statistical model where the task of indexing can be formulated as an HT problem using the criterion of minimum probability of error (MPE). In this context, the FE and SM tasks are analytically obtained in a closed form by implementing the

MPE criterion. There have been numerous extensions of this framework that have improved the characterization of the texture models [12], [19], [33]. Examples, include the modeling of the inter-band dependencies [10], [11], [13], [16], [21], and the selection of more complex representations using Curvelet, frames, and other tools from signal processing. [3], [5]–[9], [11], [14], [15], [17], [18], [20], [21], [36].

Convolutional neural networks (CNNs) have also been adopted for the task of image retrieval because of its great success in computer vision, including object detection and classification [37], [38], visual object tracking [39], [40], image segmentation [41], automatic image cropping [42] and saliency detection [39], [40], [42]–[45]).Deep learning models are often trained with larger datasets. However, in the particular case of texture retrieval, the available databases are often limited in size. For that reason, Liu *et al.* [22] adopted a pre-trained CNN to obtain features for the texture retrieval task. Despite all the advances in deep learning, the lack of bigger databases for texture retrieval has allowed traditional techniques to be still important for texture retrieval tasks [46], [47].

In this paper we follow the HT formulation presented in the seminal work by Do and Vetterli [4]. Motivated by the recent adoption of convolutional based deep representations (obtained for different layers), we revisit the FE phase by studying the rich family of Wavelet Packets (WPs). WPs are bases that offer a rich collection of decompositions of the image space in terms of space-scale components [30], [31]. This collection is efficiently constructed by several layers of convolutions (linear) and down-sampling (non-linear) operations using a two-channel filter (TCF) as a basic building block. From this hierarchical construction, WPs can be organized and indexed by a collection of embedded trees of different depths and structures. A particular example of the WP family is the Wavelet basis, which is obtained by iterating the TCF in the low frequency, therefore, producing a multi-resolution partition of the image space. However, we can construct a rich collection of sub-space decompositions of the image space by iterating the TCF in different bands [31], gaining access to information from different layers in the context of the convolution-based network of representations that is used to create WPs bases [30], [31].This rich space-scale family of decompositions of the image space (organized in a network of quad-trees) can be used to find not only new convolutional-based features that could be more effective for texture discrimination, but also new representations that can be adapted to the task (i.e., learned from some available data).

A. CONTRIBUTION
The contributions of this study include the following:
- We revisit the statistical framework in [4] to include a family of transform-based representations of the image space with a non-uniform number of transform coefficients per sub-band, and within this context, we introduce WPs for texture retrieval.

- We formulate the design of WP discriminative features (for texture retrieval) as a basis selection (BS) problem.
- BS is presented as a learning-decision task to identify the basis that offers the optimal tradeoff between approximation (feature discrimination) and estimation (complexity) errors.
- We show that BS is equivalent to a minimum cost-tree pruning problem with natural connections with the type of minimum cost-tree pruning algorithms used in classification and regression trees (CART) [48].
- A systematic experimental validation of the proposed WP solutions is presented covering synthetic data as well as four real datasets with different image sizes and texture classes. The results are promising and demonstrate the ability of our framework to obtain task-oriented discriminative features (learnt from data).
- Finally, we show that WP representations offer concrete improvements when compared with Wavelet based features in several scenarios. The new designed WP features explore new frequency bands that help to enhance the discrimination among texture classes.

B. ORGANIZATION
The rest of the paper is organized as follows: Section II introduces WPs and the setting of the HT problem. Section III presents the mathematical formulation of WP based texture retrieval and Section IV addresses the learning problem of WP basis selection. Section V summarizes the steps implemented in our methodology. Section VI presents the experimental validation with real and synthetic data. Section VII discusses the connection between our work and deep learning representations. Section VIII concludes our study providing final remarks and future directions for this research.

II. PRELIMINARIES
This section provides a brief background on WP and the statistical setting used for the indexing problem. Comprehensive expositions can be found in [30], [31], [49].

A. WAVELET PACKETS BASES
Let $\mathbb{X}$ be the image space (or pixel domain) that is assumed to have a finite *level of scale* 2^L or *resolution* 2^{-L}, where L is a fixed integer strictly greater than zero [31]. Then, $\mathbb{X}$ is equipped with an orthonormal basis $\mathbf{B}_L \equiv \left\{ (\psi_L(t_1 - 2^L n_1, t_2 - 2^L n_2))_{(t_1,t_2) \in \mathbb{R}^2} : (n_1, n_2) \in \mathbb{Z}^2 \right\}$ that is generated by the separable product of Wavelet Packet bases [30], [31], [49]. For the 2D case, the WP framework provides a way of decomposing $\mathbf{B}_L$ into four orthonormal collections that we denote by $\mathbf{B}_{L+1}^p \equiv \left\{ (\psi_{L+1}^p(t_1 - 2^{L+1} n_1, t_2 - 2^{L+1} n_2))_{(t_1,t_2) \in \mathbb{R}^2} : (n_1, n_2) \in \mathbb{Z}^2 \right\}$ for $p \in \{0, 1, 2, 3\}$.

If we denote by $U_{L+1}^p \equiv$

$$span\left\{ (\psi_{L+1}^p(t_1 - 2^{L+1} n_1, t_2 - 2^{L+1} n_2)) : (n_1, n_2) \in \mathbb{Z}^2 \right\}$$
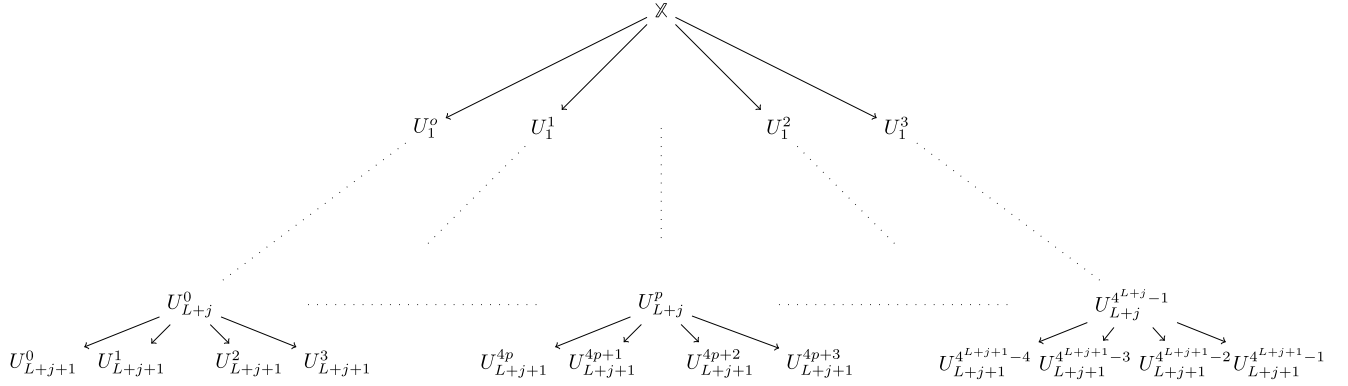
FIGURE 1. Wavelet packet tree-structured sub-space decomposition.

for any $p \in \{0, 1, 2, 3\}$, then, we have the following sub-space decomposition [31][1]

$$\mathbb{X} = U_{L+1}^0 \oplus U_{L+1}^1 \oplus U_{L+1}^2 \oplus U_{L+1}^3. \tag{1}$$

The tree-structure of the WP bases comes from the fact that $\{\mathbf{B}_{L+1}^p : p \in \{0, 1, 2, 3\}\}$ are induced by a (discrete) pair of *conjugate mirror filters* (CMF) that we denote by $(h_0(n), h_1(n))_{n \in \mathbb{N}}$ [31, Chap. 7.1.3]. In other words, the basis elements associated with the scale $L + 1$, $\psi_{L+1}(t_1, t_2)$ are induced from $\psi_L(t_1, t_2)$ and $(h_0(n), h_1(n))_{n \in \mathbb{N}}$ (details are provided in Appendix). By iterating this constructive approach on each basis element in $\{(\psi_{L+1}^p(t_1, t_2)), p \in \{0, 1, 2, 3\}\}$ [31, Th. 8.1], the process continues (in a tree-structured way) with the construction of bases and sub-space decompositions for $\mathbb{X}$. This process is illustrated in Fig. 1. In particular after j iterations, we create $\psi_{L+j}^p(t_1, t_2)$ for any $p \in \{0, \ldots, 4^j - 1\}$ where $U_{L+j}^p = span\{(\psi_{L+j}^p(t_1 - 2^{L+j}n_1, t_2 - 2^{L+j}n_2)) : (n_1, n_2) \in \mathbb{Z}^2\}$ and

$$U_{L+j}^p = U_{L+j+1}^{4p} \oplus U_{L+j+1}^{4p+1} \oplus U_{L+j+1}^{4p+2} \oplus U_{L+j+1}^{4p+3}. \tag{2}$$

As a result, the WP creates a family of tree-structured bases induced by the iteration of the TFC, a construction that is illustrated in Fig. 2b.

Finally, a basis in the WP collection is indexed by $\{(j_i, p_i) : i = 1, \ldots, M\}$, its basis elements are $B = \bigcup_{i=1}^M \mathbf{B}_{j_i}^{p_i}$, and its sub-space decomposition is $\mathbb{X} = \bigoplus_{i=1}^M U_{j_i}^{p_i}$. Importantly, for each (j_i, p_i), we can obtain an equivalent filter $H^i(z) \equiv H_{\Theta(j_i, p_i)}(z_1, z_2)$ by Eq. (22) (see the details in Proposition 1, Appendix) and, consequently, we reduce the analysis to the application (and frequency response) of M-channel filters derived from $\{(j_i, p_i) : i = 1, \ldots, M\}$. Two examples of the frequency responses are presented in Fig. 2, where we notice that for the Wavelet type of structure a solution that increases the resolution in the low frequency range is obtained.[2]

---

[1]WP sub-spaces are separable as well as their bases, which allows us to represent the 2D WP analysis as an equivalent 1D WP analysis by properly organizing the rows and columns of the image [31, Chap. 3].

[2]In general, we reduce the frequency support of the resulting sub-space by a quarter in each step that we iterate the TCF as illustrated in Fig. 2c.

B. THE HYPOTHESIS TESTING FORMULATION FOR CONTENT RETRIEVAL

A query image $\mathbf{x}$ is represented by a set of observations in a finite dimensional space $\mathbb{X} = \mathbb{R}^D$. Each candidate image in the database is represented by a statistical model (or a probability measure in $\mathbb{X}$) that we denote by $\{\mu_{\theta_i} : i = 1, \ldots, M\}$, where $\mu_{\theta_i} \in \mathcal{P}(\mathbb{X})$. Then, the (content-based) indexing of $\mathbf{x}$ reduces to find the closest N candidates in $\{\mu_{\theta_i} : i = 1, \ldots, M\}$ by applying the *maximum likelihood (ML) criterion*, i.e.,

$$k_i = \arg \max_{k \in [M] \setminus \{k_1, \ldots, k_{i-1}\}} \mu_{\theta_k}(\{\mathbf{x}\}), \quad \forall i = 1, \ldots, N, \tag{3}$$

where $[M] \equiv \{1, \ldots, M\}$.

Do and Vetterli [4, Section II] proposed an efficient two-stage principle that asymptotically (in the size of the image) approximates the ML criterion in Eq. (3). Here, we present a version of this result that considers a non-uniform rate of information (coefficients) per sub-band, which is needed for the case of WP texture analysis. Let us assume that μ_{θ_i} is equipped with a density function $f_{\theta_i}(\mathbf{x})$[3] that is fully characterized by a vector of parameters $\theta_i \in \Theta$. Then, we use the likelihood $\log f_\theta(\mathbf{x})$ for each $\theta \in \bar{\Theta} = \{\theta_i : i = 1, \ldots, M\}$ to solve Eq. (3). Furthermore, we consider that $\mathbb{X} = \bigotimes_{i=1}^T \mathbb{X}_i$, i.e., $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ where $\mathbf{x_t} = (x_{t,1}, \ldots, x_{t,D_t}) \in \mathbb{X}_t$ is a finite dimensional vector of dimension $L_t \geq 1$ and, consequently, $D = \sum_{t=1}^T D_t$. Consistent with this cartesian product splitting of $\mathbb{X}$, we assume that each f_{θ_i} with $\theta_i \in \bar{\Theta}$ decomposes in T independent (product) components indexed by $\theta_i = (\theta_i^1, \ldots, \theta_i^T)$, where

$$\log f_{\theta_i}(\mathbf{x}) = \sum_{t=1}^T \log \prod_{j=1}^{D_t} \left(f_{\theta_i^t}(\mathbf{x}_{t,j}) \right),$$
$$= \sum_{t=1}^T \left(\sum_{j=1}^{D_t} \log f_{\theta_i^t}(x_{t,j}) \right). \tag{4}$$

More specifically, Eq. (4) means that $\mathbf{x}_1, \ldots, \mathbf{x}_T$ is a set of independent but non identically distributed vectors, where

---

[3]This means that the density function $f_{\theta_i}(\mathbf{x}) = \frac{\partial \mu_{\theta_i}}{\partial \lambda}(\mathbf{x})$ is well defined.

(a) Wavelet case

(b) Balanced tree case

(c) Frequency response of equivalent filter-bank of fig.2a

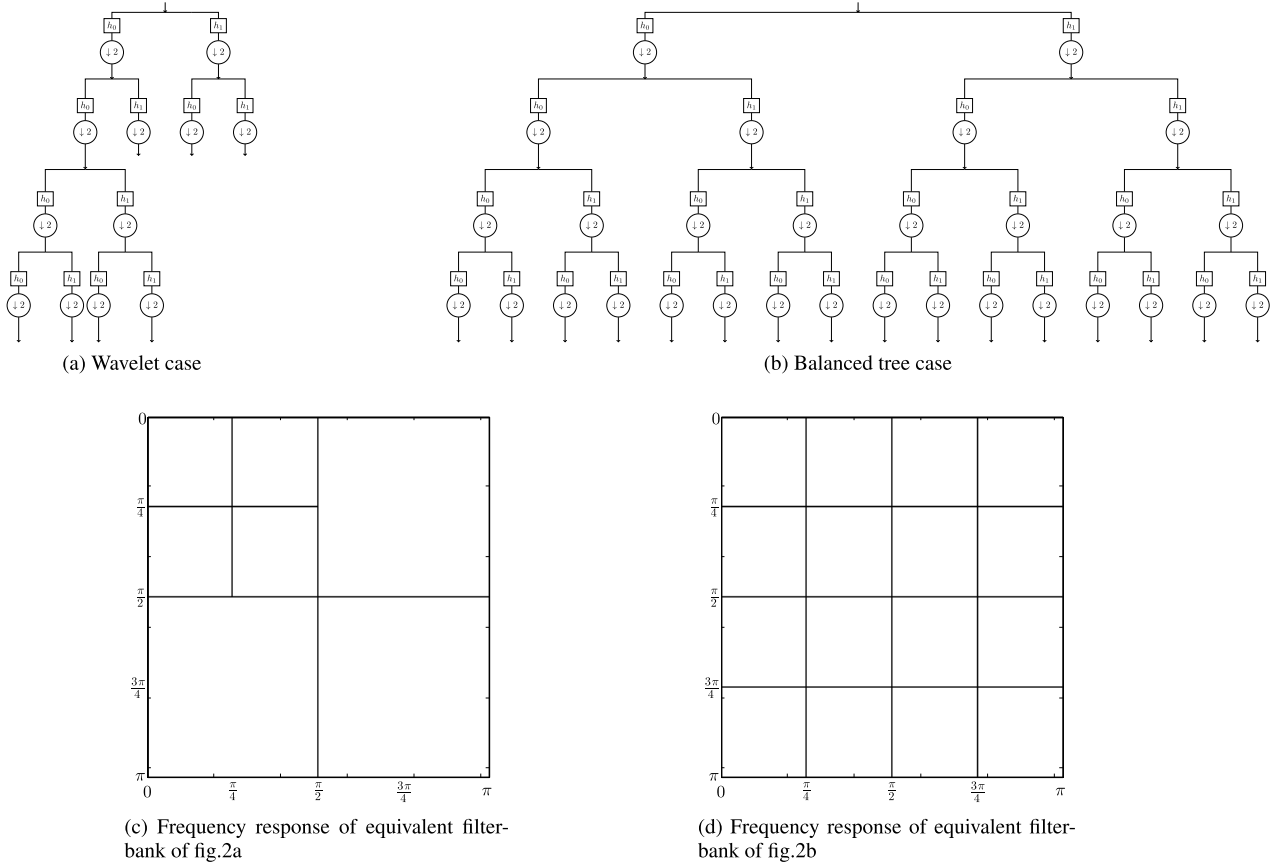(d) Frequency response of equivalent filter-bank of fig.2b

FIGURE 2. Illustration of the frequency division of Wavelet Packet bases for two tree structures. The ideal *Shannon conjugate filter pair* is considered, which provides perfect dyadic partitions of the interval $[-\pi, \pi] \times [-\pi, \pi]$. Scenario (2a-2c) shows a iteration of $H_0(z_1)H_0(z_2)$ (Wavelet type), and scenario (2b-2d) presents a balanced tree structure (uniform frequency resolution).

each t component $\mathbf{x}_t$ corresponds to i.i.d. realizations of the density $f_{\theta_i^t}$.

We assume that the query image $\mathbf{x}$ comes from an underlying parametric model $\mu_{\theta_q} \in \mathcal{P}(\mathbb{X})$ that has the same independent structure stated for the database models, i.e., $\theta_q = (\theta_q^1, \ldots, \theta_q^T)$ and $x_{t,1}, \ldots, x_{t,D_t}$ are i.i.d. realizations of $f_{\theta_q^t}$ for each $t \in \{1, \ldots, T\}$. In this context, if we fix t and we take $D_t \longrightarrow \infty$ (and consequently $D \longrightarrow \infty$), the law of large numbers implies that [50]

$$\lim_{D_t \longrightarrow \infty} -\frac{1}{D_t} \sum_{j=1}^{D_t} \log f_{\theta_i^t}(x_{t,j}) = -\mathbb{E}_{X \sim f_{\theta_q^t}} \log f_{\theta_i^t}(X)$$
$$= D(f_{\theta_q^t} || f_{\theta_i^t}) + h(f_{\theta_q^t}), \quad (5)$$

almost surely. Here $D(f_{\theta_q^t} || f_{\theta_i^t})$ is the Kullback-Leibler divergence [51] of $f_{\theta_q^t}$ with respect to $f_{\theta_i^t}$ and $h(f_{\theta_q^t})$ is the differential entropy [51] of $f_{\theta_q^t}$. Finally, assuming a non-uniform rate of convergence, we define

$$\mathbf{w}_t = \lim_{D \longrightarrow \infty} \frac{D_t}{D}, \quad (6)$$

for all $t \in \{1, \ldots, T\}$ such that $\sum_{t=1}^{T} \mathbf{w}_t = 1$. Then, asymptotically (as the number of observations goes to infinity) the ML principle in Eq. (3) reduces to the *minimum weighted*

divergence decision:

$$k_i = \arg \min_{k \in [M] \backslash \{k_1, \ldots, k_{i-1}\}} \sum_{t=1}^{T} \mathbf{w}_t \cdot D(f_{\theta_q^t} || f_{\theta_k^t}), \quad (7)$$

$\forall i = 1, .., N$.

We will show that the non-uniform rate assumption in Eq. (6) is important for the adoption of this HT framework in the context of WPs.

III. WAVELET PACKET BASED TEXTURE RETRIEVAL

We contextualize the framework in Section II-B for texture indexing when images are represented in a WP domain. For that, we propose an extension of the inter-band independence texture model structure in [4], [16], [33] considering the sub-space partitions induced by WPs. First, we introduce some important notations and nomenclature to formalize the idea.

A. ROOTED TREE-REPRESENTATION

We adopt a rooted tree notation for the WPs family. A tree-structure creates a particular WP basis by iterating the four-channel filter (FCF) as illustrated in Fig. 2b. Let J denotes the maximum number of iterations of the sub-band

decomposition and let $G = (V, E)$ be a graph with

$$V = \{(0, 0), (1, 0), \ldots, (1, 3), \ldots, (J, 0), \ldots, (J, 4^J - 1)\}$$

and E be the collection of arcs on $V \times V$ that characterizes the full balanced rooted tree with root $v_{root} = (0, 0)$ in Fig. 1. Instead of representing a tree as a collection of arcs in G, we use the convention used by Breiman *et al.* [48] in which sub-graphs are represented as a subset of nodes of the full graph. In particular, a *rooted quad-tree* $\mathcal{T} = \{v_0, v_1, \ldots\} \subset V$ is defined as a collection of nodes: the root, internal nodes and leaf nodes. We define $\mathcal{L}(\mathcal{T})$ as the set of leaves of $\mathcal{T}$ and $\mathcal{I}(\mathcal{T})$ as the set of internal nodes, where consequently $\mathcal{T} = \mathcal{L}(\mathcal{T}) \cup \mathcal{I}(\mathcal{T})$. We say that a rooted quad-tree S is a subtree of $\mathcal{T}$ if $S \subset \mathcal{T}$, and if the root of S and $\mathcal{T}$ are the same. Then, S is a pruned version of $\mathcal{T}$, denoted by $S \ll \mathcal{T}$. If the root of S is an internal node of $\mathcal{T}$, then S is a branch of $\mathcal{T}$. For any $v \in T$, we denote the largest branch of $\mathcal{T}$ rooted at v by $\mathcal{T}_v$. The size of a rooted quad-tree $\mathcal{T}$ is the cardinality of $\mathcal{L}(\mathcal{T})$ and is denoted by $|\mathcal{T}|$. Finally, $\mathcal{T}_{full} \equiv V$ denotes the full quad-tree, and consequently, the collection of WP bases will be indexed by $\{\mathcal{T} \subset V : \mathcal{T} \ll \mathcal{T}_{full}\}$.

Any pruned version of $\mathcal{T}_{full}$ in Fig. 2b represents a particular WP basis by the iteration of the FCF. In particular, for an arbitrary rooted tree $\mathcal{T} \ll \mathcal{T}_{full}$ of size M, each of its leaf nodes $\mathcal{L}(\mathcal{T}) = \{(j_k, p_k) : k = 1, \ldots, M\} \subset V$ represents a sub-space generated by the application of WP. Then, the sub-space decomposition produced by WP with tree-representation $\mathcal{T}$ is given by:

$$\mathbf{U}_{\mathcal{T}} = \left\{ U_{j_k}^{p_k} : k = 1, \ldots, M \right\}, \tag{8}$$

where $\mathbb{X} = \bigoplus_{k=1}^{M} U_{j_k}^{p_k}$. Each of these sub-spaces is induced by a basis $\mathbf{B}_{j_k}^{p_k}$ with $k = 1, \cdots, M$, so the WP basis induced by $\mathcal{T}$ is given by,

$$\mathbf{B}_{\mathcal{T}} = \cup_{k=1}^{M} \mathbf{B}_{j_k}^{p_k}, \tag{9}$$

Finally, for any $x \in \mathbb{X}$, we can determine the transformed coefficients by projecting the image in the elements of $\mathbf{B}_{\mathcal{T}}$ (see Eq. (20) in Appendix). In particular, considering the realistic finite dimensional case, where $dim(\mathbb{X}) = 4^J$ (2D dyadic image)[4] for any (j_k, p_k), the projection of x in $U_{p_k}^{j_k}$ is determined by 4^{J-j_k} transformed coefficients that are obtained by Eq.(21). For simplicity, we map the index $(n_1, n_2) \in \{1, \ldots, 2^{J-j_k}\}^2$ to $n \in \{1, \ldots, 4^{J-j_k}\}$ to represent the transformed coefficients as a 1D vector. Thus, the transformed coefficients of x in $U_{j_k}^{p_k}$ are denoted by $D_{j_k}^{p_k}(x) = \left(d_{j_k}^{p_k}(x, n) \right)_{n=1, \ldots, 4^{J-j_k}}$ and the transformed coefficients for the basis $\mathbf{B}_{\mathcal{T}}$ are given and represented by

$$D_{\mathcal{T}}(x) = (D_{j_k}^{p_k}(x))_{k=1, \ldots, M}. \tag{10}$$

It is worth noting that the number of transformed coefficients of the node $(j, p) \in V$ scales like 4^{J-j}.

---

[4]Without loss of generality the pixel based representation of x corresponds to the transformed coefficients of the trivial WP basis $\{v_{root}\} \ll \mathcal{T}_{full}$, i.e., $(x(n)) = (d_0^0(x, n) : n = 1, \ldots, 4^J)$.

B. TEXTURE RETRIEVAL

Let us consider an arbitrary tree $\mathcal{T} \ll \mathcal{T}_{full}$ with $\mathcal{L}(\mathcal{T}) = \{(j_k, p_k) : k = 1, \ldots, K\} \subset V$ and X be a random image in $\mathbb{X}$. The texture model of X is determined in the transformed domain $\mathbf{B}_{\mathcal{T}}$. More precisely, the transform-based random vector $D_{\mathcal{T}}(X) = (D_{j_k}^{p_k}(X))_{k=1, \ldots, K}$ in Eq. (10) follows a parametric model $\mu_{\bar{\theta}}$ with density $f_{\bar{\theta}}$, where $\bar{\theta} = (\theta^{((p_k, j_k))} : k = 1, \ldots, K) \in \Theta$ and for any $x \in \mathbb{X}$ its pdf has the following product of marginal structure:

$$\underbrace{\log f_{\bar{\theta}}(D_{\mathcal{T}}(x))}_{\text{joint pdf in the transform domain}}$$

$$= \sum_{k=1}^{K} \log \underbrace{\prod_{n=1}^{4^{(J-j_k)}} f_{\theta^{((p_k, j_k))}}(d_{j_k}^{p_k}(x, n))}_{\text{marginal pdf of the sub-band } U_{j_k}^{p_k}}$$

$$= \sum_{k=1}^{K} \left(\underbrace{\sum_{n=1}^{4^{(J-j_k)}} \log f_{\theta^{((p_k, j_k))}}(d_{j_k}^{p_k}(x, n))}_{i.i.d. \text{ assumption for the transform coeff. at } U_{j_k}^{p_k}} \right), \tag{11}$$

where $f_{\theta^{((p_k, j_k))}}$ is a pdf in $\mathbb{R}$ for each $k = 1, \ldots, K$. From Eq. (11), the components of the image projected at the sub-bands of WP are independent, and within each sub-band its transformed coefficients are i.i.d. characterized by a pdf. For this pdf, we consider the simplest model adopted in [4], i.e., we consider the Generalized Gaussian model (GGM) with zero mean and parametrized by $\theta = (\alpha, \beta) \in \mathbb{R}^2$ in the following way:

$$f_{\theta}(d) = \frac{\beta}{2\alpha \Gamma(1/\beta)} e^{-\left(\frac{|d|}{\alpha}\right)^{\beta}}. \tag{12}$$

Finally, it is important to note the scaling on the number of transformed coefficients for an arbitrary node $(j_k, p_k) \in \mathcal{L}(\mathcal{T})$. If we denote the size of the image by $L = 4^J$ for some $J > 0$ (the dyadic case studied in Section III-A) then the size of the vector $D_{j_k}^{p_k}(X)$ is $L/4^{j_k}$, which is an exclusive function of j_k (the number of arcs that connects (j_k, p_k) with the root $(0, 0)$). If $\mathcal{T}$ is not a balanced tree [48], [52], we have an asymmetric number of transformed coefficients per sub-band. This should be considered in the asymptotic connection derived between ML and the divergence principle in Eq. (7).

In the context of the texture indexing problem, we have M probability models representing the texture database $\{f_{\bar{\theta}_i} : i = 1, \ldots, M\}$, each of them following the model in Eq. (11) and, consequently, they are fully characterized by

$$\bar{\Theta} = \left\{ \bar{\theta}_i = (\theta_i^{((j_k, p_k))})_{k=1, \ldots, K} : i = 1, \ldots, M \right\} \subset \Theta,$$

with $\Theta = \mathbb{R}^{2 \cdot |\mathcal{T}|}$. In addition, we assume an underlying (hidden) query model $f_{\bar{\theta}_q}$ consistent with Eq. (11) and parametrized by $\bar{\theta}_q = (\theta_q^{((j_k, p_k))})_{k=1, \ldots, K} \in \Theta$ that produces

a realization x. Then, the solution of (3), considering the regime $L \longrightarrow \infty$ and $\mathcal{T} \ll \mathcal{T}_{full}$, reduces to:

$$
\begin{aligned}
k_i &= \arg \min_{l \in [M] \backslash \{k_1, \ldots, k_{i-1}\}} \sum_{k=1}^{K} \mathbf{w}_{(j_k, p_k)} D(f_{\theta_q^{((j_k, p_k))}} \| f_{\theta_l^{(j_k, p_k)}}), \\
&= \arg \min_{l \in [M] \backslash \{k_1, \ldots, k_{i-1}\}} \sum_{k=1}^{K} \mathbf{w}_{(j_k, p_k)} \\
&\quad \cdot \left[\log \left(\frac{\beta_q^{((j_k, p_k))} \alpha_l^{((j_k, p_k))} \Gamma(1/\beta_l^{((j_k, p_k))})}{\beta_l^{((j_k, p_k))} \alpha_q^{((j_k, p_k))} \Gamma(1/\beta_q^{((j_k, p_k))}))} \right) \right. \\
&\quad + \left(\frac{\alpha_q^{((j_k, p_k))}}{\alpha_l^{((j_k, p_k))}} \right)^{\beta_l^{((j_k, p_k))}} \frac{\Gamma((\beta_l^{((j_k, p_k))} + 1)/\beta_q^{((j_k, p_k))})}{\Gamma(1/\beta_q^{((j_k, p_k))})} \\
&\quad \left. - \frac{1}{\beta_q^{((j_k, p_k))}} \right],
\end{aligned}
\tag{13}
$$

from Eqs. (7) and (12).

Finally, to implement Eq. (13) a first-stage (feature extraction) is conducted to obtain a sufficient statistic that summarizes the information of x in each of the sub-band indexed by (j_k, p_k). In particular, the ML criterion is used to estimate $\theta_q^{((j_k, p_k))} = (\alpha_q^{((j_k, p_k))}, \beta_q^{((j_k, p_k))})$ from the information of $D_{j_k}^{p_k}(x)$ in Eq.(10) for each $k = 1, \ldots, K$.

Remark 1: It is worth noting the importance of the non-uniform scaling on the number of transformed coefficients considered in Section II-B, because it follows that $\mathbf{w}_{(j_k, p_k)} = 4^{-j_k}$ for all $k = 1, \ldots, K$ in (13). Consequently, for the selection of the closest M models, the terminal nodes that are closer to the root are more significant in the decision than nodes that are deeper in the tree. This observation is due to the number of coefficients of the first groups, which is orders of magnitude greater than the number of coefficients of the second group.

IV. WAVELET PACKET BASIS SELECTION

Any tree $\mathcal{T} \ll \mathcal{T}_{full}$ in the family of WPs provides a valid representations for the indexing problem. This rises the problem of basis selection (BS). A clear objective for this task is seeking the basis that maximizes the discrimination among the texture classes considering the HT in Section III-B. However, texture discrimination is not an exclusive criterion for this task. The complexity of the tree also needs to be considered, as a large tree (in terms of the number of leaves) implies deeper leaves with their reduced number of transformed coefficients to estimate the parameters in the FE phase of the indexing task (see Eq. (13)). This issue rises the existence of a non-trivial estimation error in the FE phase that needs to be considered for BS. For that reason, the BS can be posed as a statistical learning problem that finds an optimal balance between an estimation and an approximation error [48], [52]–[54]. In particular, we state the following regularization problem

$$
\mathcal{T}^*(\lambda) = \arg \min_{\mathcal{T} \ll \mathcal{T}_{full}} -\hat{R}(\mathcal{T}) + \lambda \cdot \Phi(\mathcal{T}),
\tag{14}
$$

where $\hat{R}(\mathcal{T})$ models the discrimination quality of the features induced by $\mathcal{T}$ and $\Phi(\mathcal{T})$ represents its learning complexity. λ is a regularization parameter that models the compromise between the fidelity and cost in this context. In particular, we adopt the tree size as a metric for $\Phi(\mathcal{T}) = |\mathcal{T}|$, as it has been used in CART [48] and other tree learning problems [52], [55]–[58] to model estimation errors. The assumption here is that the deviation of the estimated parameters in Eq. (13) from the true parameters is proportional to the size of the tree [48], [52].

For the fidelity measure, $\hat{R}(\mathcal{T})$, we consider a global indicator of pair-wise weighted divergence, used in Eq. (13), between classes given by

$$
\frac{1}{M(M-1)} \sum_{c=1}^{M} \sum_{\substack{k=1 \\ k \neq c}}^{M} \left(\sum_{t=1}^{T} w_{(j_t, p_t)} D(f_{\theta_c^{((j_t, p_t))}} \| f_{\theta_k^{((j_t, p_t))}}) \right), \tag{15}
$$

where $\overline{\Theta} = \left\{ \bar{\theta}_c = \left(\theta_c^{((j_t, p_t))} \right)_{t=1, \ldots, T} : c = 1, \ldots, M \right\}$ denotes (in the simplest case) the selection of one model per class from the database. The use of the weighted divergence as an indicator of the discrimination capacity of the indexing task is justified from the *Stein's lemma* [51, Th. 12.8.1], where the weighted divergence determines the error exponent of the type 2 error given a fixed type 1 error in a two class (hypothesis testing) problem.

A. MINIMUM COST-TREE PRUNING ALGORITHM

The type of regularization problem stated in Eq. (14) has been addressed in the context of decision trees for which efficient solutions are available [48], [52]. For the application of these results in our context, the fidelity measure $\hat{R}(\mathcal{T})$ must be additive [48], [52],[5] which follows from its construction as

$$
\hat{R}(\mathcal{T}) = \sum_{t \in \mathcal{L}(\mathcal{T})} \hat{R}((j_t, p_t)), \tag{16}
$$

with $\hat{R}((j_t, p_t))$ defined as

$$
\frac{1}{M(M-1)} \sum_{c=1}^{M} \sum_{\substack{k=1 \\ k \neq c}}^{M} \mathbf{w}_{(j_t, p_t)} D(f_{\bar{\theta}_c^{((j_t, p_t))}} \| f_{\bar{\theta}_k^{((j_t, p_t))}}).
$$

Then, we can use the following result:

Theorem 1 (Scott [52, Th. 1] and Chou et al. [58, Th. 3]): There is a sequence of embedded tree-structures $\mathcal{T}_{full} \gg R_1 \gg R_2 \gg \ldots \gg R_m = \{root\}$ and weights $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m = \infty$ such that for any $l \in \{1, \ldots, m\}$, $\mathcal{T}^*(\lambda) = R_l$ for all $\lambda \in [\lambda_{l-1}, \lambda_l)$.

This result states that there is a family of embedded trees that offers all the admissible solutions for Eq. (14), in the sense that $\{\mathcal{T}^*(\lambda), \lambda \geq 0\} = \{R_1, \ldots, R_m\}$. The specification of the values of $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m = \infty$ and their respective trees $\{R_1, \ldots, R_m\}$ are called the solution of the *family pruning problem* [48], [52].

---

[5]A functional $f(\mathcal{T})$ is additive if it is the sum of components of the leaves of $\mathcal{T}$ [52].

In addition, it is well-known from [57] that the set of possible solutions of Eq. (14), i.e., $\{\mathcal{T}^*(\lambda), \lambda \geq 0\}$, is given by the following *cost-fidelity problem* [52], [57]

$$\mathcal{T}^{k*} = \arg \max_{\{\mathcal{T} \ll \mathcal{T}_{full} : |\mathcal{T}| \leq k\}} \hat{R}(\mathcal{T}), \qquad (17)$$

for all $k = 1, 4, \ldots, |\mathcal{T}_{full}|$. For any admissible quad-tree size $k > 0$, Eq. (17) finds the tree of size equal to or smaller than k that maximizes the fidelity measure. By connecting these two characterizations (cost-fidelity solutions and the family pruning problem), it is direct to show that $\{\mathcal{T}^{k*}, k = 3\,l + 1 \text{ with } l = 0, \ldots, (|\mathcal{T}_{full}| - 1)/3\} = \{R_1, \ldots, R_m\}$ and, consequently, the solutions of the cost-fidelity problem in Eq. (17) are embedded in the sense that $\{root\} = \mathcal{T}^{1*} \ll \mathcal{T}^{4*} \ll \ldots \ll \mathcal{T}^{(3\,l+1)*}$, where l defines the number of tree branches. This embedded structure is the key to obtain implementable algorithms to solve $\{\mathcal{T}^{k*}, k = 3 \cdot l + 1 \text{ with } l = 0, \ldots, (|\mathcal{T}_{full}| - 1)/3\}$ and then $\{\mathcal{T}^*(\lambda), \lambda \geq 0\}$. For completeness, Algorithm 1 presents this solution.[6]

Algorithm 1 Proposed Minimum Cost-Tree Algorithm

Require: Tree-structure $\mathcal{T}^k$
  maxNumberLeaves
  while $k \leq$ maxNumberLeaves **do**
    maxCost$= 0$
    $\Delta = 0$
    for $t \in \mathcal{L}(\mathcal{T}^k)$ **do**
      $\mathcal{T}_t = \text{split}(\mathcal{T}^k, t)$
      $\Delta(\mathcal{T}_t^4) = \hat{R}(\mathcal{T}_t^4)$
      if maxCost $< \Delta$ **then**
        maxCost $\leftarrow \Delta$
        $\mathcal{T}^{k+3} \leftarrow \{\mathcal{T}^k \cup \mathcal{L}(\mathcal{T}_t)\} \setminus \{t\}$
      end if
    end for
    Set $k = k + 3$
  end while
Ensure: Sequence of tree-structure decomposition of $\mathcal{T}^k$

Finally, solving Eq. (14) requires to know the true tradeoff between fidelity and cost functions that we denote it by λ^*. In this work, the selection of λ^* is done considering an empirical risk minimization (ERM) approach over the admissible set of tree solutions given by $\{R_1, \ldots, R_m\}$.

V. SUMMARY OF THE MODELING STAGES
The different stages presented in this work in Sections III and IV are summarized in Algorithm 2.

VI. EXPERIMENTAL ANALYSIS
In this section, we present the results for the family of adaptive trees obtained as the solutions of Eq. (17).

---

[6]The complexity of this algorithm is $O(|\mathcal{T}_{full}| \cdot \log(|\mathcal{T}_{full}|))$. See details in [52] and reference therein.

Algorithm 2 Wavelet Packet Texture Retrieval Procedure

  Step 1: Get images from textures without overlap.
  Step 2: Compute Wavelet Packet coefficients from a full decomposition.
  Step 3: Wavelet Packet decomposition is represented as a tree-structure, where each node contains coefficients.
  Step 4: Compute the parameters α and β of the GGM that represent the Wavelet Packet coefficients.
  Step 5: Compute the best basis representation $\mathcal{T}^*$, using Algorithm 1.
  Step 6: Retrieve textures using $\mathcal{T}^*$.

Results will be presented for different datasets to evaluate the adaptive nature of the framework and its performance as a function of the tree size. In particular, we will evaluate the family of WP solutions indexed by $\{\mathcal{T}^{k*}, k = 3\,l + 1 \text{ with } l = 0, \ldots, (|\mathcal{T}_{full}| - 1)/3\}$ in terms of the performance as well as the structure of their filter-bank decomposition and partition of the 2D frequency plane.

A. SYNTHETIC TEXTURE SCENARIO
Before presenting the results on real texture databases, we evaluate the adaptive capacity of our framework on a controlled (synthetic) two texture indexing problem. For this purpose, WP statistical models (following the GGM presented in Section III) were selected by simply choosing a balanced tree of depth 3. In this context, the statistics of the GGMs for each of their leaves were considered the same except for one of the terminal nodes, which is the node that (by design) offers the discrimination power for the task (see the expression in Eq. (7)). The idea of this evaluation was to localize all the texture discrimination information in a specific sub-band to see if the solutions in Eq. (17) promote more resolutions in this target frequency band as k grows. Synthetic samples were created by simulating the transformed coefficients of the two models. From these data, the fidelity measure was estimated using Eq. (15) considering the weighted divergence for 16 examples per class. Fig. 3 reports the estimation of the weighted divergence as well as the energy that was considered in this analysis as a reference (non-discriminative) fidelity indicator. In particular, these figures plots the term $w_{(j_t, p_t)} D(f_{\theta_c}((j_t, p_t)) \| f_{\theta_k}((j_t, p_t)))$ in Eq. (15) associated with the additive contribution of the node (j_t, p_t) indexed by row j_t and column p_t. From these figures, it is possible to see how the weighted divergence appropriately captures the frequency bands that are more discriminative for the task. For the illustration, we consider two contexts where the discriminative bands are indexed by the pair $(3, 21)$ and $(3, 49)$, respectively. From the result, we see that our method captures the discrimination of this task in the right bands and, consequently, $\{\mathcal{T}^{k*}, k = 3\,l + 1 \text{ with } l = 0, \ldots, (|\mathcal{T}_{full}| - 1)/3\}$ increases the resolution in the bands that are more informative. In contrast, Figure 3 shows that the amount of energy in the leaves is not always mapped to the parents and also
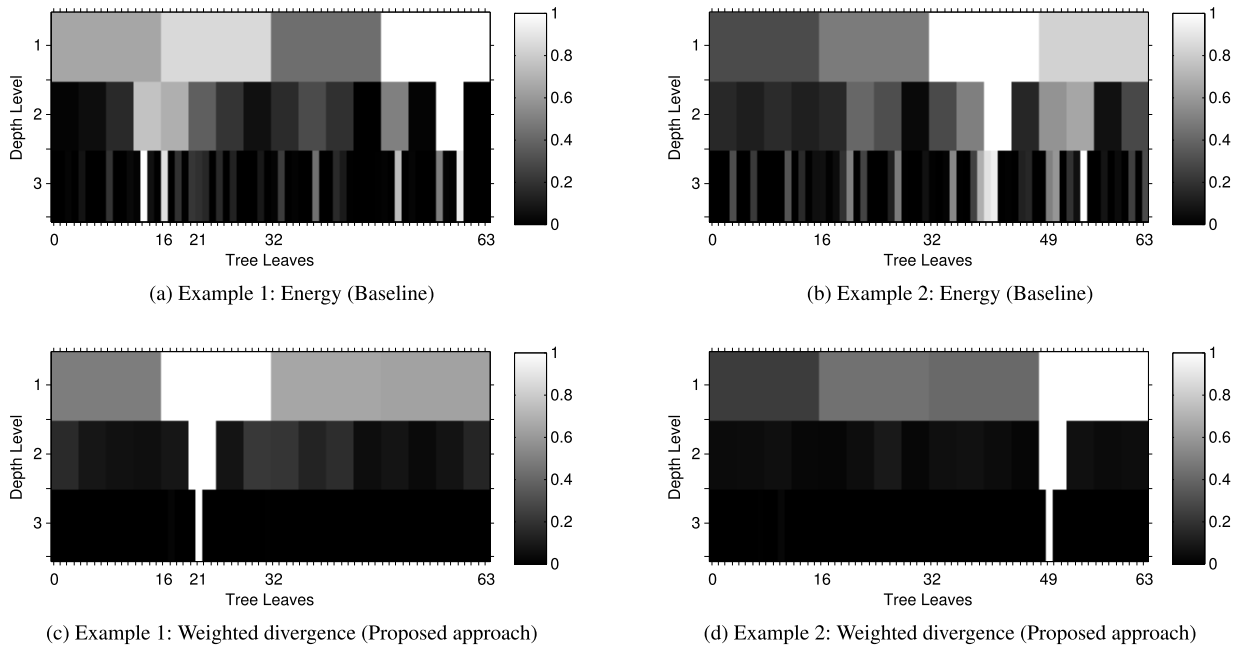
(a) Example 1: Energy (Baseline)

(b) Example 2: Energy (Baseline)

(c) Example 1: Weighted divergence (Proposed approach)

(d) Example 2: Weighted divergence (Proposed approach)

FIGURE 3. These maps show the weighted divergence and energy term indexed by a sub-band in the WP decomposition. For the examples 1 and 2, the discriminative information is in the leaves 21 and 49 (depth level 3), respectively. In both cases, we observe that our proposed approach finds the discriminative sub-bands.

the modified node is not always the most energetic one. As expected, those results indicate that energy as a fidelity measure does not capture the most informative bands for the task.

B. REAL TEXTURE SCENARIOS

For the analysis on real texture datasets, we consider subsets from the VisTex [59], Brodatz [60], and STex [61] datasets. We also consider the ALOT [62] dataset. The subset of the VisTex database corresponds to the 40 color textures (of 512×512 pixels) used in [4]. The subset of Brodatz consists of 30 gray-scale textures of 640×640 pixels each, which is the setting used in [32]. Moving on to larger sets, we extract two collection of images from the STex and ALOT databases, respectively, to create the full STex, full ALOT, reduced STex, and reduced ALOT. The full STex and full ALOT contain 436 color and 250 gray-scale textures of 1024×1024 pixels and 1536×1024 pixels, respectively. The reduced STex and reduced ALOT consist of 40 color and gray-scales textures of 1024×1024 pixels and 1536×1024 pixels, respectively. Each texture from the VisTex and Brodatz datasets is divided into 16 and 25 non-overlapping textures of 128×128 pixels, respectively. Each texture of the Full STex, Full ALOT, Reduced STex, and Reduced ALOT is divided into 16 non-overlapping textures of 256×256 pixels. After these divisions, the color images are transformed into gray-scale versions and, subsequently, all images in the collections are normalized to zero mean and unit variance, which is a standard normalization used in previous studies [4], [5], [8], [9], [63]. In total, we use six databases with different image

sizes and number of textures, which offer a rich context to evaluate the potential of WP based texture indexing.

For each context, we divide the dataset in training and testing in a proportion of $3:7$, where 30% of the data is used to find the tree-structure solving Eq. (17) and the remaining 70% is used for the evaluation of the texture retrieval. The training phase involves computing the fidelity measure in Eq. (15) to obtain the family $\{\mathcal{T}^{k*}, k\}$. The testing phase computes the performances of this family as a function of the size of the tree. For the retrieval performance, we use one query example per class in the dataset and we use the standard average recall metric adopted in [4] given by

$$\text{Recall} = \frac{1}{M \cdot C} \sum_{c=1}^{C} \sum_{m=1}^{M} \mathbb{1}_{\{k_{m,c}=c\}},$$

where C is the number of total classes in the dataset (in our case $C = \{40, 30, 436, 250\}$), M is the number of examples we have per class (16 and 25 depending on the dataset) and $k_{m,c}$ denotes the true class of the M closest retrieved image when the query image belongs to the class c. Finally, the complete analysis uses *Daubechies 4* as the mother Wavelet basis because it is the one that shows the best performances for our analysis.

1) ANALYSIS OF THE SUB-BAND MODEL FITTING

Considering that WPs offer a rich range of filter-bank decomposition, we numerically evaluate the fitting of the GGM considered in Eq. (12) to model the statistical dependencies of the transform coefficients in each of the induced sub-bands as considered in Section III. We use the ML estimator
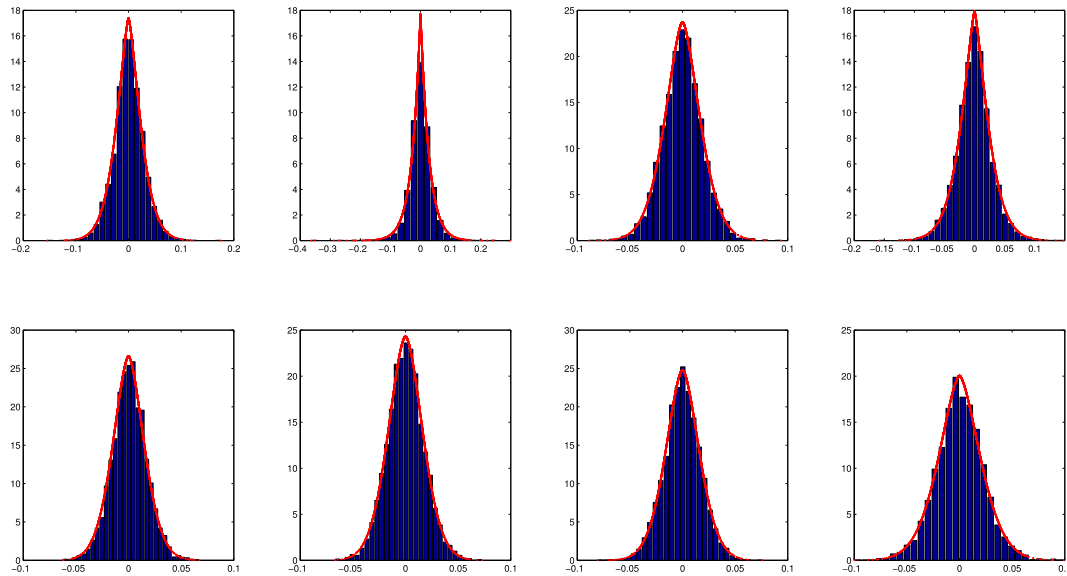
FIGURE 4. Histograms and ML fitting of the WP sub-band coefficients associated with the nodes indexed by (2, 4), (2, 5), (2, 6), (2, 7), (2, 12), (2, 13), (2, 14), (2, 15). The Generalized Gaussian Model in Eq. (12) is used for the histogram fitting. The image *20_c1l1* from ALOT database is used for illustration.

of its parameters for numerous sub-bands and textures in our rich collection of datasets. In general, it was observed that the GGM model captures the marginal statistics of the transformed coefficients of any arbitrary band and, consequently, the modeling extension adopted in Section III seems reasonable. For illustration, we present some of these fittings in Fig. 4.

2) WAVELET PACKETS RETRIEVAL PERFORMANCES

We present the retrieval performance associated with the solutions of Eq. (17) for each of the datasets adopting the weighted divergence in Eq. (15). For completeness, we also consider the tree solutions obtained for the same regularization problem but adopting the non-weighted divergence and the energy per-band as alternatives fidelity measures.

Fig. 5 shows the retrieval performance as a function of the size of the tree for the three fidelity measures. The dashed lines correspond to the wavelet decomposition, which serves as our baseline. Overall, we observe the expected tradeoff between estimation and approximation errors in the evolution of each of the performance curves. At the beginning, more complex trees significantly improve the retrieval accuracy. Then, the estimation error dominates, implying a saturation, which leads to a drop of the retrieval performances as k increases. It is important to note that these changes in the regime of the performance curves are a function of the complexity of the task. For the smallest (VisTex and Brodatz) and largest (STex and ALOT) image size datasets, these changes occur in the range $[10 - 34]$ of the tree sizes.

From these curves, we can determine the trees that offer the optimal balance between estimation and approximation errors and, consequently, the best performance for the tasks. In general, the solution obtained with the proposed weighted

divergence shows one of the best performance in almost all the scenarios. There are three exceptions that occur for the Full ALOT, Reduced ALOT and Reduced STex. However, the performances deviate from the best solution relative for each case (obtained with the energy fidelity and non-weighted divergence) by only 0.76%, 1.58%, and 0.31%, respectively. It is also interesting to note that WP solutions obtained with the energy as a fidelity measure present competitive results in all the scenarios. The exception is the VisTex and Reduced ALOT datasets where the non-weighted and weighted divergences show the best performance, respectively.

Finally, we provide the performance of the Wavelet representation as a baseline for performance comparison (see dashed lines in Fig. 5). Most notably, we confirm our conjecture that the family of WPs and their richer sub-band decompositions offer representations with relevant improvements in indexing performance. This gain is more prominent in the case of the largest datasets, where alternative WP trees show more significant gains with respect to the conventional Wavelet solution. For the case of smaller databases (VisTex and Brodatz), the improvements are not significant, which can be explained from the observation that in the context of smaller texture images WPs do not have room to take advantage of the discrimination power of non-conventional (Wavelet-type) frequency bands as we rapidly move to the regime were estimation error dominates the performance curves. Interestingly in this context, the smaller-sized trees (sizes 4, 7 and 10) matches the Wavelet solution and, consequently, we recover as part of our basic selection formulation the conventional Wavelet solution [4]. Table 1 summarizes the best performance obtained (optimal trees) for each fidelity measure and the performances of the Wavelet solution. The last column of Table 1 shows the gain in relative percentage
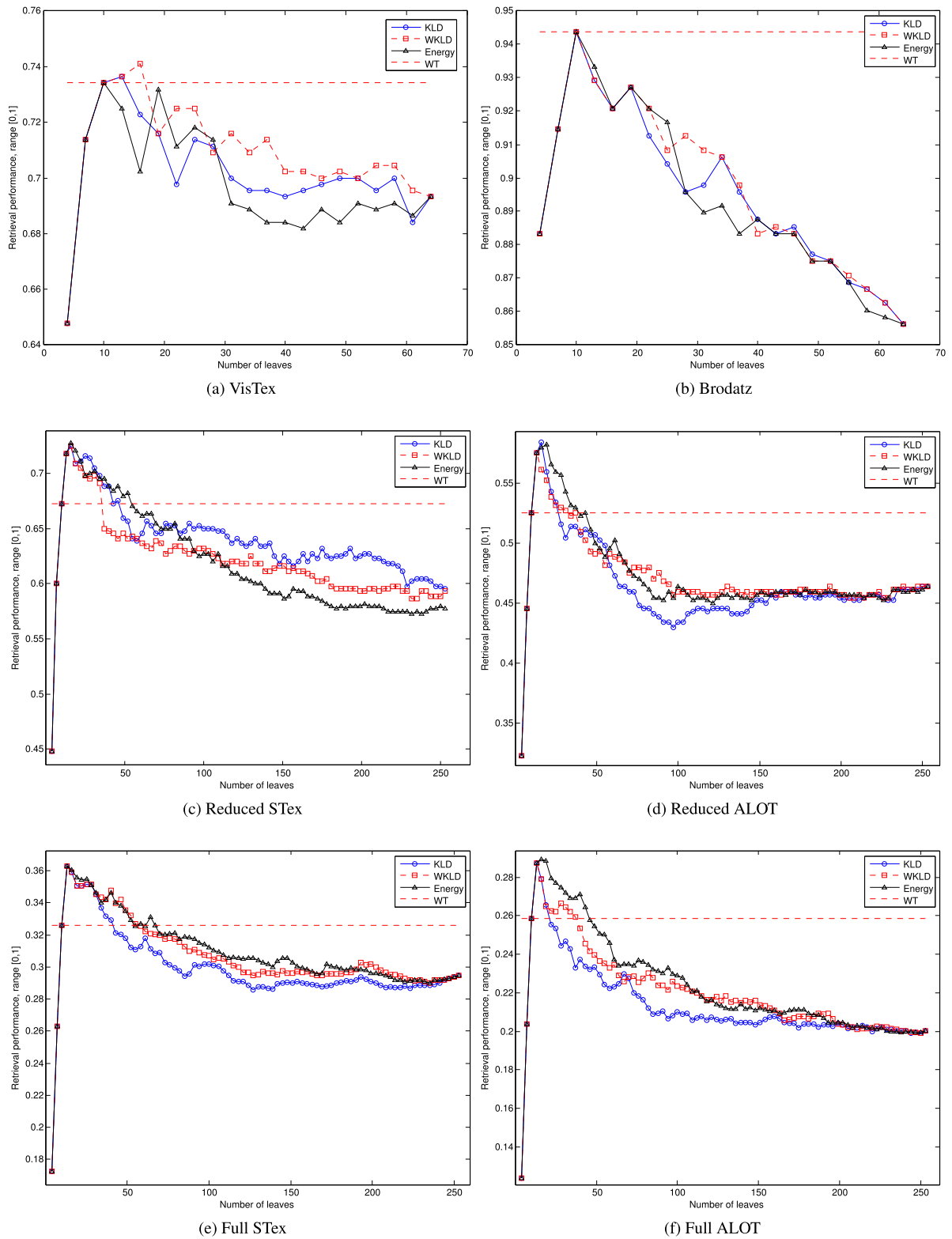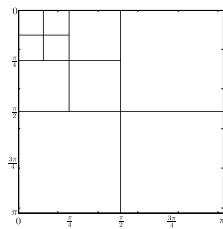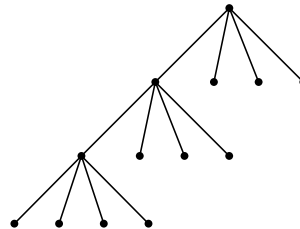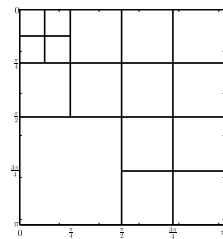
(a) VisTex

(b) Brodatz

(c) Reduced STex

(d) Reduced ALOT

(e) Full STex

(f) Full ALOT

FIGURE 5. Retrieval performance for the family of Wavelet Packets solutions of the regularized problem in Eq. (14) using the weighted divergence, the divergence and the energy as fidelity measures. Results are presented independently for the databases: VisTex, Brodatz, Full STex, Full ALOT, Reduced STex and Reduced ALOT.
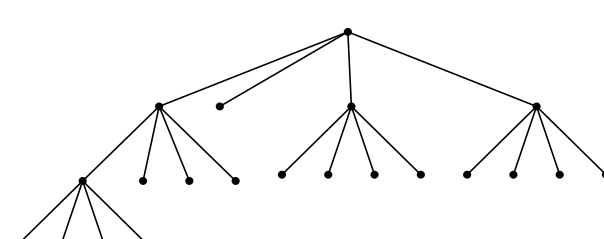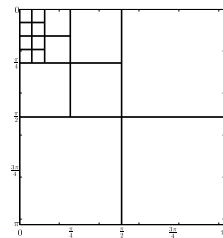
(a) Brodatz frequency partition.
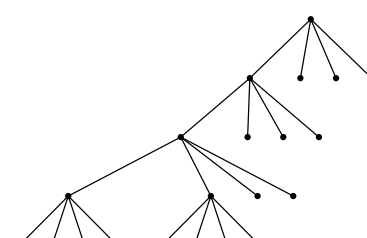

(b) Brodatz tree-structure.
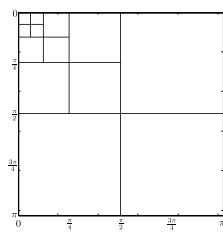

(c) VisTex frequency partition.

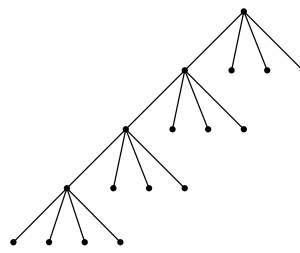(d) VisTex tree-structure.

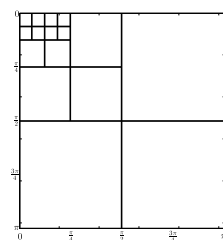(e) Reduced STex and Reduced ALOT frequency partition.

(f) Reduced STex and Reduced ALOT tree-structure.
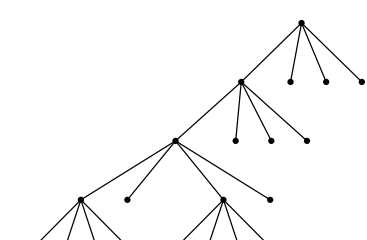

(g) Full STex frequency partition.

(h) Full STex tree-structure.

(i) Full ALOT frequency partition.

(j) Full ALOT tree-structure.

FIGURE 6. First column shows the frequency partitions induced by the best WP solution, which is represented by the tree in the second column.

TABLE 1. Best retrieval performance.

Texture	WT	WP Energy	WP KLD	WP WKLD	Best Relative Gain
VisTex	73.41%	73.41%	73.64%	**74.09%**	0.92%
Brodatz	**94.38%**	**94.38%**	**94.38%**	**94.38%**	0%
Reduced STex	67.27%	**72.73%**	72.50%	72.50%	7.77%
Reduced ALOT	52.50%	58.18%	**58.41%**	57.50%	11.25%
Full STex	32.60%	**36.27%**	**36.27%**	**36.27%**	11.25%
Full ALOT	25.85%	**28.95%**	28.73%	28.73%	11.99%

with respect to the Wavelet representation, which shows that the best improvements in retrieval performance are achieved for the larger-size texture datasets.

3) ANALYSIS OF THE OPTIMAL TREE STRUCTURE OF WPS

The filter-bank decomposition of the best WP solutions for each database are presented in Fig. 6. We observe that the solutions follow a Wavelet type of path as low frequencies are iterated in the decomposition at the beginning of the WP decomposition. However, for many of the scenarios, there are non-trivial deviations from the Wavelet structure as other frequency bands are iterated in the process of creating the optimal trees. These non-Wavelet type bands offer better discrimination than the recursive iteration of the low frequency that defines a Wavelet decomposition.

VII. DISCUSSION: CONNECTION WITH CNN

CNN can be seen as a filter-bank strategy (convolution-based), equipped with a deep architecture and some nonlinear stages designed to learn from data representations that captures salient features in low dimensionality. In the standard and ideal use of CNN, the representations are learned from supervised data in an end-to-end manner by minimizing a loss function (for example the cross entropy or Kullback-Leibler divergence) [37], [41], [43].

The WP framework presented in this work offers some interesting connections with deep learning features obtained from a pre-trained CNN. In terms of structure, our WP-based features share both the convolutional (linear) stage and the nonlinear stage (attributed to a dimensionality reduction phase) presented in CNNs. In our setting, however, a statistical-based reduction is performed for each of the WP sub-bands by estimating the parameters α and β of the GMM models to reduce dimensionality. This lossy reduction per sub-space comes from our model assumption, in the sense that the α and β parameters are sufficient statistics when minimum probability of error criterion is adopted. Much work remains to be done to fully explore the connection between our framework constructed over the idea of WP basis selection with an MPE criterion and the data-driven approaches used to adapt pre-trained CNN for image indexing [22], [23], [26], [29].

VIII. CONCLUSION AND FUTURE WORK

This work shows performance improvements for texture retrieval from the design of new filter-bank discriminatory features using the rich collection of Wavelet Packet (WP) bases. The tree-indexed WP collection was used to find an

adequate balance between feature discrimination and learning (over-fitting) complexity. It should be noted that the WP filter-bank structure was central to address the problem of optimal representation as a minimum cost-tree pruning algorithm, which is reminiscent of the solution proposed in the context of classification and regression trees (CART). We show how adaptive the proposed WP solution is to the nature of the problem in terms of the number of classes and the size of the image (aspects of the problem that are tightly related with over-fitting). Overall, WPs offer features that outperform the Wavelet representation as evidence that the exploration of different sub-bands offers better texture discrimination than the standard Wavelet filter-bank partition.

This work is focused on the signal representation aspects of the transform-based representation. For representation analysis, we adopted one of the basic model structures used in [16], [33]. We leave the investigation of more complex and sophisticated models for future work. Along these lines, we conjecture that the methodology presented will benefit from the adoption of more sophisticated texture models that consider non-trivial spatial intra-band dependencies, inter-band dependencies [13], [16] and more complex parametric distributions [12], [33], [64], [65]. Another interesting direction for future work would be to derive precise expressions for the parameter estimation error taking into account the observation that deeper nodes have less data for the estimation of the model parameters.

APPENDIX
WAVELET PACKETS ANALYSIS

Complementing the exposition in Section II-A, the elements of the basis associated with the scale $L + 1$, $\psi_{L+1}^p(t_1, t_2)$, are induced by $\psi_L(t_1, t_2)$ in the following way: $\psi_{L+1}^p(t_1, t_2) =$

$$\sum_{(n_1,n_2)\in\mathbb{Z}^2} h_{f_1(p),f_2(p)}(n_1, n_2) \cdot \psi_L(t_1 - 2^L n_1, t_2 - 2^L n_2), \quad (18)$$

where $p \in \{0, 1, 2, 3\}$, and $f_1(\cdot)$ and $f_2(\cdot)$ are discrete functions given by [31, Prop. 8.4]:

$$f_1(b) = \begin{cases} 0 & b = 0 \vee b = 1, \\ 1 & b = 2 \vee b = 3. \end{cases}$$

$$f_2(b) = \begin{cases} 0 & b = 0 \vee b = 2, \\ 1 & b = 1 \vee b = 3. \end{cases}$$

and $h_0(n)$ and $h_1(n)$ are related by the *perfect reconstruction property*.[7] By iterating this filter bank approach on each basis element in $\{(\psi_{L+1}^p(t_1, t_2)), p \in \{0, 1, 2, 3\}\}$ [31, Th. 8.1], we can continue, in a tree-structured way. For instance after j iterations, we create $\psi_{L+j}^p(t_1, t_2)$ for any $p \in \{0, \ldots, 4^j - 1\}$ where $U_{L+j}^p = span\{(\psi_{L+j}^p(t_1 - 2^{L+j}n_1, t_2 - 2^{L+j}n_2)) : (n_1, n_2) \in \mathbb{Z}^2\}$, see Fig. 1, where $\forall j \geq 1$,

---

[7] $h_1(n) = (-1)^{1-n} h_0(1 - n), \forall n \in \mathbb{Z}$ [66], [31, Th. 8.1].

$\forall p \in \{0, \ldots, 4^j - 1\}$, $\psi_{L+j+1}^{4p+b}(t_1, t_2)$ is given by:

$$\sum_{(n_1, n_2) \in \mathbb{Z}^2} h_{f_1(b), f_2(b)}(n_1, n_2) \cdot \psi_{L+j}^p(t_1 - 2^{L+j} n_1, t_2 - 2^{L+j} n_2),$$

(19)

with $b \in \{0, 1, 2, 3\}$.

A key property of WP is the inter-scale relationship induced in Eq. (19) among the WP transform coefficients obtained across scales [31]. More precisely, the transform coefficients of $(x(t_1, t_2))$ that belong to $U_j^p \subset \mathbb{X}$ are given by

$$d_j^p(x, n_1, n_2) \equiv \langle x(t_1, t_2), \psi_j^p(t_1 - 2^j n_1, t_2 - 2^j n_2) \rangle, \quad (20)$$

$\forall (n_1, n_2) \in \mathbb{Z}^2$. Using Eq. (2), there is an alternative analysis of $(x(t_1, t_2))$, which is based on the bases associated with U_{j+1}^{4p}, U_{j+1}^{4p+1}, U_{j+1}^{4p+2} and U_{j+1}^{4p+3}, where we have that the transform coefficients $d_{j+1}^{4p+b}(x, n_1, n_2)$ are given by:

$$\langle x(t_1, t_2), \psi_{j+1}^{4p+b}(t_1 - 2^{j+1} n_1, t_2 - 2^{j+1} n_2) \rangle,$$
$$= \sum_{(k_1, k_2) \in \mathbb{Z}^2} h_{f_1(b), f_2(b)}(k_1 - 2n_1, k_2 - 2n_2) \cdot d_j^p(x, k_1, k_2).$$

(21)

for all $b \in \{0, 1, 2, 3\}$. A key aspect of this property is that we move from an analysis in continuous time in Eq. (20), to an analysis (algorithm) in discrete time in Eq. (21).

FILTER-BANK IMPLEMENTATION AND THE WP SUB-SPACE FREQUENCY DECOMPOSITION

The analysis step in Eq. (21) can be implemented by a two channel filter (TCF), followed by a down-sampler operation by 2 (non-linear) [30], [31]. This basic step is extended in the following result:

Proposition 1 [67, Chap. 11.3.3]: Let $x(t_1, t_2)$ be in a finite 2^L scale space $\mathbb{X}$, with transform coefficients $(d_L^0(x, n_1, n_2))_{(n_1, n_2) \in \mathbb{Z}^2}$. Let us consider an arbitrary sub-space U_j^p with $j > L$ and $p \in \{0, \ldots, 4^{j-L} - 1\}$. Let us denote by $(h_0(n))_{n \in \mathbb{Z}}$ and $(h_1(n))_{n \in \mathbb{Z}}$ the conjugate mirror filter pair,[8] by $U_{L+1}^{p_1}, \ldots, U_{j-1}^{p_{j-L-1}}$ the sequence of intermediate sub-spaces used to go from $\mathbb{X}$ to U_j^p, and by $\Theta(j, p) = (\bar{\theta}_1, \ldots, \bar{\theta}_{j-L}) \in (\{0, 1\} \times \{0, 1\})^{j-L}$ a binary code. In the last notation, $\bar{\theta}_k$ represents filtering with $H_{\bar{\theta}_k(1)}(z_1) \cdot H_{\bar{\theta}_k(2)}(z_2)$ and then applying the down-sampler by 2 on each coordinate at step k of the iteration. Then $(d_j^p(x, n_1, n_2))_{(n_1, n_2) \in \mathbb{Z}^2}$ is obtained by passing $(d_L^0(x, n_1, n_2))_{(n_1, n_2) \in \mathbb{Z}^2}$ through the following discrete time filter

$$H_{\Theta(j, p)}(z_1, z_2) = \prod_{i=1}^{j-L} H_{\bar{\theta}_k(1)}(z_1^{2^{i-1}}) \cdot H_{\bar{\theta}_k(2)}(z_2^{2^{i-1}}), \quad (22)$$

and then applying the down-sampler by 2^{j-L}.

Remark 2: Note that the process that relates $(d_L^0(x, n_1, n_2))_{(n_1, n_2) \in \mathbb{Z}^2}$ with $(d_j^p(x, n_1, n_2))_{(n_1, n_2) \in \mathbb{Z}^2}$ in

---

[8]With transfer function $H_0(z)$ and $H_1(z)$, respectively.

Eq. (22) is linear but not time invariant. Therefore, it is inaccurate to talk about the frequency response associated with the process of projecting $(x(t_1, t_2))$ into the WP sub-space U_j^p. Pavez and Silva [55] addressed this issue by considering only the equivalent filtering part in Eq. (22). This consideration offers a characterization of the frequency content associated with each sub-space, from which we can define the frequency decomposition achieved by a given WP basis.

REFERENCES

[1] J. C. Felipe, A. J. M. Traina, and C. Traina, Jr., "Retrieval by content of medical images using texture for tissue identification," in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2003, pp. 175–180.

[2] F. Kayitakire, C. Hamel, and P. Defourny, "Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery," *Remote Sens. Environ.*, vol. 102, nos. 3–4, pp. 390–401, 2006.

[3] P. de Rivaz and N. Kingsbury, "Complex wavelet features for fast texture image retrieval," in *Proc. Int. Conf. Image Process.*, vol. 1, Oct. 1999, pp. 109–113.

[4] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[5] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 517–527, Dec. 2002.

[6] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1168–1178, Dec. 2005.

[7] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Rotation-invariant texture image retrieval using rotated complex wavelet filters," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 6, pp. 1273–1282, Dec. 2006.

[8] 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.

[9] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval via signature alignment based on steerable sub-Gaussian modeling," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1212–1225, Jul. 2008.

[10] S. Sakji-Nsibi and A. Benazza-Benyahia, "Copula-based statistical models for multicomponent image retrieval in thewavelet transform domain," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 253–256.

[11] R. Kwitt and A. Uhl, "A joint model of complex wavelet coefficients for texture retrieval," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1877–1880.

[12] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb. 2010.

[13] R. Kwitt, P. Meerwald, and A. Uhl, "Efficient texture image retrieval using copulas in a Bayesian framework," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2063–2077, Jul. 2011.

[14] Y. Dong and J. Ma, "Statistical contourlet subband characterization for texture image retrieval," in *Intelligent Computing Theories and Applications*. Berlin, Germany: Springer, 2012, pp. 495–502.

[15] M. S. Allili, N. Baaziz, and M. Mejri, "Texture modeling using contourlets and finite mixtures of generalized Gaussian distributions and applications," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 772–784, Apr. 2014.

[16] N.-E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2246–2261, May 2014.

[17] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, "Texture classification and retrieval using shearlets and linear regression," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 358–369, Mar. 2015.

[18] C. Li, G. Duan, and F. Zhong, "Rotation invariant texture retrieval considering the scale dependence of Gabor wavelet," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2344–2354, Aug. 2015.

[19] H. Rami, L. Belmerhnia, A. D. El Maliani, and M. El Hassouni, "Texture retrieval using mixtures of generalized Gaussian distribution and Cauchy–Schwarz divergence in wavelet domain," *Signal Process., Image Commun.*, vol. 42, pp. 45–58, Mar. 2016.

[20] Y. Wang, M. Shi, S. You, and C. Xu, "DCT inspired feature transform for image retrieval and reconstruction," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4406–4420, Sep. 2016.

[21] C. Li, Y. Huang, and L. Zhu, "Color texture image retrieval based on Gaussian copula models of Gabor wavelets," *Pattern Recognit.*, vol. 64, pp. 118–129, Apr. 2017.

[22] P. Liu, J.-M. Guo, C.-Y. Wu, and D. Cai, "Fusion of deep learning and compressed domain features for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 5706–5717, Dec. 2017.

[23] R. Fu, B. Li, Y. Gao, and P. Wang, "Content-based image retrieval based on CNN and SVM," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Oct. 2016, pp. 638–642.

[24] K. Lin, J. Lu, C. Chen, J. Zhou, and M. Sun, "Unsupervised deep learning of compact binary descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1501–1514, Jun. 2019.

[25] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. 22nd ACM Int. Conf. Multimedia*, New York, NY, USA, 2014, pp. 157–166. [Online]. Available: http://doi.acm.org/10.1145/2647868.2654948

[26] J. Y.-H. Ng, F. Yang, and L. S. Davis, "Exploiting local features from deep networks for image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jul. 2015, pp. 53–61.

[27] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 27–35.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[29] K. Lin, J. Lu, C.-S. Chen, and J. Zhou, "Learning compact binary descriptors with unsupervised deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1183–1192.

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

[31] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. New York, NY, USA: Academic, 2009.

[32] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 429–441, Oct. 1993.

[33] M. S. Allili, "Wavelet modeling using finite mixtures of generalized Gaussian distributions: Application to texture discrimination and retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1452–1464, Apr. 2012.

[34] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.

[35] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2545–2558, Jul. 2013.

[36] M.-T. Pham, G. Mercier, and L. Bombrun, "Color texture image retrieval based on local extrema features and Riemannian distance," *J. Imag.*, vol. 3, no. 4, p. 43, 2017.

[37] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[39] X. Dong, J. Shen, D. Wu, K. Guo, X. Jin, and F. Porikli, "Quadruplet network with one-shot learning for fast visual object tracking," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3516–3527, Jul. 2019.

[40] W. Wang and J. Shen, "Deep visual attention prediction," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2368–2378, May 2018.

[41] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[42] W. Wang, J. Shen, and H. Ling, "A deep network solution for attention and aesthetics aware photo cropping," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1531–1544, Jul. 2019.

[43] W. Wang, J. Shen, and L. Shao, "Video salient object detection via fully convolutional networks," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 38–49, Jan. 2018.

[44] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1265–1274.

[45] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 478–487.

[46] M.-T. Pham, "Efficient texture retrieval using multiscale local extrema descriptors and covariance embedding," in *Computer Vision—ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds. Cham, Switzerland: Springer, 2019, pp. 564–579.

[47] M. Alkhatib and A. Hafiane, "Robust adaptive median binary pattern for noisy texture classification and retrieval," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5407–5418, Nov. 2019.

[48] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

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

[50] L. Breiman, *Probability*. Reading, MA, USA: Addison-Wesley, 1968.

[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[52] C. Scott, "Tree pruning with subadditive penalties," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4518–4525, Dec. 2005.

[53] G. Lugosi and K. Zeger, "Concept learning using complexity regularization," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 48–54, Jan. 1996.

[54] C. Scott and R. D. Nowak, "Minimax-optimal classification with dyadic decision trees," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1335–1353, Apr. 2006.

[55] E. Pavez and J. F. Silva, "Analysis and design of wavelet-packet cepstral coefficients for automatic speech recognition," *Speech Commun.*, vol. 54, pp. 814–835, Jul. 2012.

[56] J. F. Silva and S. S. Narayanan, "On signal representations within the Bayes decision framework," *Pattern Recognit.*, vol. 45, no. 5, pp. 1853–1865, May 2012.

[57] J. Silva and S. S. Narayanan, "Discriminative wavelet packet filter bank selection for pattern recognition," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1796–1810, May 2009.

[58] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal pruning with applications to tree-structured source coding and modeling," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 299–315, Mar. 1989.

[59] R. Pickard, C. Graszyk, S. Mann, J. Wachman, L. Pickard, and L. Campbell, *Vistex Database*. Cambridge, MA, USA: MIT Media Lab, 1995.

[60] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, vol. 66. New York, NY, USA: Dover, 1966.

[61] R. Kwitt and P. Meerwald. *Salzburg Texture Image Database*. [Online]. Available: http://www.wavelab.at/sources/STex/

[62] G. J. Burghouts and J.-M. Geusebroek, "Material-specific adaptation of color invariant features," *Pattern Recognit. Lett.*, vol. 30, no. 3, pp. 306–313, 2009.

[63] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 933–936.

[64] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture discrimination using multimodal wavelet packet subbands," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 3, Oct. 2004, pp. 1493–1496.

[65] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture analysis using probabilistic models of the unimodal and multimodal statistics of adaptive wavelet packet coefficients," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, May 2004, pp. III-597–III-600.

[66] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, B. Ruskai, Ed. Burlington, MA, USA: Jones & Bartlett Learning, 1992, pp. 153–178.

[67] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

ANDREA VIDAL (S'17) received the B.S. and M.S. degrees (Hons.) in electrical engineering from the University of Chile, Santiago, Chile, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department, The University of Texas at Dallas. Her current research interests include entrainment, emotion recognition, lip synthesis, and facial animation.

JORGE F. SILVA received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), in 2005 and 2008, respectively. He is currently an Associate Professor with the Electrical Engineering Department and the Director of the Information and Decision Systems (IDS) Group, University of Chile, Santiago, Chile. His current research interests include statistical signal processing, machine learning and information processing, universal source coding, and compressible models for inference and decision. He is an IEEE member of the Signal Processing and Information Theory Societies. He received the Viterbi Doctoral Fellowship, from 2007 to 2008, and Simon Ramo Scholarship, from 2007 to 2008, at USC. He has participated as a reviewer in various IEEE journals on signal processing. He was an Associate Editor of the IEEE Transactions on Signal Processing, from 2015 to 2017.

CARLOS BUSSO (S'02–M'09–SM'13) received the B.S. and M.S. degrees (Hons.) in electrical engineering from the University of Chile, Santiago, Chile, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 2008. He leads the Multimodal Signal Processing (MSP) Laboratory, UTD. He is currently an Associate Professor with the Electrical Engineering Department, The University of Texas at Dallas (UTD). His work has direct implication in many practical domains, including national security, health care, entertainment, transportation systems, and education. He was selected by the School of Engineering of Chile as the best Electrical Engineer graduated, in 2003, across Chilean universities. He is the coauthor of the winner paper of the Classifier Sub-Challenge event at the Interspeech 2009 emotion challenge. His research interest is in human-centered multimodal machine intelligence and applications. His current research interests include the broad areas of affective computing, multimodal human-machine interfaces, nonverbal behaviors for conversational agents, in-vehicle active safety systems, and machine learning methods for multimodal processing. He is a member of ISCA, AAAC, and ACM. At USC, he received a provost doctoral fellowship, from 2003 to 2005, and a fellowship in Digital Scholarship, from 2007 to 2008. He was a recipient of an NSF CAREER Award. In 2014, he received the ICMI Ten-Year Technical Impact Award. In 2015, his student received the third prize IEEE ITSS Best Dissertation Award (N. Li). He also received the Hewlett Packard Best Paper Award at the IEEE ICME 2011 (with J. Jain), and the Best Paper Award at the AAAC ACII 2017 (with Yannakakis and Cowie). He was the General Chair of ACII 2017.

• • •