UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

WAVELET PACKET FILTER BANK SELECTION FOR TEXTURE RETRIEVAL

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA

ANDREA VIDAL SALAZAR

PROFESOR GUÍA:
JORGE SILVA SÁNCHEZ

MIEMBROS DE LA COMISIÓN:
MARTIN ADAMS
CARLOS BUSSO RECABARREN

SANTIAGO DE CHILE
2017

RESUMEN DE TESIS PARA OPTAR AL
TÍTULO DE MAGISTER EN CIENCIAS
DE LA INGENIERÍA, MENCIÓN ELÉCTRICA
RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE INGENIERA CIVIL ELÉCTRICA
POR: ANDREA VIDAL SALAZAR
FECHA: DICIEMBRE 2017
PROFESOR GUÍA: JORGE SILVA SÁNCHEZ

WAVELET PACKET FILTER BANK SELECTION FOR TEXTURE RETRIEVAL

Durante los últimos años, el avance de la tecnología de captura y almacenamiento ha generado un volumen sin precedentes de imágenes digitales almacenadas en las bases de datos. Esto plantea el desafío de desarrollar sistemas autónomos que sean eficientes en la búsqueda y organización del contenido digital. Como problema emblemático surgió Content-Based Image Retrieval como área de investigación. Un sistema de indexación de imágenes busca encontrar las imágenes más similares a una en particular y está compuesto de dos etapas: extracción de características y medición de similitud. La primera etapa busca la forma de representar la imagen extrayendo las características más discriminativas, mientras que la segunda etapa es usada para ordenar las imágenes de acuerdo a su similitud.
Esta tesis propone el uso de Wavelet Packet para abordar el problema de indexación de imágenes de texturas. Wavelet Packet es una herramienta del procesamiento de señales que no ha sido usada en el estado del arte para enfrentar el problema de indexación y, además, es capaz del proveer distintas representaciones para una imagen. Para seleccionar la mejor representación de Wavelet Packet, este trabajo propone una nueva metodología para el problema de indexación que aborda el problema de selección de bases para la familia de Wavelet Packets utilizando el criterio de Mínima Probabilidad de Error. Como resultado de la implementación de la metodología propuesta, se muestra que las soluciones provistas por Wavelet Packet son adaptivas y mejoran el desempeño del sistema de indexación con respecto a la solución Wavelet, bajo condiciones similares y modelos estadísticos.

Summary

Over the last years, the advance on technology has produced an unprecedented amount of digital content stored in databases. This situation motivates the development of systems that can organize and search digital content automatically in real time. As a solution of this challenge, the area of Content-Based Image Retrieval has emerged as a very relevant topic on image processing. In a nutshell, an image retrieval system attempts to find the most similar images to a query image. It is composed of two main stages: feature extraction and similarity measurement. The first stage aims to find a way to represent the image extracting discriminant characteristics from it, while the second stage ranks images in terms of their similarity.

This thesis proposes the use of Wavelet Packets to address the texture image retrieval problem. Wavelet Packet is a signal processing tool that has not been used on the literature to deal with the retrieval problem and can provide multiple representations for an image. A novel methodology is proposed to select the best Wavelet Packet representation using the Minimum Probability of Error criterion. As a result of the implementation of the proposed methodology, it is shown that the adaptive nature of the proposed Wavelet Packets solution improves the retrieval performance with respect to a Wavelet solution used under the same conditions and model assumptions.

A mi Goli.
A mi madre y a mi padre.

Acknowledgements

Nueve años duró esta travesía por la universidad que está llegando a su fin. Durante esos años la universidad se convirtió en mi segundo hogar, hogar que fue comformado por personas que hicieron de mi paso por la universidad unas de las mejores experiencias de mi vida. Me encantaría poder agradecer a todo el mundo que conocí, sin embargo en virtud del espacio que poseo, espero que no se sientan si no son nombrados y de verdad decirles gracias por todo.

En una primera instancia, me gustaría agradecer a los chicos del Laboratorio IDS. A quienes lo conforman hoy en día y quienes ya terminaron su paso por la universidad. No olvidaré esas interminables conversaciones durante y después almuerzo (algunas serias y otras jugosas), gracias por la buena onda y aguantar mis canciones y spams. Quisiera hacer mención especial a "Los Coquetos", Sebita y Felipe, por hacer los días más llevaderos tirando tallas, molestándonos mutuamente y darme su confianza. También agradecer al lado Colombiano del laboratorio por adaptarse, entrar en confianza con nosotros, y no menos importante compartir su cultura.

No puedo dejar de agradecer a mi profesor guía, Jorge Silva, por todo lo que me ha enseñado a través de estos años, por mostrarme el maravilloso mundo de la investigación, y por sus conversaciones motivacionales que siempre llegaban en el momento justo. No obstante, el desarrollo de la tesis no sólo involucra la parte académica, es por ello que no puedo dejar de agradecer a Milena por recibirme siempre en su oficina, por su buena onda, conversaciones varias, sus ayudas y humor.

Finalmente pero no menos importante, quisiera agradecer a mi familia que siempre estuvo presente en los buenos y malos momentos de esta travesía. Madre gracias por apoyarme. Padre gracias por hacerme fuerte. Hermanos gracias por quererme como soy y aguantarme. También quisiera hacer una mención especial para mi Goli, que en donde quiera que estés, gracias por todo, en especial por lo que me has enseñado que son cosas tan simples pero valiosas como lo es siempre sonreír ante la vida que nos toca vivir y valorar lo que tenemos.

Se agradece a Advanced Center for Electrical and Electronic Engineering (AC3E) y al Departamento de Postgrado y Postítulo de la Vicerrectoría de Asuntos Académicos de la Universidad de Chile por su apoyo económico para el desarrollo de esta tesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Nowadays, retrieving information is a challenging problem because of increasing amounts of data produced by technology. In particular, online data storage and digital cameras have compelled search engines to develop new and efficient techniques to search images. Thus, Content-Based Image Retrieval (CBIR) emerged as a research area. CBIR's purpose is to recover the most similar images (indexing task) from a database as compared to a particular image of interest. An important problem within CBIR framework is texture image retrieval, which is relevant in certain areas such as medical [1] and satellite [2] imaging. In the first area, texture image retrieval is important to decrease the number of required procedures to diagnose illnesses, while the second area, texture image retrieval is used to perform forest inventory analysis. CBIR is composed of two main stages: Feature Extraction (FE) and Similarity Measurement (SM). The first is used to extract the image features, while the second measures the similarity among images using the extracted features.

One of the main studies in the CBIR area is the work of Do *et al.* [3] in which the authors proposed a statistical framework, which entails a comparison of statistical models to improve retrieval performance. Here, the retrieval problem is formulated as Multiple Hypothesis Testing (MHT) providing a joint FE and SM scheme. They considered a texture image as a realization of a statistical image model on the Wavelet domain where consequently the task of indexing reduces to a MHT problem. Both the FE and the SM are obtained in closed form, implementing the optimal decision rule based on the minimum probability of error criterion.

Since the work of Do *et al.* [3], CBIR experienced a breakthrough. From that point onwards, the use of the statistical retrieval framework has grown significantly because it reduced the computational complexity and the dimensionality of the image retrieval problem. Within this framework, different techniques have been applied to enhance the retrieval performance. Researchers have usually made improvements in the FE stage in which different signal processing (SP) tools have been used such as dual-tree complex wavelet transform (DT-CWT) [4, 5, 6, 7], steerable filters [8, 9, 10], and gabor wavelets [11, 12]. These SP tools capture details with different texture orientations. Another SP tool used in image retrieval problem is the Discrete Cosine Transform [13] that chooses the most informative orientations of details. Recently, new transforms such as contourlets [14] and shearlets [15] have been applied to improve the retrieval performance [16, 17, 18]. Contourlets describe

the contour of an object, while shearlets explore a wider variety of orientations of details than other transforms. However, the changes on the FE stage not only comprise changes in the feature represention, but also in the improvement on the characterization of the texture models [19, 20, 21], for instance, modeling the inter-band dependencies [22, 7, 23, 24, 12]. The changes in the FE stage has also produced changes in the SM stage, because the goal is to find closed expression to make more efficient the similarity calculation of image retrieval [4, 5, 10, 25, 16, 26, 21, 27].

Based on the work of Do *et al.* [3], this work introduces the use of Wavelet Packet (WP) in a texture retrieval problem. Some authors have studied WP qualities and effectiveness in analysis, classification, and segmentation of textures [28, 29, 30, 31, 32, 33, 34]; however, WP has not been explored in a retrieval context. Wavelet Packets (WPs) are a collection of bases that offer many ways of decomposing the image spaces in terms of space-scale information [35, 36]. This collection is induced by a two-channel filter (TCF) and organized in terms of an embedded tree-structure. One emblematic case of this family is the Wavelet basis that is constructed by iterating the TCF in the high scale band (or low frequency information) producing a multi-resolution partition of the image space. However, by iterating the TCF in different bands [36], a very rich collection of sub-space decompositions of the image space can be constructed, which this thesis hypothesizes that it can benefit image texture retrieval tasks.

1.1 Hypothesis

The rich collection of space-scale partitions of the image space can be used to find descriptions that offer better texture discrimination and, consequently, better performance for the indexing task. Furthermore, the adoption of a collection of filter bank representations allows the possibility to formulate a learning criterion to adapt the problem to the nature of the task in terms of discrimination and learning complexity.

1.2 Structure of the thesis

This thesis is organized in 8 chapters. In Chapter 2 presents some WP, which are essential to understand the development of this work. In Chapter 3, the statistical framework proposed in [3] is extended to include a more general family of transform based representations of the image space, where the indexing task is reduced to the problem of joint parameter estimation and minimum KL divergence similarity. This extension is used in Chapter 4 to provide a novel framework based on WP basis selection. Furthermore, the basis selection approach presented in Chapter 5 can be formulated as a trade-off between feature discrimination and estimation errors. This trade-off appears since the indexing phase implies a parameter estimation problem with a finite number of samples. Using the tree-structure of the WP family, the WP selection is reduced to an equivalent minimum cost-tree pruning problem (Chapter 5.1). Remarkably this problem has natural connection with classification and regression trees

(CART) algorithms [37], which offers a computational efficient solution. Chapter 6 presents a systematic experimental analysis of the proposed WP solutions on synthetic data as well as on six real datasets with different images sizes and texture classes. The achieved results are very promising, demonstrating the ability of the proposed framework to adapt to the task, providing more discriminative features. It is shown that WP representations offer important improvements when it is compared with the standard Wavelet solutions in several scenarios, where the obtained WP solutions explore new sub-spaces and scales that help to enhance the discrimination among images. Finally, Chapter 7 presents the conclusions and future work.

Chapter 2

Wavelet Packet Analysis

Wavelet Packets (WPs) offer different space-scale representations and the potential to adapt to complex phenomena such as natural images and acoustic processes [38, 39]. This section provides a brief presentation of WP with focus on its filter-bank characteristics, sub-band frequency decomposition and indexed structure (tree-structure)[1].

2.1 WP Bases: A Tree-Structured Collection

Let $\mathbb{X}$ be a raw image space (pixel domain) that, without loss of generality, is assumed to have a finite *level of scale* 2^L or *resolution* 2^{-L}, L being an integer strictly greater than zero [36]. 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 [36, 35, 40], where ψ_L is the mother wavelet function. For the 2D case, the WP framework provides a way of decomposing $\mathbf{B}_L$ into four orthonormal collections denoted by $\mathbf{B}^p_{L+1} \equiv \left\{(\psi^p_{L+1}(t_1 - 2^{L+1} n_1, t_2 - 2^{L+1} n_2))_{(t_1,t_2)\in\mathbb{R}^2} : (n_1, n_2) \in \mathbb{Z}^2\right\}$ for $p \in \{0, 1, 2, 3\}$. Then, denoting by $U^p_{L+1} \equiv span\left\{(\psi^p_{L+1}(t_1 - 2^{L+1} n_1, t_2 - 2^{L+1} n_2)) : (n_1, n_2) \in \mathbb{Z}^2\right\}$ with $p \in \{0, 1, 2, 3\}$, it has the following sub-space decomposition [36]

$$\mathbb{X} = U^0_{L+1} \oplus U^1_{L+1} \oplus U^2_{L+1} \oplus U^3_{L+1}. \tag{2.1}$$

These 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, see details on [36, Chap. 3]. The tree-structure of the WP bases comes from the fact that $\left\{\mathbf{B}^p_{L+1} : p \in \{0, 1, 2, 3\}\right\}$ are induced by a (discrete) pair of *conjugate mirror filters* (CMF) that are denotes by $(h_0(n), h_1(n))_{n\in\mathbb{N}}$ [36, Chap. 7.1.3]. More precisely, the basis elements associated with the scale $L + 1$ are induced from $\psi_L(t_1, t_2)$ in the following

[1] Comprehensive expositions can be found in [36, 35, 40].

way:

$$\begin{aligned}
\psi^0_{L+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}} h_0(n_1) \sum_{n_2\in\mathbb{Z}} h_0(n_2)\cdot\psi_L(t_1-2^L n_1, t_2-2^L n_2),\\
\psi^1_{L+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}} h_1(n_1) \sum_{n_2\in\mathbb{Z}} h_0(n_2)\cdot\psi_L(t_1-2^L n_1, t_2-2^L n_2),\\
\psi^2_{L+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}} h_0(n_1) \sum_{n_2\in\mathbb{Z}} h_1(n_1)\cdot\psi_L(t_1-2^L n_1, t_2-2^L n_2),\\
\psi^3_{L+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}} h_1(n_1) \sum_{n_2\in\mathbb{Z}} h_1(n_2)\cdot\psi_L(t_1-2^L n_1, t_2-2^L n_2),
\end{aligned} \tag{2.2}$$

where $h_0(n)$ and $h_1(n)$ are related by the *perfect reconstruction property*[2].

By iterating this filter bank approach on each basis element in $\{(\psi^p_{L+1}(t_1,t_2)), p\in\{0,1,2,3\}\}$ [36, Th. 8.1], it can continue, in a tree-structured way, with the construction of bases and sub-space decompositions for $\mathbb{X}$. For example, after j iterations, it creates $\psi^p_{L+j}(t_1,t_2)$ for any $p\in\{0,\ldots,4^j-1\}$ where $U^p_{L+j} = span\left\{(\psi^p_{L+j}(t_1-2^{L+j}n_1, t_2-2^{L+j}n_2)) : (n_1 n_2)\in\mathbb{Z}^2\right\}$, see Figure 2.1. Furthermore by construction, $\forall j\geq 1$, $\forall p\in\{0,\ldots,4^j-1\}$,

$$U^p_{L+j} = U^{4p}_{L+j+1}\oplus U^{4p+1}_{L+j+1}\oplus U^{4p+2}_{L+j+1}\oplus U^{4p+3}_{L+j+1}, \tag{2.3}$$

where

$$\begin{aligned}
\psi^{4p}_{L+j+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}}\sum_{n_2\in\mathbb{Z}} h_0(n_1)h_0(n_2)\cdot\psi^p_{L+j}(t_1-2^{L+j}n_1, t_2-2^{L+j}n_2),\\
\psi^{4p+1}_{L+j+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}}\sum_{n_2\in\mathbb{Z}} h_1(n_1)h_0(n_2)\cdot\psi^p_{L+j}(t_1-2^{L+j}n_1, t_2-2^{L+j}n_2),\\
\psi^{4p+2}_{L+j+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}}\sum_{n_2\in\mathbb{Z}} h_0(n_1)h_1(n_1)\cdot\psi^p_{L+j}(t_1-2^{L+j}n_1, t_2-2^{L+j}n_2),\\
\psi^{4p+3}_{L+j+1}(t_1,t_2) &= \sum_{n_1\in\mathbb{Z}}\sum_{n_2\in\mathbb{Z}} h_1(n_1)h_1(n_2)\cdot\psi^p_{L+j}(t_1-2^{L+j}n_1, t_2-2^{L+j}n_2).
\end{aligned} \tag{2.4}$$

As a result, the WP creates a family of tree-structured bases induced by the iteration of the *two-channel filter* (TCF) as illustrated in Figure 2.2.

A key property of WP is the inter-scale relationship induced in Eq. (2.4) among the WP transform coefficients obtained across scales [36]. More precisely, the transform coefficients of $(x(t_1,t_2))$ belong to $U^p_j\subset\mathbb{X}$ are given by

$$\mathrm{d}^p_j(x,n_1,n_2)\equiv\langle x(t_1,t_2),\psi^p_j(t_1-2^j n_1, t_2-2^j n_2)\rangle,\ \forall(n_1,n_2)\in\mathbb{Z}^2. \tag{2.5}$$

This decomposition allows an alternative analysis of $(x(t_1,t_2))$, which is based on the bases associated with U^{4p}_{j+1}, U^{4p+1}_{j+1}, U^{4p+2}_{j+1} and U^{4p+3}_{j+1} using Eq. (2.3):

$$\begin{aligned}
\mathrm{d}^{4p+b}_{j+1}(x,n_1,n_2) &= \langle x(t_1,t_2),\psi^{4p+b}_{j+1}(t_1-2^{j+1}n_1, t_2-2^{j+1}n_2)\rangle,\\
&= \sum_{k_1\in\mathbb{Z}}\sum_{k_2\in\mathbb{Z}} h_{f_1(b)}(k_1-2n_1)h_{f_2(b)}(k_2-2n_2)\cdot\mathrm{d}^p_j(x,k_1,k_2),
\end{aligned} \tag{2.6}$$

[2] $h_1(n) = (-1)^{1-n}h_0(1-n)$, $\forall n\in\mathbb{Z}$ [41], [36, Th. 8.1].

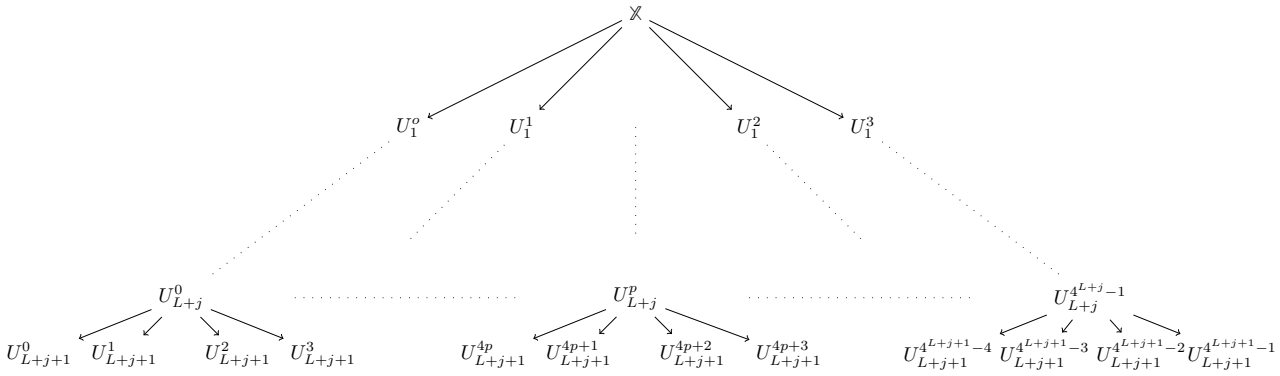


Figure 2.1: Wavelet Packet Tree-structured Sub-space decomposition.

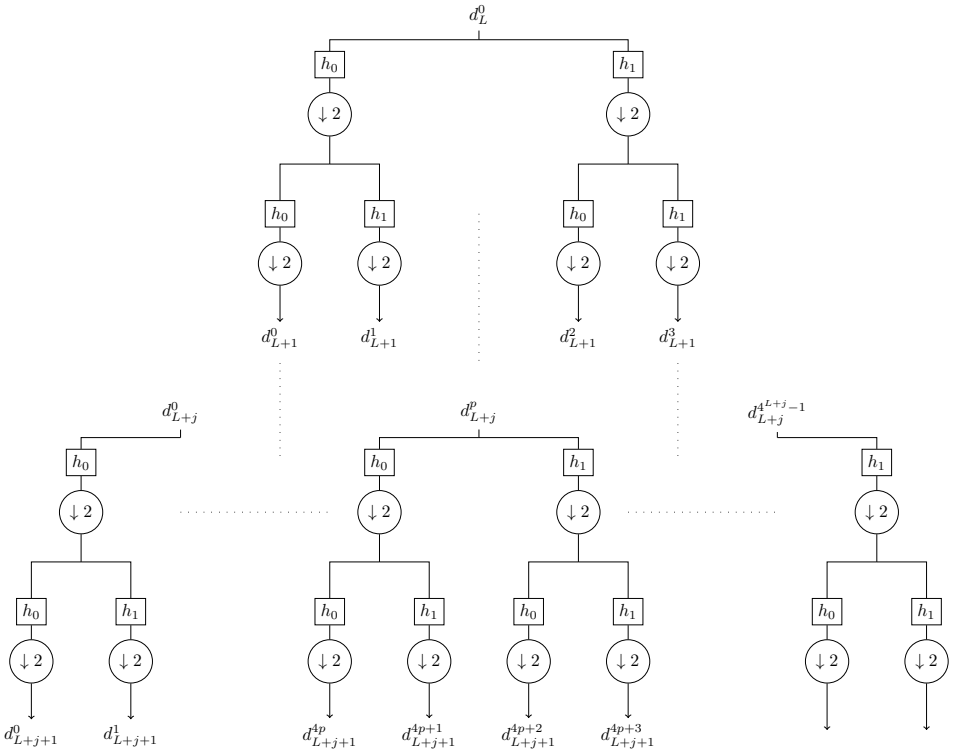


Figure 2.2: Wavelet Packet filter-bank implementation.

for all $b \in \{0,1,2,3\}$, where $f_1(\cdot)$ and $f_2(\cdot)$ are discrete functions given by [36, Prop. 8.4]:

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

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

Therefore a closed-form relationship in the transform coefficients can be obtained for every pair of basis elements in the WP, as illustrated in Figure 2.2. A key aspect of this property is that the analysis in continuous time in Eq. (2.5) is moved to an analysis (algorithm) in discrete time in Eq. (2.6). Furthermore, the WP quaternary structure presented in Eq. (2.6) permits a fast algorithm implementation of the WP analysis [36, Chap. 7.3]. Because of

the aforementioned inter-scale relationship of the WP coefficients, which allows a recursive splitting of the TCFs creating the quad-tree structure (Figure 2.1).

2.2 Filter-Bank Implementation and the WP Sub-Space Frequency Decomposition

The analysis step in Eq. (2.6) can be implemented by a two-channel filter (TCF), followed by a down-sampling operation by a factor of 2 (non-linear)[35, 36]. This basic step is extended in the following result:

Proposition 2.1 *[42, Chap. 11.3.3] Let $x(t_1, t_2)$ be in a finite 2^L scale space $\mathbb{X}$, with transform coefficients $(\mathrm{d}_L^0(x, n_1, n_2))_{(n_1,n_2)\in\mathbb{Z}^2}$ obtained from Eq. (2.5). Considering an arbitrary sub-space U_j^p induced from the WP filter bank decomposition with $j > L$ and $p \in \{0, \ldots, 4^{j-L} - 1\}$. Denoting by $(h_0(n))_{n\in\mathbb{Z}}$ and $(h_1(n))_{n\in\mathbb{Z}}$, the conjugate mirror filter pair (with transfer function $H_0(z)$ and $H_1(z)$), by $U_{L+1}^{p_1}, \ldots, U_j^{p_{j-L-1}}$ the sequence of intermediate sub-spaces used to go from $\mathbb{X}$ to U_j^p, and by $\Theta(j, p) = (\bar{\theta}_1, \ldots, \bar{\theta}_{j-L}) \in (\{0, 1\} \times \{0, 1\})^{j-L}$ the binary path code. In the last definition, choosing $\bar{\theta}_k$ implies 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 $(\mathrm{d}_j^p(x, n_1, n_2))_{(n_1,n_2)\in\mathbb{Z}^2}$ is obtained by passing $(\mathrm{d}_L^0(x, n_1, n_2))_{(n_1,n_2)\in\mathbb{Z}^2}$ through the following discrete time filter*

$$H_{\Theta(j,p)}(z_1, z_2) = \prod_{\mathrm{i}=1}^{j-L} H_{\bar{\theta}_k(1)}(z_1^{2^{\mathrm{i}-1}}) \cdot H_{\bar{\theta}_k(2)}(z_2^{2^{\mathrm{i}-1}}), \tag{2.7}$$

and then applying the down-sampler by 2^{j-L} operator on each coordinate.

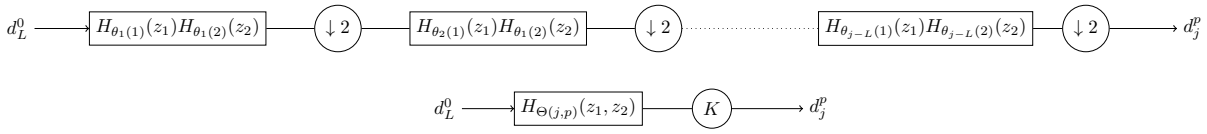
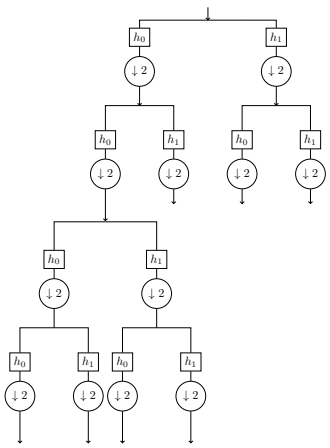


Figure 2.3: The system view process to determine the transform coefficients of the sub-space U_j^p, presented in Proposition 2.1. The aggregated down-sampler is by $K = 2^{j-L}$.

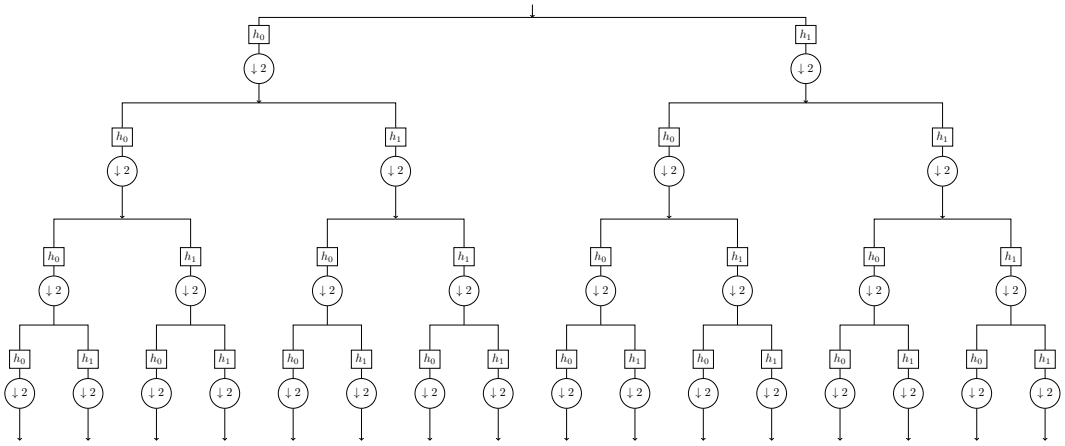
Remark 1 Note that the process that relates the coefficients $(\mathrm{d}_L^0(x, n_1, n_2))_{(n_1,n_2)\in\mathbb{Z}^2}$ with $(\mathrm{d}_j^p(x, n_1, n_2))_{(n_1,n_2)\in\mathbb{Z}^2}$ in Eq. (2.7) is linear but not time invariant. Therefore, it is inaccurate to talk about the frequency response associated with the process of projecting $(x(t_1, t_2))$ into the WP sub-space U_j^p. Pavez *et al.* [39] addressed this issue by considering only the equivalent filtering part in Eq. (2.7) and avoiding the last down-sampling stage. This consideration offers a characterization of the frequency content associated with each sub-space (or sub-band), from which can be defined the frequency decomposition achieved by a given WP basis. For illustration, this frequency partition is shown for the ideal *Shannon two-channel*

filter [36][3]. A basis is indexed by $\{(j_{\rm i}, p_{\rm i}) : {\rm i} = 1, \ldots, M\}$ associated with the basis element $B = \bigcup_{{\rm i}=1}^{M} \mathbf{B}_{j_{\rm i}}^{p_{\rm i}}$ and sub-space decomposition $\mathbb{X} = \bigoplus_{{\rm i}=1}^{M} U_{j_{\rm i}}^{p_{\rm i}}$. For each $(j_{\rm i}, p_{\rm i})$, its equivalent filters $H^{\rm i}(z) \equiv H_{\Theta(j_{\rm i},p_{\rm i})}(z_1, z_2)$ can be obtained by Eq. (2.7) and, consequently, reduces the analysis to the frequency response of an M-channel filter-bank. Examples of the frequency response before the down-sampling stage are presented in Figures 2.4 and 2.5. From these examples, it can notice that the Wavelet type of structure, which is produced by iterating $H_{0,0}(\mathrm{e}^{jw_1}, \mathrm{e}^{jw_2})$ in every step, it obtains a solution that increases the resolution in the low frequency range. In general, the frequency support of the resulting sub-space is reduced by half on each dimension in each step of iterating the TCF as illustrated in Figure 2.4c.

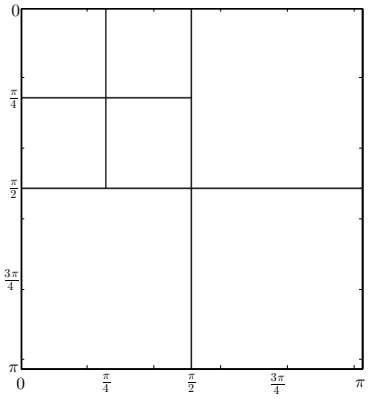
[3]The frequency partition generated by filter bank iterations is not ordered in frequency [36, Chap. 8] and for this reason the Gray Code is used to find the correct frequency partition. An example of this is illustrated in Figure 2.5, where the ordered frequency partitions generated by the filter bank iterations in Figures 2.5a and 2.5b are given by Figures 2.5c and 2.5d, respectively. Details are presented in Chapter 8.1 for completness.



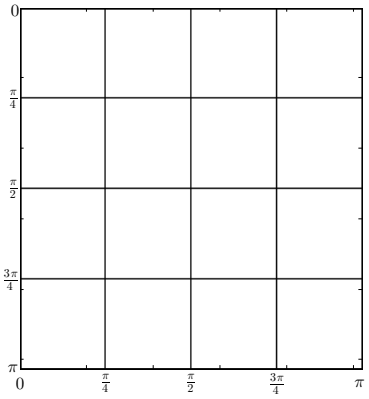
(a) Wavelet case



(b) Balanced tree case

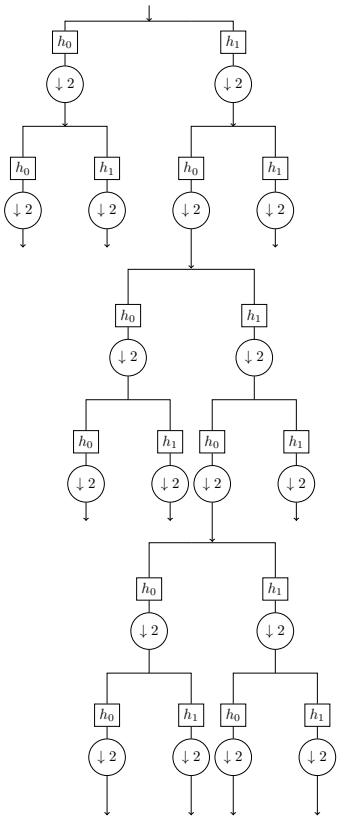


(c) Frequency response of equivalent filterbank of figure 2.4a

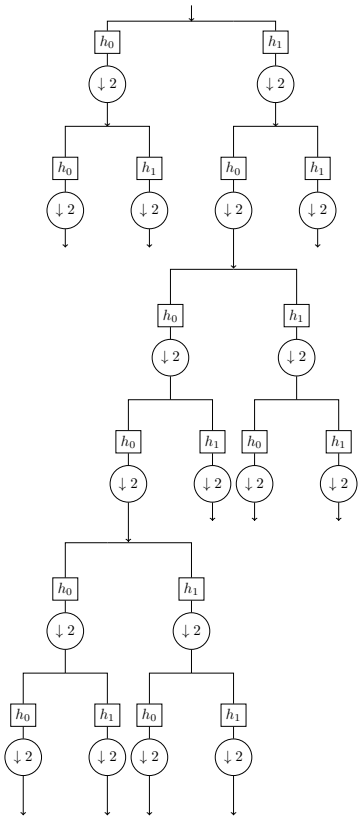


(d) Frequency response of equivalent filterbank of figure 2.4b

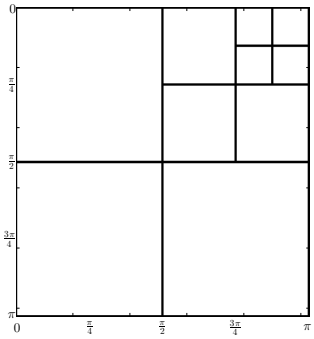
Figure 2.4: Illustration of the frequency division of Wavelet Packet bases for two tree structures. The ideal *Shannon conjugate filter pair* is considered, which provides perfect dyadic partitions of the interval $[-\pi, \pi] \times [-\pi, \pi]$. Scenario (2.4a-2.4c) shows a iteration of $H_0(z_1)H_0(z_2)$ (Wavelet type), and scenario (2.4b-2.4d) presents a balanced tree structure (uniform frequency resolution).



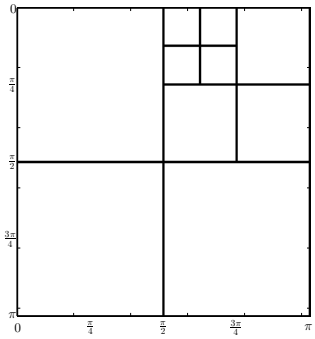
(a) Center-right iterated channel



(b) Left iterated channel



(c) Frequency response of equivalent filterbank of figure 2.5a



(d) Frequency response of equivalent filterbank of figure 2.5b

Figure 2.5: The same scenario as in Figure 2.4. Scenario (2.5a-2.5c) shows a double iteration of $H_0(z_1)H_1(z_2)$. Meanwhile, scenario (2.5b-2.5d) is a iteration of $H_0(z_1)H_1(z_2)$ and $H_0(z_1)H_0(z_2)$ filters.

2.3 Tree Indexing

To formulate the problem of WP basis selection, the rooted tree notation used in Silva *et al.* [38] is adopted. The tree-structure creates a particular WP basis by iterating the four-channel filter bank as illustrated in Figure 2.2. Let J denotes the maximum number of iterations of the sub-band decomposition and let $G = (V, E)$ be a graph with

$$V = \{(0,0), (1,0), (1,1), (1,2), (1,3), \ldots, (J,0), \ldots, (J, 4^J - 1)\}$$

and E the collection of arcs on $V \times V$ that characterizes the full balanced rooted tree with root $v_{root} = (0,0)$ in Figure 2.1. Instead of representing a tree as a collection of arcs in G, this work uses the convention used by Breiman *et al.* in CART [37] 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. $\mathcal{L}(\mathcal{T})$ is defined as the set of leaves of $\mathcal{T}$ and $\mathcal{I}(\mathcal{T})$ as the set of internal nodes, where consequently $\mathcal{T} = \mathcal{L}(\mathcal{T}) \cup \mathcal{I}(\mathcal{T})$. 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 $v \in T$, the largest branch of $\mathcal{T}$ rooted at v is denoted 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, $\mathcal{T}_{full} \equiv V$ denote the full quad-tree, and consequently, the collection of WP bases will be indexed by the collection $\{\mathcal{T} \subset V : \mathcal{T} \ll \mathcal{T}_{full}\}$.

Any pruned version of $\mathcal{T}_{full}$ in Figure 2.2 represents a particular WP basis by the iteration of the TCF. 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:

$$\mathbf{U}_{\mathcal{T}} = \left\{U_{j_k}^{p_k} : k = 1, \ldots, M\right\}, \tag{2.8}$$

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,

$$\mathbf{B}_{\mathcal{T}} = \cup_{k=1}^{M} \mathbf{B}_{j_k}^{p_k}, \tag{2.9}$$

Finally for any $x \in \mathbb{X}$, the transform coefficients with respect to $\mathbf{B}_{\mathcal{T}}$ can be determined by Eq. (2.5). 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} transform coefficients that are obtained by the discrete time signal processing equation in (2.6). For simplicity, the index $(n_1, n_2) \in \left\{1, \ldots, 2^{J-j_k}\right\}^2$ is mapped into $n \in \left\{1, \ldots, 4^{J-j_k}\right\}$ to represent the transform coefficients as a 1D vector. Thus, the transform coefficients of x in $U_{j_k}^{p_k}$ are denoted by $D_{j_k}^{p_k}(x) = \left(\mathrm{d}_{j_k}^{p_k}(x, n)\right)_{n=1,\ldots,4^{J-j_k}}$ and the transform coefficients for the basis $\mathbf{B}_{\mathcal{T}}$ are given and represented by

$$D_{\mathcal{T}}(x) = (D_{j_k}^{p_k}(x))_{k=1,\ldots,M}. \tag{2.10}$$

Note that the number of transform coefficients of the node $(j, p) \in V$ scales like 4^{J-j}.

[4]Without loss of generality the pixel based representation of x corresponding to the transform coefficients of the trivial WP basis $\{v_{root}\} \ll \mathcal{T}_{full}$, i.e., $(x(n)) = (\mathrm{d}_0^0(n) : n = 1, \ldots, 4^J)$ is the representation at the smaller scale.

Chapter 3

The Multiple Hypothesis Problem for Image Retrieval

Following the approach proposed by Do *et al.* [3], the indexing problem is formulated as a multiple hypothesis testing (MHT) problem. A query image $\mathbf{x}$ is represented by a set of observations in a finite dimensional space $\mathbb{X} = \mathbb{R}^L$. Each candidate image in the database is presented by a statistical model (or a probability measure in $\mathbb{X}$) that it denotes by $\{\mu_{\theta_\mathrm{i}} : \mathrm{i} = 1, \ldots, M\}$, where $\mu_{\theta_\mathrm{i}} \in \mathcal{P}(\mathbb{X})$. Then, the (content-based) indexing of $\mathbf{x}$ reduces to find the closest N candidates in $\{\mu_{\theta_\mathrm{i}} : \mathrm{i} = 1, \ldots, M\}$ by applying the *maximum likelihood (ML) principle*, i.e., to find the N models $\left\{\mu_{\theta_{k_\mathrm{i}}} : \mathrm{i} = 1, \ldots, N\right\}$ with the highest likelihood of observing $\mathbf{x}$, i.e.,

$$k_\mathrm{i} = \arg \max_{k \in [M] \setminus \{k_1, \ldots, k_{\mathrm{i}-1}\}} \mu_{\theta_k}(\{\mathbf{x}\}), \ \forall \mathrm{i} = 1, \ldots, N, \tag{3.1}$$

where $[M] \equiv \{1, \ldots, M\}$. The ML criterion in Eq. (3.1) is well-known to be the optimal from the probability of error viewpoint [43]. However, in this context, it is computationally expensive and intractable in general. For that reason, Do *et al.* [3, Section II] proposed the adoption of an efficient two-stage principle that asymptotically (in the size of the image) approximates Eq. (3.1). This chapter elaborates a slightly more general version of this approach to apply it to the case of WP texture analysis.

First, this work focuses on a parametric context, i.e., μ_{θ_i} is equipped with a density function $f_{\theta_\mathrm{i}}(\mathbf{x})$[1] that is fully characterized by a vector of parameters $\theta_\mathrm{i} \in \Theta$. Then, in order to evaluate the probability, the likelihood $\log f_\theta(\mathbf{x})$ for each $\theta \in \bar{\Theta} = \{\theta_\mathrm{i} : \mathrm{i} = 1, \ldots, M\}$ is used to solve Eq. (3.1). Furthermore, it is considered that $\mathbb{X} = \bigotimes_{\mathrm{i}=1}^{T} \mathbb{X}_\mathrm{i}$, i.e., $\mathbf{x} = (\mathbf{x_1}, \ldots, \mathbf{x_T})$ where $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,L_t}) \in \mathbb{X}_t$ is a finite dimensional vector of dimension $L_t \geq 1$ and, consequently, $L = \sum_{t=1}^{T} L_t$. Consistent with this cartesian product splitting of $\mathbb{X}$, it is assumed that each f_{θ_i} with $\theta_\mathrm{i} \in \bar{\Theta}$ decomposes in T independent (product) components indexed by $\theta_\mathrm{i} = (\theta_\mathrm{i}^1, \ldots, \theta_\mathrm{i}^T)$, where

$$\log f_{\theta_\mathrm{i}}(\mathbf{x}) = \sum_{t=1}^{T} \log \prod_{j=1}^{L_t} \left(f_{\theta_\mathrm{i}^t}(\mathbf{x}_{t,j}) \right) = \sum_{t=1}^{T} \left(\sum_{j=1}^{L_t} \log f_{\theta_\mathrm{i}^t}(x_{t,j}) \right). \tag{3.2}$$

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

Therefore, $f_{\theta_{\mathrm{i}}}$ in $\mathbb{X}$ is fully determined by a set of marginals pdfs $\left\{f_{\theta_{\mathrm{i}}^1}, \ldots, f_{\theta_{\mathrm{i}}^T}\right\}$ in $\mathbb{R}$. More specifically, Eq. (3.2) means that under the hypothesis that $\mathbf{x_1}, \ldots, \mathbf{x_T}$ follows $\mu_{\theta_{\mathrm{i}}}$, $\mathbf{x_1}, \ldots, \mathbf{x_T}$ is a set of independent but non identically distributed vectors, where each t component $\mathbf{x_t}$ corresponds to i.i.d. realizations of the density $f_{\theta_{\mathrm{i}}^t}$. This assumption is important for the texture model adopted in this work.

It is assumed 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 $x_{t,1}, \ldots, x_{t,L_t}$ are i.i.d. realizations of $f_{\theta_q^t}$ for each $t \in \{1, \ldots, T\}$. In this context, if t is fixed and $L_t \longrightarrow \infty$ is taken (and consequently $L \longrightarrow \infty$), the law of large numbers implies that [44]

$$\lim_{L_t \longrightarrow \infty} -\frac{1}{L_t} \sum_{j=1}^{L_t} \log f_{\theta_{\mathrm{i}}^t}(x_{t,j}) = -\mathbb{E}_{X \sim f_{\theta_q^t}} \log f_{\theta_{\mathrm{i}}^t}(X) = D(f_{\theta_q^t} || f_{\theta_{\mathrm{i}}^t}) + h(f_{\theta_q^t}), \tag{3.3}$$

almost surely. Here $D(f_{\theta_q^t} || f_{\theta_{\mathrm{i}}^t})$ is the Kullback-Leibler divergence of $f_{\theta_q^t}$ with respect to $f_{\theta_{\mathrm{i}}^t}$ [45] and $h(f_{\theta_q^t})$ is the differential entropy of $f_{\theta_q^t}$[45]. Finally, assuming a non-uniform rate of convergence, it defines [2]

$$\mathbf{w}_t = \lim_{L \longrightarrow \infty} \frac{L_t}{L}, \tag{3.4}$$

for all $t \in \{1, \ldots, T\}$ such that $\sum_{t=1}^{T} \mathbf{w}_t = 1$. The result in Eq. (3.3) can be used to show that the global likelihood rate convergences to

$$\lim_{L \longrightarrow \infty} -\frac{1}{L} \log f_{\theta_{\mathrm{i}}}(\mathbf{x}) = \sum_{t=1}^{T} \mathbf{w}_t \left(D(f_{\theta_q^t} || f_{\theta_{\mathrm{i}}^t}) + h(f_{\theta_q^t}) \right) \tag{3.5}$$

almost surely. Then, asymptotically (as the number of observations goes to infinity) the ML principle in Eq. (3.1) reduces to the minimum weighted divergence decision:

$$k_{\mathrm{i}} = \arg \min_{k \in [M] \setminus \{k_1, \ldots, k_{\mathrm{i}-1}\}} \sum_{t=1}^{T} \mathbf{w}_t \cdot D(f_{\theta_q^t} || f_{\theta_k^t}), \ \forall \mathrm{i} = 1, ..., N. \tag{3.6}$$

As stated in [3], the asymptotic equivalence between Eqs. (3.6) and (3.1) (when L goes to infinity) is instrumental because the ML decision reduces to find a consistent estimate $\hat{\theta}_q(\mathbf{x})$ of θ_q from $\mathbf{x}$, for instance using the ML estimator, and then using the weighted divergence in Eq. (3.6) to select the M closest models in $\bar{\Theta}$ of the target density $f_{\hat{\theta}_q(\mathbf{x})}$. The first stage is reminiscent of a feature extraction phase (or estimation phase) that goes from the image space to parameters, and the second stage is a selection phase using the weighted divergence (as a similarity metric) at the parameter level [3][3]. The next chapter contextualizes this framework for the case of texture indexing when the images are represented in a WP domain.

[2]It will be shown that the non-uniform rate assumption in Eq. (3.4) is important for the adoption of this MHT framework in the context of WPs.

[3]Under the assumption of a consistent estimator of θ_q, asymptotic connection between Eqs. (3.1) and (3.6) is achieved. Interestingly for the case of a discrete or quantized problem where the ML estimator is the empirical distributions, the connection between Eqs. (3.1) and (3.6) is achieved for any finite L [3, Section II-A].

Chapter 4

Wavelet Packet Texture Retrieval

This study extends the statistical model proposed by Do *et al.* [3] for texture image in the WP domain. Their idea was that the texture image can be represented in the transform domain as the concatenation of independent vectors (the coefficients obtained for each sub-band of the Wavelet basis), where each of these vectors has an invariant behavior in the sense that they are i.i.d. realizations of a given parametric distribution. This work extends this transform domain texture model when it is considered an arbitrary Wavelet Packet basis and, consequently, for a wider range of sub-band decomposition for $\mathbb{X}$ (see Eq. (2.8)). The purpose of this extension from wavelet transform to WP is that WP is capable to explore other space-scale domains. This allows us to obtain details of the texture that can improve the performance of the retrieval system. This work will demonstrate that using WP instead of wavelet basis improves the texture retrieval performance.

Considering $\mathcal{T} \ll \mathcal{T}_{full}$ with $\mathcal{L}(\mathcal{T}) = \{(j_k, p_k) : k = 1, \ldots, K\} \subset V$, then the texture model for a random image $X \in \mathbb{X}$ is determined in the transform domain of $\mathbf{B}_{\mathcal{T}}$. More precisely, the transform-based random vector $D_{\mathcal{T}}(X) = (D_{j_k}^{p_k}(X))_{k=1,\ldots,K}$ in Eq. (2.10) 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:

$$\begin{aligned} \underbrace{\log f_{\bar{\theta}}(D_{\mathcal{T}}(x))}_{\text{joint pdf in the transform domain}} &= \sum_{k=1}^{K} \log \underbrace{\prod_{n=1}^{4^{(J-j_k)}} f_{\theta^{((p_k,j_k))}}(\mathrm{d}_{j_k}^{p_k}(x,n))}_{\text{marginal pdf of the sub-band } U_{j_k}^{p_k}} \\ &= \sum_{k=1}^{K} \left(\underbrace{\sum_{n=1}^{4^{(J-j_k)}} \log f_{\theta^{((p_k,j_k))}}(\mathrm{d}_{j_k}^{p_k}(x,n))}_{\text{i.i.d.assumption for the transform coeff. at } U_{j_k}^{p_k}} \right), \end{aligned} \tag{4.1}$$

where $f_{\theta^{((p_k,j_k))}}$ is a pdf in $\mathbb{R}$ for each $k = 1, \ldots, K$. From Eq. (4.1), the components of the image projected at the sub-bands of WP are independent, and within each sub-band its transform coefficients are i.i.d. characterized by a pdf. For this last pdf, this work considers the Generalized Gaussian model (GGM) adopted in [3] with zero mean and parametrized by

$\theta = (\alpha, \beta) \in \mathbb{R}^2$ in the following way:

$$f_\theta(\mathrm{d}) = \frac{\beta}{2\alpha\Gamma\left(1/\beta\right)} \mathrm{e}^{-\left(\frac{|\mathrm{d}|}{\alpha}\right)^\beta}. \tag{4.2}$$

Finally, it is important to note the scaling on the number of transform coefficients for an arbitrary node $(j_k, p_k) \in \mathcal{L}(\mathcal{T})$. If the size of the image is denoted by $L = 4^J$ for some $J > 0$ (the dyadic case studied in Chapter 2.3), then the size of the vector $D_{j_k}^{p_k}(X)$ is $L/4^{j_k}$, which is a exclusive function of j_k (the number of arcs that connects (j_k, p_k) with the root $(0, 0)$). Then, it is clear that if $\mathcal{T}$ is not a balanced tree [37, 46], an asymmetric number of transform coefficients is obtained per sub-band. This should be considered in the asymptotic connection derived between ML and the divergence principle in Eq. (3.5).

In the context of the texture indexing problem, M probability models represent the texture database $\left\{f_{\bar{\theta}_\mathrm{i}} : \mathrm{i} = 1, \ldots, M\right\}$, each of them following the model in Eq. (4.1) and, consequently, they are fully characterized by

$$\bar{\Theta} = \left\{\bar{\theta}_\mathrm{i} = (\theta_\mathrm{i}^{((j_k,p_k))})_{k=1,\ldots,K} : \mathrm{i} = 1, \ldots, M\right\} \subset \Theta = \mathbb{R}^{2\cdot|\mathcal{T}|}.$$

In addition, it is assumed an underlying (hidden) query model $f_{\bar{\theta}_q}$ consistent with Eq. (4.1) 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 the MHT problem (3.1), considering the regime $L \longrightarrow \infty$ and $\mathcal{T} \ll \mathcal{T}_{full}$, reduces to:

$$\begin{aligned}
k_\mathrm{i} &= \arg \min_{l \in [M] \setminus \{k_1, \ldots, k_{\mathrm{i}-1}\}} \sum_{k=1}^{K} \mathbf{w}_{(j_k,p_k)} \cdot D(f_{\theta_q^{((j_k,p_k))}} || f_{\theta_l^{((j_k,p_k))}}), \\
&= \arg \min_{l \in [M] \setminus \{k_1, \ldots, k_{\mathrm{i}-1}\}} \sum_{k=1}^{K} \mathbf{w}_{(j_k,p_k)} \cdot \left[\log\left(\frac{\beta_q^{((j_k,p_k))} \alpha_l^{((j_k,p_k))} \Gamma(1/\beta_l^{((j_k,p_k))})}{\beta_l^{((j_k,p_k))} \alpha_q^{((j_k,p_k))} \Gamma(1/\beta_q^{((j_k,p_k))}))} \right) \right. \\
&\left. + \left(\frac{\alpha_q^{((j_k,p_k))}}{\alpha_l^{((j_k,p_k))}} \right)^{\beta_l^{((j_k,p_k))}} \frac{\Gamma((\beta_l^{((j_k,p_k))} + 1)/\beta_q^{((j_k,p_k))})}{\Gamma(1/\beta_q^{((j_k,p_k))})} - \frac{1}{\beta_q^{((j_k,p_k))}} \right],
\end{aligned} \tag{4.3}$$

from Eqs. (3.5), (3.6) and (4.2). Given the scaling on the number of transform coefficients described above, it is simple to show that $\mathbf{w}_{(j_k,p_k)} = 4^{-j_k}$ for all $k = 1, \ldots, K$. Then, 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 deeper in the tree, because the number of coefficients of the first groups is orders of magnitude greater than the number of coefficients of the second group. Finally as mentioned in Chapter 3, to implement Eq. (4.3) a first-stage (feature extraction) is conducted where the ML criterion is used to estimate θ_q from x.

Chapter 5

Wavelet Packet Basis Selection

WPs provide a wide range of representations for the indexing problem that 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 MHT formulation. However, texture discrimination is not the exclusive criterion in 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 transform coefficients to estimate the parameters in the FE phase of the indexing task (see Eq. (4.3)). This issue rises the existence of 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 errors [47, 37, 46, 48]. In particular, the following regularization problem is stated,

$$\mathcal{T}^*(\lambda) = \arg \min_{\mathcal{T} \ll \mathcal{T}_{full}} -\hat{R}(\mathcal{T}) + \lambda \cdot \Phi(\mathcal{T}), \tag{5.1}$$

where $\hat{R}(\mathcal{T})$ models the discrimination quality of the features induced by 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, the tree size is adopted for $\Phi(\mathcal{T}) = |\mathcal{T}|$ as it has been used in CART [37] and other tree learning problems [39, 49, 38, 50, 46] to model estimation errors. The assumption here is that the deviation of the estimated parameters in Eq. (4.3) from the true parameters is proportional to the size of the tree [37, 46]. For the fidelity measure, a global indicator of pair-wise weighted divergence is considered, used in Eq. (4.3), between classes given by

$$\hat{R}(\mathcal{T}) = \frac{1}{M(M-1)} \sum_{c=1}^{M} \sum_{\substack{k=1 \\ k \neq c}}^{M} \left(\sum_{t=1}^{T} \mathbf{w}_{(j_t,p_t)} D(f_{\theta_c^{((j_t,p_t))}} \| f_{\theta_k^{((j_t,p_t))}}) \right), \tag{5.2}$$

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 [45, 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[1].

5.1 Minimum Cost-Tree Pruning Algorithm

The type of regularization problem stated in Eq. (5.1) has been addressed by Breiman *et al.* [37] and Scott [46] in the context of decision trees for which efficient algorithms are available [46]. For the application of these results in the thesis context, the fidelity measure $\hat{R}(\mathcal{T})$ must be additive with respect to the tree [37, 46][2], which follows from construction as

$$\hat{R}(\mathcal{T}) = \sum_{t \in \mathcal{L}(\mathcal{T})} \hat{R}((j_t, p_t)), \tag{5.3}$$

with $\hat{R}((j_t,p_t)) \equiv \frac{1}{M(M-1)} \sum_{c=1}^{M} \sum_{\substack{k=1 \\ k \neq c}}^{M} \mathbf{w}_{(j_t,p_t)} D(f_{\bar{\theta_c}((j_t,p_t))} \| f_{\bar{\theta_k}((j_t,p_t))})$. Considering that both the fidelity and the cost terms are additive functionals of the tree and the fidelity function is nondecreasing, the result of Theorem 5.1 can be used to solve Eq. (5.1).

Theorem 5.1 *(*Scott [46, Th. 1] and Chou *et al.* [50, Th. 3]*) The complete solution of Eq. (5.1) is characterized by 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* $l \in \{1, .., m\}$, $\mathcal{T}^*(\lambda) = R_l$ *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. (5.1)[3]. The specification of the values of $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m = \infty$ and their respective trees $\{R_1, \ldots, R_m\}$ is called the solution of the *family pruning problem* [37, 46]. In addition, it is well-known from [38] that the set of possible solutions of Eq. (5.1), i.e., $\{\mathcal{T}^*(\lambda), \lambda \geq 0\}$, is given by the following *cost-fidelity problem* [46, 38]:

$$\mathcal{T}^{k*} = \arg \max_{\left\{\mathcal{T} \ll \mathcal{T}_{full} : |\mathcal{T}| \leq k\right\}} \hat{R}(\mathcal{T}), \tag{5.4}$$

for all $k = 1, 4, \ldots, |\mathcal{T}_{full}|$. For any admissible quad-tree size $k > 0$, Eq. (5.4) finds the tree of size equal to or smaller than k that maximizes the fidelity measure. Connecting these two characterizations (cost-fidelity solutions and the family pruning problem), it is direct to show that $\left\{\mathcal{T}^{k*}, k = 3l+1 \text{ with } l = 0, \ldots, \left(|\mathcal{T}_{full}| - 1\right)/3\right\} = \{R_1, \ldots, R_m\}$ and, consequently, the solutions of the cost-fidelity problem in Eq. (5.4) are embedded in the sense that $\{root\} = \mathcal{T}^{1*} \ll \mathcal{T}^{4*} \ll \ldots \ll \mathcal{T}^{(3 \cdot l+1)*}$, where l defines the number of tree branches. This embedded structure is the key to obtain implementable algorithms to solve $\left\{\mathcal{T}^{k*}, k = 3l+1 \text{ with } l = 0, \ldots, \left(|\mathcal{T}_{full}| - 1\right)/3\right\}$ and then $\{\mathcal{T}^*(\lambda), \lambda \geq 0\}$. For completeness, Algorithm 1 presents this solution[4].

[1] The details of this result are presented in Chapter 8.2.

[2] A functional $f(T)$ is additive if it is the sum of components of the leaves of T [46].

[3] In the sense that $\{\mathcal{T}^*(\lambda), \lambda \geq 0\} = \{R_1, \ldots, R_m\}$.

[4] The complexity of this algorithm is $O(|\mathcal{T}_{full}| \cdot \log(|\mathcal{T}_{full}|))$. See details in [46] and reference therein.

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, the solution of the original problem in Eq. (5.1) requires to know the true trade-off between fidelity and cost functions that is denoted by λ^*. In practice, this value can be determined experimentally via cross-validation or from a validation and test setting. The choice of one approach or another depends on the amount of data available. In this work, the selection of λ^* will be done considering an empirical risk minimization (ERM) approach over the admissible set of tree solutions given by $\{R_1, \ldots, R_m\}$.

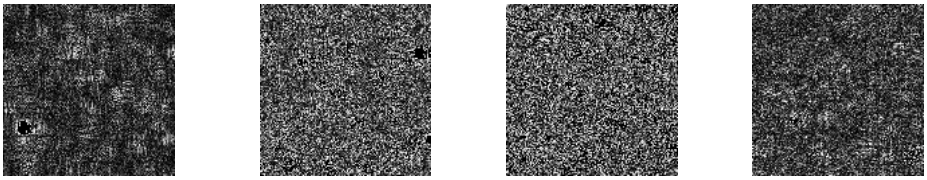
Chapter 6

Experimental Analysis

This chapter presents the results for the family of adaptive trees obtained as the solutions of Eq. (5.4). 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, the family of WP solutions indexed by $\left\{\mathcal{T}^{k*}, k = 3l + 1 \text{ with } l = 0, \ldots, \left(|\mathcal{T}_{full}| - 1\right)/3\right\}$ will be evaluated in terms of performance as well as the structure of their filter-bank decomposition and partition of the 2D frequency plane.

6.1 Synthetic Scenario

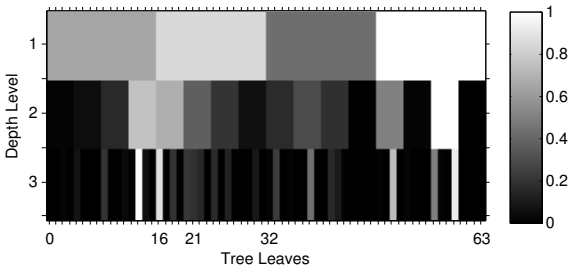
Before presenting the results on real texture databases, the adaptive capacity of the proposed framework will be evaluated on a controlled (synthetic) two texture indexing problem. For that, WP statistical models following the GGM model presented in Chapter 4 were selected by simply choosing a balanced tree of depth 3. In this context, the statistics of the GGM 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. (3.6). The idea was to put all the texture discrimination information in a specific sub-band to see if the solutions in Eq. (5.4) promote better frequency resolutions in this target band as k grows. Synthetic samples were created by simulating the transform coefficients of the two models (see examples of synthetic textures in Figure 6.1). From these data, the fidelity measure was estimated using Eq. (5.2) considering the weighted divergence for 16 examples per class. Figure 6.2 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 $\mathbf{w}_{(j_t,p_t)} D(f_{\theta_c((j_t,p_t))} \| f_{\theta_k((j_t,p_t))})$ in Eq. (5.2) associated with the additive contribution of the node (j_t, p_t) organized by raw j_t and column p_t. From these graphics, it is possible to see how the weighted divergence appropriately captures the frequency bands that are more discriminative for the task. For the illustration, two contexts are considered where the discriminative bands are indexed by the pair $(3, 21)$ and $(3, 49)$, respectively. From the result, it is observed that the proposed method captures the discrimination of this task on the right bands and, consequently,



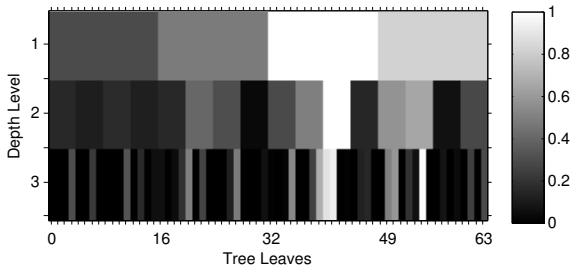
(a) (b) (c) (d)

Figure 6.1: Examples of synthetic textures.

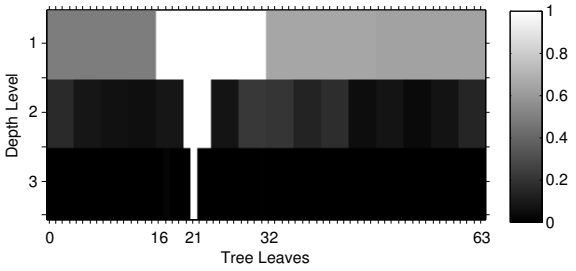
$\left\{\mathcal{T}^{k*}, k = 3l + 1 \text{ with } l = 0, \ldots, \left(|\mathcal{T}_{full}| - 1\right)/3\right\}$ increases the resolution on the bands that are more informative. In contrast, it is shown that the amount of energy of the leaves is not always mapped to the parents and also the modified node is not always the more energetic one. Those factors produce that energy as a fidelity measure does not capture the more discriminative bands for the task.



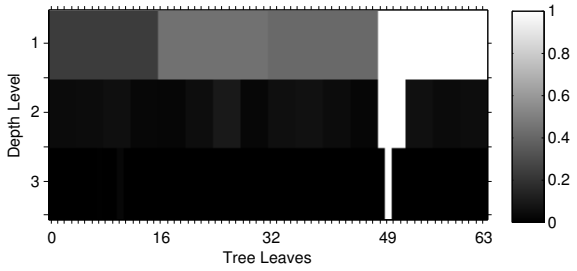
(a) Example 1: Energy (Baseline)



(b) Example 2: Energy (Baseline)



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



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

Figure 6.2: These maps shows the weighted divergence and energy term indexed by a sub-band in the WP decomposition.

6.2 Texture Databases

For the analysis on real texture images, the evaluation considers the VisTex [51] (Figure 6.3), Brodatz [52] (Figure 6.4), STex [53] (Figure 6.5), and ALOT [54] (Figure 6.6) datasets. The VisTex database is composed of 40 color textures of 512×512 pixels, which is considered in [3]. The Brodatz consists of 30 gray-scale textures of 640×640 pixels each, which is the setting used in [28]. Moving on larger sets, it is extracted two collection of images from STex and ALOT databases respectively to create the full STex, full ALOT, reduced STex, and reduced ALOT. On the one hand, the full STex and full ALOT contain 436 color and 250 gray-scale textures of 1024×1024 pixels and 1536×1024 pixels, respectively. On the other hand, 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 VisTex and Brodatz is divided in 16 and 25 non-overlapping textures of 128×128 pixels, respectively. However, 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 this division, the color images are transformed into gray-scale versions and, subsequently, all images on the collections are normalized to zero mean and unit variance, which is a standard normalization used on previous studies [3, 8, 9, 55, 10]. In summary, this work uses six databases with different images sizes and number of textures, which offer a rich context to evaluate the potential of WP based texture indexing.

For each context, the dataset is divided in training and testing in a proportion of $3:7$, where 30% of the data is used to find the tree-structure solving Eq. (5.4) and the remaining 70% is used for the evaluation of the texture retrieval. The training phase involves computing the fidelity measure in Eq. (5.2) to obtain the family $\left\{\mathcal{T}^{k*}, k\right\}$. The testing phase computes the performances of this family as a function of the size of the tree. For the retrieval performance, it is used one query example per class in the dataset and the standard average recall metric adopted in [3] 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 the case of this work $C = \{40, 30, 436, 250\}$), M is the number of examples 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 in all the analysis, *Daubechies 4* is used as the mother Wavelet basis, because it is the one considered by Do *et al.* [3] and the one that shows the best performances for analysis made in this thesis.

Figure 6.3: Texture images VisTex used for the experiments. Bark0, Bark6, Bark8, Bark9, Brick1, Brick4, Brick5, Buildings9, Fabric0, Fabric4, Fabric7, Fabric9, Fabric11, Fabric14, Fabric15, Fabric17, Fabric18, Flowers5, Food0, Food5, Food8, Grass1, Leaves8, Leaves10, Leaves11, Leaves12, Leaves16, Metal0, Metal2, Misc2, Sand0, Stone1, Stone4, Terrain10, Tile1, Tile4, Tile7, Water5, Wood1 y Wood2.

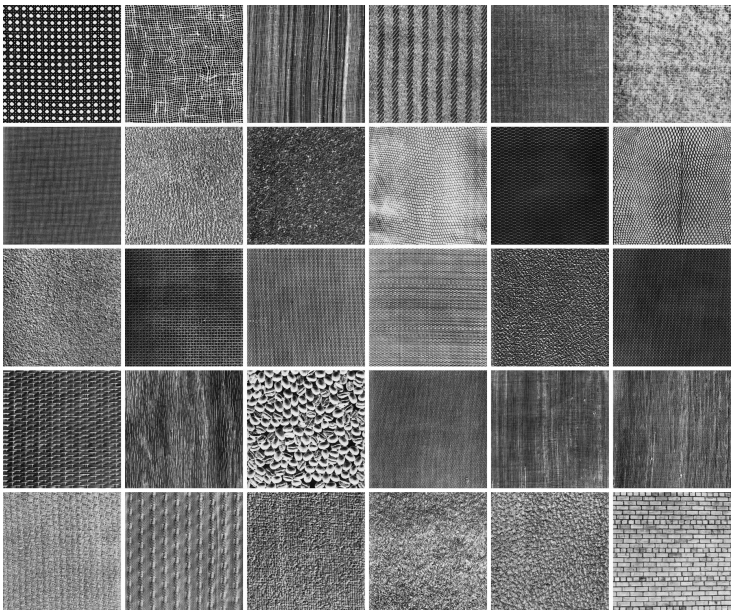


Figure 6.4: Texture images Brodatz used for the experiments. D102, D103, D105, D11, D16, D19, D21, D24, D29, D3, D34, D36, D4, D52, D53, D55, D57, D6, D65, D68, D74, D77, D78, D79, D82, D83, D84, D9, D92 y D95.

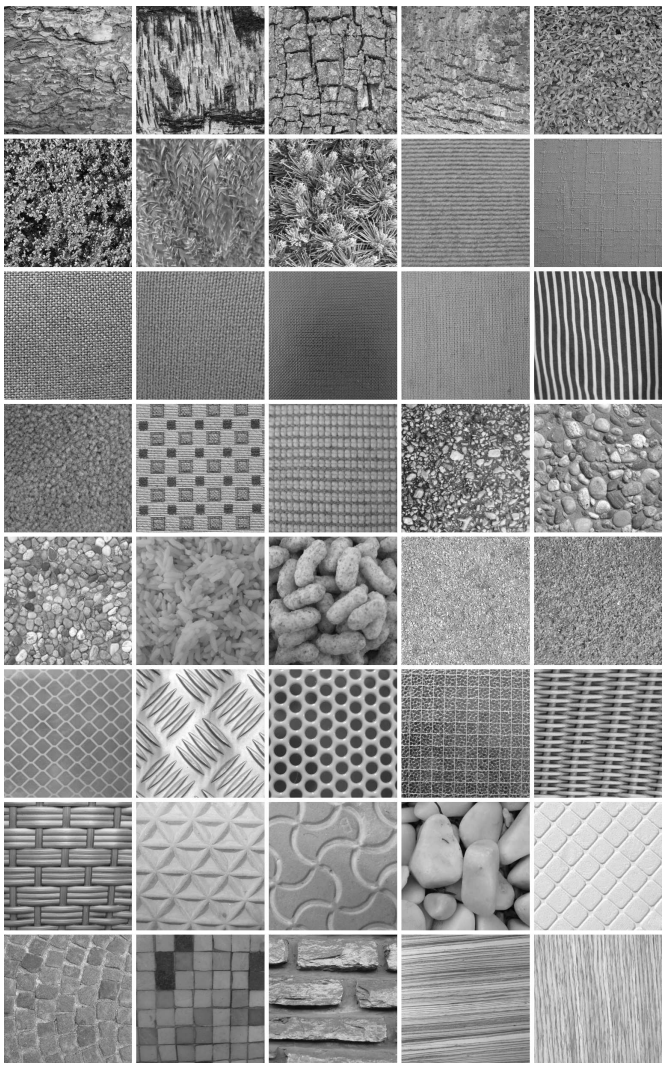


Figure 6.5: Example of texture images STex used for the experiments. Bark0000, Bark0003, Bark0004, Bark0012, Bush0000, Bush0003, Bush0009, Bush0015, Fabric0002, Fabric0003, Fabric0008, Fabric0027, Fabric0028, Fabric0030, Fabric0031, Fabric0052, Fabric0055, Fabric0073, Floor0002, Floor0003, Floor0004, Food0007, Food0008, Gravel0005, Gravel0011, Metal0000, Metal0010, Metal0018, Misc0018, Misc0021, Rattan0000, Rattan0002, Rubber0005, Rubber0007, Stone0005, Styrofoam0000, Tile0002 Tile0003, Wall0006, Wood0034 y Wood0035.

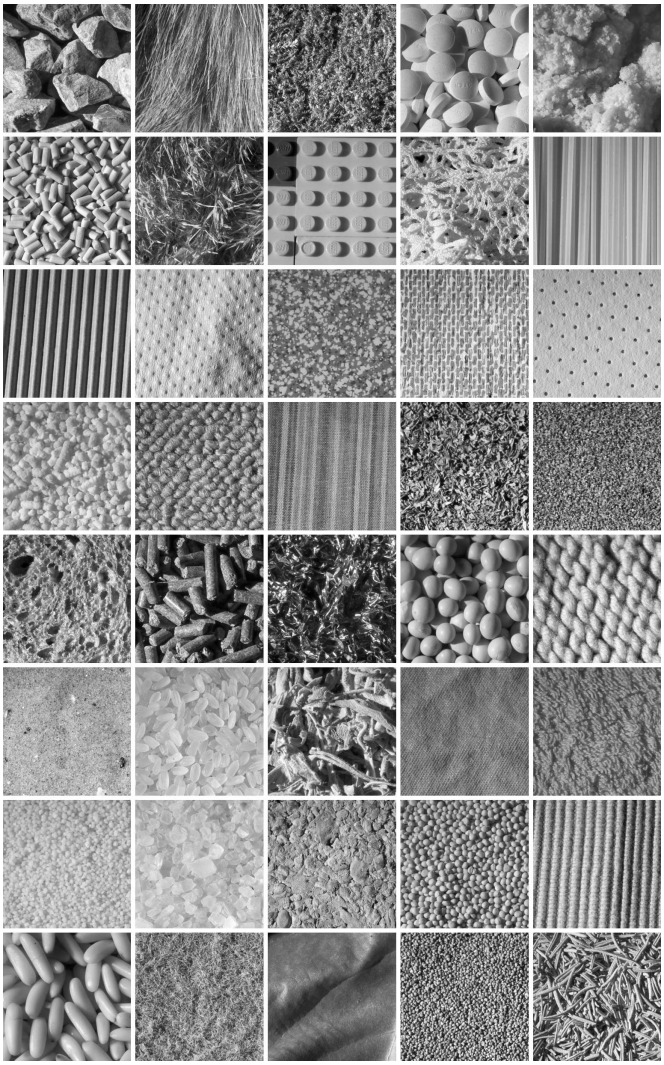


Figure 6.6: Example of texture images ALOT used for the experiments. Fruit sprinkles, couscous, toilet paper, spearmint, bread (wholewheat), silver foil (heavily cramped), soya beans, wool knitwear, sand (wet), ricotta rice, sage, cotton (color orange), terry cloth (orange), rock salt, cork, mustard seed (yellow) ribbed cotton (yellow), pine seeds, scourer (rough side), lef (brown), poppy seed, rosemary leaves, stones (gray level), fake fur, moss (green), vitamin C pills (roter), flan (apple), snail poison, reed (plumes), lego (plates yellow and blue), lace, spaguetti (regularly ordered), ribbed cardboard, carpet (blue), wallpaper (beige motive), chamois (punched), carpet (beige), cotton (blue and purples stripes), sand paper (roughness 40), and rabbit food.

6.3 Analysis of the Sub-Band Model Fitting

Considering that WPs offer a rich range of filter-bank decomposition, this chapter evaluates numerically the fitting of the GGM model considered in Eq. (4.2) to model the statistical dependencies of the transform coefficients in each of the induced sub-bands as considered in Chapter 4. For that it is used the ML estimator of its parameters for numerous sub-bands and texture in the rich collection of datasets. In general, it was observed that the GGM model captures the marginal statistics of transforms coefficients of any arbitrary band and, consequently, the modeling extension adopted in Chapter 4 seems reasonable. For illustration, Figure 6.7 presents some of these fittings.

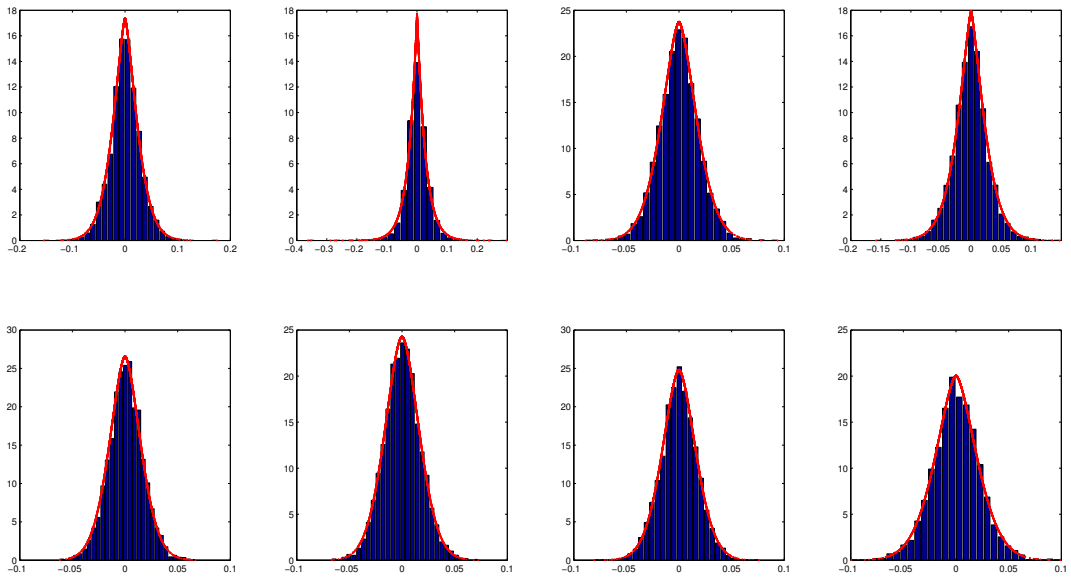


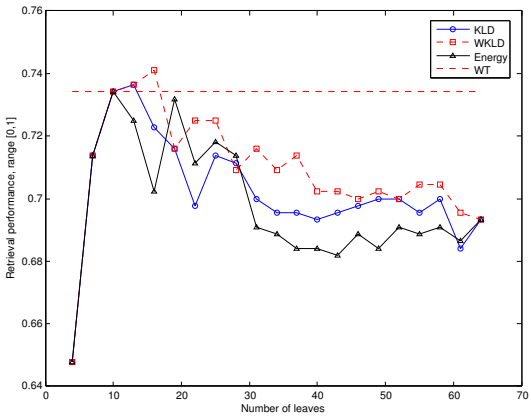
Figure 6.7: Histograms and ML fitting of the WP sub-band coefficients associated with the nodes indexed by $(2,4)$, $(2,5)$, $(2,6)$, $(2,7)$, $(2,12)$, $(2,13)$, $(2,14)$, $(2,15)$. The Generalized Gaussian Model in Eq. (4.2) is used for the histogram fitting. The image *20_c1l1* from ALOT database is used for illustration.

6.4 Wavelet Packets Retrieval Performances

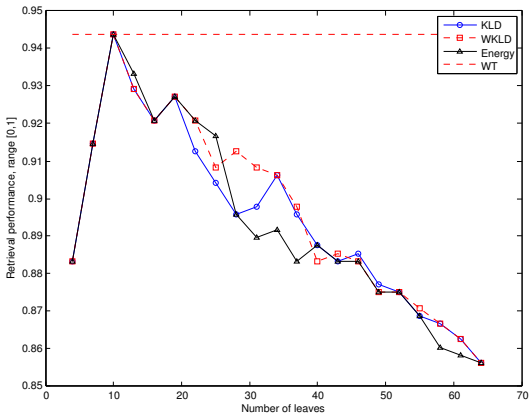
This work proposes a novel methodology for the WP basis selection for texture retrieval. Because of that, this chapter presents the retrieval performance associated with the solutions of Eq. (5.4) for each of the datasets adopting the weighted divergence in Eq. (5.2), and then comparing the retrieval performance of different WP bases with the wavelet solution. For completeness, it is also considered the tree solutions obtained for the same regularization problem but adopting the non-weighted divergence and the energy per-band used in [28] as alternatives fidelity measures.

Figure 6.8 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 baseline and is denoted by WT. Overall, it is observed the expected trade-off between estimation and approximation errors in the evolution of each of the performance curves. At the beginning, more complex trees improves dramatically the retrieval accuracy. Then, the estimation error dominates, implying a saturation which leads to a deterioration on the retrieval performances as k increases. It is important to note that these changes on the regime of the performance curves is a function of the complexity of the task. For the smaller (VisTex and Brodatz) and larger image size datasets (STex and ALOT) these changes happen in the range $[10-34]$ of the tree sizes. From these curves, it can be determined 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 happen for the Full ALOT, Reduced ALOT and Reduced STex. However, the performances deviate from the best solution (obtained with the energy fidelity and non-weighted divergence) very slightly, which are 0.76%, 1.58%, and 0.31% relative for each case, respectively. It is also interesting to note that WP solutions associated with the energy as a fidelity measure present very 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. This suggests that energy is a very good proxy of the discrimination power of the WP bands in natural texture indexing, which is something that was also observed on the analysis of acoustic signal on automatic speech recognition (ASR) [39].

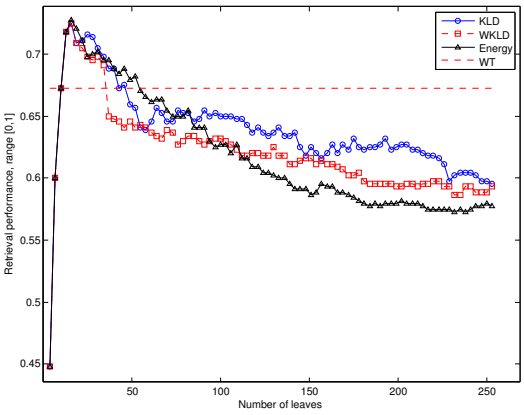
Finally, it is provided the performance of the Wavelet basis used in [3] for performance comparison (see dashed lines in Figure 6.8). Remarkably, the conjecture of this work is confirmed that stated that the family of WPs and their richer sub-band decompositions offer descriptions with relevant improvements in indexing performance. This gain is more prominent in the case of the largest size datasets, where alternative WP trees show very significant gains with respect to the conventional Wavelet solution. Nevertheless, the improvements were not so significant for the case of smaller databases (VisTex and Brodatz). A possible explanation is that in the context of smaller texture images WPs do not have room to take advantage the discrimination power of non-conventional (Wavelet type) frequency bands. Rapidly moving to the regime were estimation error dominates the performance curves. Interestingly in these contexts, the smaller sizes trees (sizes 4, 7 and 10) matches the Wavelet solution and, consequently, it is recovered as part of this work formulation the solution proposed in [3] as the optimal basis for these datasets. Table 6.1 presents the best performance obtained (optimal trees) for each fidelity measure and the performances of the Wavelet solution which is denoted by WT. The last column of Table 6.1 shows the gain in relative percentage with respect to the Wavelet solution, which shows that the best improvements of the retrieval performance are achieved for the larger size texture datasets.



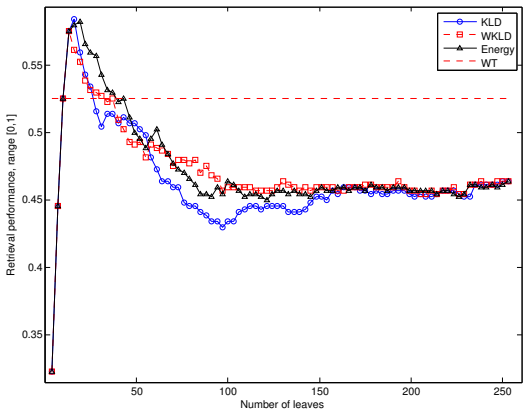
(a) VisTex



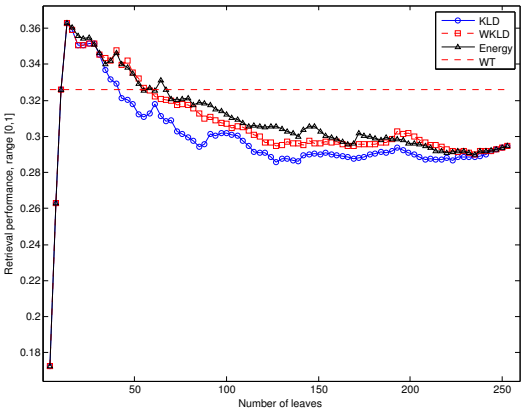
(b) Brodatz



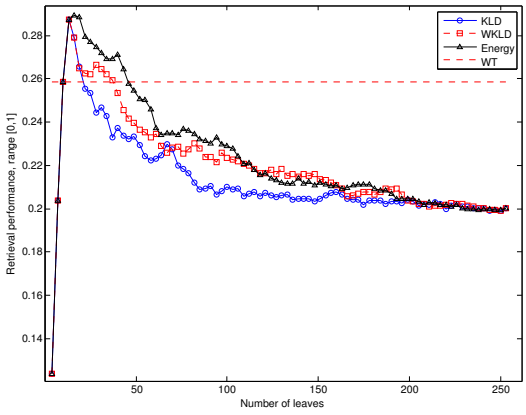
(c) Reduced STex



(d) Reduced ALOT



(e) Full STex



(f) Full ALOT

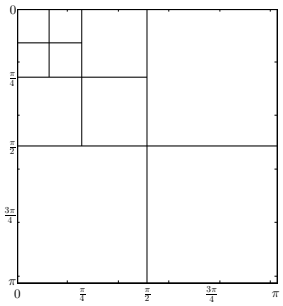
Figure 6.8: Retrieval performance for the family of Wavelet Packets solutions of the regularized problem in Eq. (5.1) using the weighted divergence, the divergence and the energy as fidelity measures. Results are presented independently for the databases: VisTex, Brodatz, Full STex, Full ALOT, Reduced STex and Reduced ALOT.

Table 6.1: Best Retrieval Performance

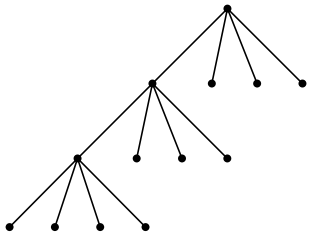
Texture	WT	WP Energy	WP KLD	WP WKLD	Best Relative Gain
VisTex	73.41%	73.41%	73.64%	**74.09%**	0.92%
Brodatz	**94.38%**	**94.38%**	**94.38%**	**94.38%**	0%
Reduced STex	67.27%	**72.73%**	72.50%	72.50%	7.77%
Reduced ALOT	52.50%	58.18%	**58.41%**	57.50%	11.25%
Full STex	32.60%	**36.27%**	**36.27%**	**36.27%**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

6.5 Analysis of the Optimal Tree Structure of WPs

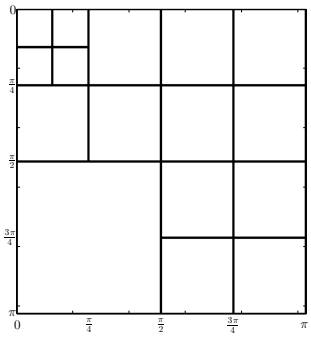
The filter-bank decomposition of the best WPs solutions for each datasets are presented in Figures 6.9 and 6.10. It is observed that the solutions follow a Wavelet type of path as low frequencies are iterated in the decomposition at the beginning of the WP decomposition. However, for many of the scenarios, there are non-trivial deviations from the Wavelet path as other frequency bands are iterated on the process of creating the optimal trees. These non-Wavelet type of bands offer better discrimination than the recursive iteration of the low frequency that defines a Wavelet case.



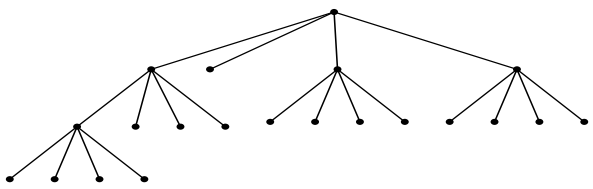
(a) Brodatz frequency partition.



(b) Brodatz tree-structure.

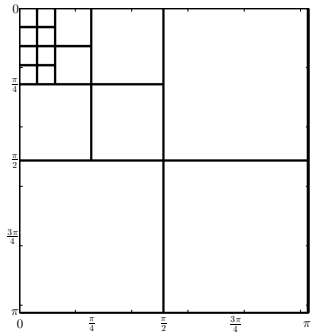


(c) VisTex frequency partition.

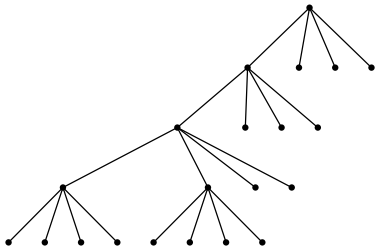


(d) VisTex tree-structure.

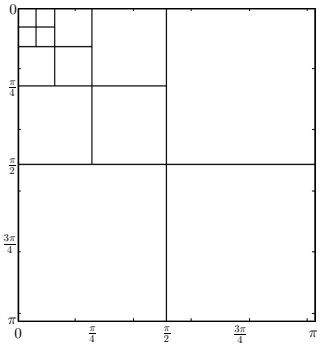
Figure 6.9: First column shows the frequency partitions induced by the best WP solution, which is represented by the tree in the second column.



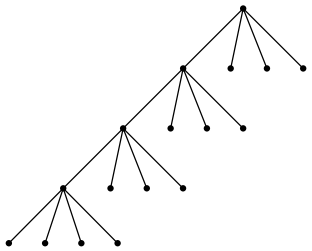
(a) Reduced STex and Reduced ALOT frequency partition.



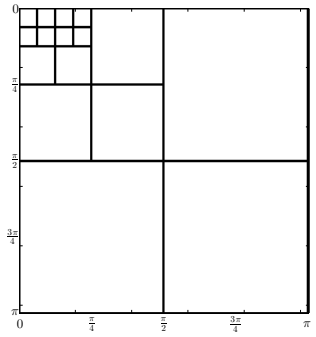
(b) Reduced STex and Reduced ALOT tree-structure.



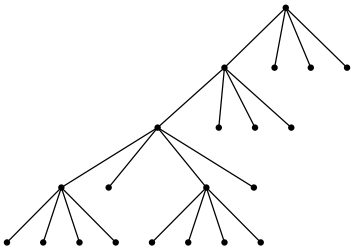
(c) Full STex frequency partition.



(d) Full STex tree-structure.



(e) Full ALOT frequency partition.



(f) Full ALOT tree-structure.

Figure 6.10: First column shows the frequency partitions induced by the best WP solution, which is represented by the tree in the second column.

Chapter 7

Conclusions and Future Work

This work shows performance improvements for texture retrieval from the design of a new set of filter-bank discriminative features. The tree-indexed WP collection was used to find an adequate balance between feature discrimination and some learning over-fitting effects. On the implementation of this idea, 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. It is shown 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 (aspect of the problem that is tightly related with over-fitting). Overall, WPs offer features that outperforms the Wavelet representation as an evidence that the exploration of different sub-bands offers a better texture discrimination than the standard Wavelet filter-bank partition.

As a direction of future research, it is conjectured that the idea of adaptive WP filter-banks will benefit from the adoption of better texture models, in particular by adding non-trivial inter-band dependencies for instance copulas [23, 24], changing the wavelet packets coefficients models [19, 20, 32, 31] and other improvements such as the consideration of the parameter error estimation in the formulation of the problem because deeper nodes have less data for the parameter estimation. Furthermore, the adaptive nature of the proposed methodology provides an opportunity to incorporate the WP methodology in diverse kind of problems such as face recognition and emotion recognition. In the face recognition problem, the main idea would be to find the best WP tree-structure (basis) that represents a face providing an improvement on the performance of face-no face discrimination tasks. Finally, this methodology can be used beyond faces detection to achieve recognition of human emotions. For that, an initial representation of the facial landmarks with WP tree-structure could result in an advanced characterization capable of discriminating among emotions.

Chapter 8

Appendix

8.1 The Gray Code

The frequency partition generated by filter bank iterations is not ordered in the frequency domain [36, Chap. 8] and for this reason the Gray Code is used to find the ordered frequency partition. Thus, to find the ordered frequency partition, it is needed to evaluate the range of frequencies that contributes with the image texture discrimination and comparing WP basis frequency partition with the frequency partition generated by the Wavelet solution.

Proposition 8.1 *[36, Chap. 8.1.2] Let (j,p) be an admissible sub-band of the Shannon WP decomposition. Associating to $p \in \{0,\ldots,4^j-1\}$ a pair $(q_1,q_2) \in \{0,\ldots,2^j-1\}^2$. Where q_1, q_2 are the indexes of the 1D subbands that generate the sub-band U_j^p. Then, its equivalent frequency-ordered label (j,k) is constructed by the following rule*

$$k_{\mathrm{i}} = G(q_{\mathrm{i}}) \equiv \sum_{h=1}^{j-L} \bar{q}_{\mathrm{i},h} \cdot 2^h \in \left\{0,\ldots,2^{j-L}-1\right\}, \qquad \mathrm{i} \in \{1,2\}$$

where $\bar{q}_{\mathrm{i},h} \equiv \left(\sum_{l=h}^{j-L} q_{\mathrm{i},l}\right) \bmod 2 \in \{0,1\}$, $\forall h \in \{1,\ldots,j-L\}$. Hence, the ordered frequency partition of a 2D signal is given by the following intervals

$$\begin{aligned}&\left[-(k_1+1)2^{j-L}\pi, -k_1 2^{j-L}\pi\right] \times \left[-(k_2+1)2^{j-L}\pi, -k_2 2^{j-L}\pi\right] \cup \\ &\left[k_1 2^{j-L}\pi, (k_1+1)2^{j-L}\pi\right] \times \left[k_2 2^{j-L}\pi, (k_2+1)2^{j-L}\pi\right].\end{aligned}$$

Because of the symmetry of these intervals, it is possible to do the WP frequency analysis using the positive part of the intervals.

$$\left[k_1 2^{j-L}\pi, (k_1+1)2^{j-L}\pi\right] \times \left[k_2 2^{j-L}\pi, (k_2+1)2^{j-L}\pi\right].$$

8.2 Weighted divergence and Stein's Lemma

It is proposed the extension of *Stein's Lemma* [45, Th. 12.8.1] to the weighted divergence. The purpose is to support the decision of considering weighted divergence as fidelity measure. The Stein's Lemma is restated for the proposed problem as follows,

Theorem 8.2 *(*Stein's Lemma restated*) Let* $D_{\mathcal{T}}(x) = \left(D_{j_t}^{p_t}(x)\right)_{t=1,\dots,T.}$ *be i.i.d.* $\sim f_{\bar{\theta}}$ *with* $\bar{\theta} = \left(\theta_{\mathrm{i}}^{((j_t,p_t))}\right)_{t=1,\dots,T}$. *Consider the hypothesis test between two alternatives,* $f_{\bar{\theta}} = f_{\bar{\theta}_1}$ *and* $f_{\bar{\theta}} = f_{\bar{\theta}_2}$, *where* $D(f_{\bar{\theta}_1} \| f_{\bar{\theta}_2}) < \infty$. *Let* $A_L \subset \mathcal{X}^L$ *be an acceptance region for hypothesis* H_1. *Let the probabilities of error be*

$$\alpha_{L,A_L} = P_{\bar{\theta}_1}^L(A_L^c), \qquad \beta_{L,A_L} = P_{\bar{\theta}_2}^L(A_L).$$

And for $0 < \epsilon < \frac{1}{2}$, *define*

$$\beta_{L,A_L}^{\epsilon^*} = \min_{\substack{A_{L,A_L} \subset \mathcal{X}^L \\ \alpha_{L,A_L} < \epsilon}} \beta_{L,A_L}.$$

Then,

$$\lim_{L\to\infty} \frac{1}{L} \log\left(\beta_{L,A_L}^{\epsilon^*}\right) = -\sum_{t=1}^{T} \boldsymbol{w}_t D(f_{\theta_1}^{((j_t,p_t))} \| f_{\theta_2}^{((j_t,p_t))}).$$

Proof. The proof is based on the *Stein's Lemma* [45, Th. 12.8.1]. Let $\mathbb{X}$ be a raw image and be $x \in \mathbb{X}$, in which transformed coefficients are given by $D_{\mathcal{T}}(x) = \left(D_{j_t}^{p_t}(x)\right)_{t=1,\dots,T.}$. Each node (j_t, p_t) contains transformed coefficients $\left(\mathrm{d}_{j_t}^{p_t}(x,n)\right)_{n=1,\dots,L_t}$, which the sum of their size is the total number of coefficients $\dim(\mathbb{X}) = L = \sum_{t=1}^{T} L_t$.
From Chapter 3, it is known that all of the data in a leaf is i.i.d. and each leaf is independent from the others. Then, let $\left(\theta^{((j_t,p_t))}\right)_{t=1,\dots,T}$ be the parameters of the statistical model of $\left(D_{j_t}^{p_t}(x)\right)_{t=1,\dots,T.}$, where each statistical model μ_θ is equipped with a density function f_θ . It is stated the following hypothesis test,

$$\begin{aligned} H_1 &: f_\theta = f_{\theta_1}, \\ H_2 &: f_\theta = f_{\theta_2}. \end{aligned}$$

With $\theta_{\mathrm{i}} = \left\{\left(\theta_{\mathrm{i}}^{((j_t,p_t))}\right)_{t=1,\dots,T}\right\}_{\mathrm{i}=1,2}$ and $\{f_{\theta_{\mathrm{i}}}\}_{\mathrm{i}=1,2}$ their respective density functions, each one equipped with statistical models $\{\mu_{\theta_{\mathrm{i}}}\}_{\mathrm{i}=1,2}$. For the achievability part, it is considered the log-likelihood ratio as follows,

$$\frac{1}{L} \log\left(\frac{f_{\theta_1}(D_{j_1}^{p_1}(x), \dots, D_{j_T}^{p_T}(x))}{f_{\theta_2}(D_{j_1}^{p_1}(x), \dots, D_{j_T}^{p_T}(x))}\right). \tag{8.1}$$

Because the sub-bands are independent, Eq. (8.1) can be written as

$$\frac{1}{L} \sum_{t=1}^{T} \log\left(\frac{f_{\theta_1}(D_{j_t}^{p_t}(x))}{f_{\theta_2}(D_{j_t}^{p_t}(x))}\right). \tag{8.2}$$

Furthermore, the transformed coefficients are considered i.i.d. then,

$$\frac{1}{L}\sum_{t=1}^{T}\sum_{j=1}^{L_t}\log\left(\frac{f_{\theta_1}(D_{j_t}^{p_t}(x_{t,j}))}{f_{\theta_2}(D_{j_t}^{p_t}(x_{t,j}))}\right) = \sum_{t=1}^{T}\frac{L_t}{L}\sum_{j=1}^{L_t}\frac{1}{L_t}\log\left(\frac{f_{\theta_1}(D_{j_t}^{p_t}(x_{t,j}))}{f_{\theta_2}(D_{j_t}^{p_t}(x_{t,j}))}\right), \tag{8.3}$$

where $x_{t,j}$ is the j coefficient of the node or sub-band t. Similar to Chapter 3, if $\mathbf{w}_t = \lim_{L\longrightarrow\infty}\frac{L_t}{L}$ is considered, then when $L\longrightarrow\infty$ Eq. (8.3) is equivalent to:

$$\sum_{t=1}^{T}\mathbf{w}_t\mathbb{E}_{D_{j_t}^{p_t}(x_{t,j})\sim f_{\theta_1}^{((j_t,p_t))}}\log\left(\frac{f_{\theta_1}(D_{j_t}^{p_t}(x_{t,j}))}{f_{\theta_2}(D_{j_t}^{p_t}(x_{t,j}))}\right) = \sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}}). \tag{8.4}$$

Thus,

$$\frac{1}{L}\log\left(\frac{f_{\theta_1}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}{f_{\theta_2}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}\right)\longrightarrow\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}}) \tag{8.5}$$

in probability one. The sequence of acceptance region can de defined as follows,

$$\begin{aligned}A_L(\delta) = \Bigg\{&D_{\mathcal{T}}(x): 2^{L(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})-\delta)}\leq\frac{f_{\theta_1}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}{f_{\theta_2}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))} \quad\wedge\\ &\frac{f_{\theta_1}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}{f_{\theta_2}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}\leq 2^{T(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})+\delta)}\Bigg\}\end{aligned}$$

The acceptance probability is:

$$\begin{aligned}P_{\theta_1}^{L}(A_L(\delta)) &= P_{\theta_1}^{L}\left(\left|\frac{1}{L}\frac{f_{\theta_1}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}{f_{\theta_2}(D_{j_1}^{p_1}(x),\ldots,D_{j_T}^{p_T}(x))}-\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})\right|<\delta\right),\\ &= 1-\alpha_{L,A_L(\delta)}.\end{aligned}$$

Also, it has the following:

$$\begin{aligned}P_{\theta_2}^{L}(A_L(\delta)) &= \sum_{A_L(\delta)} f_{\theta_2}(D_{\mathcal{T}}(x)),\\ &\leq \sum_{A_L(\delta)} f_{\theta_1}(D_{\mathcal{T}}(x))2^{-L(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})-\delta)},\\ &= 2^{-L(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})-\delta)}(1-\alpha_{L,A_L(\delta)}),\end{aligned}$$

and,

$$\begin{aligned}P_{\theta_2}^{L}(A_L(\delta)) &= \sum_{A_L(\delta)} f_{\theta_2}(D_{\mathcal{T}}(x)),\\ &\geq \sum_{A_L(\delta)} f_{\theta_1}(D_{\mathcal{T}}(x))2^{-L(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})+\delta)},\\ &= 2^{-L(\sum_{t=1}^{T}\mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}}\|f_{\theta_2^{((j_t,p_t))}})+\delta)}(1-\alpha_{L,A_L(\delta)}).\end{aligned}$$

Therefore,

$$\begin{aligned} 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})+\delta)}(1-\alpha_{L,A_L(\delta)}) &\leq \beta_{L,A_L(\delta)}, \\ 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})-\delta)}(1-\alpha_{L,A_L(\delta)}) &\geq \beta_{L,A_L(\delta)}. \end{aligned}$$

Then,

$$\begin{aligned} -\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}}) - \delta + \frac{\log(1-\alpha_{L,A_L(\delta)})}{L} &\leq \frac{1}{L}\log(\beta_{L,A_L(\delta)}), \\ -\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}}) + \delta + \frac{\log(1-\alpha_{L,A_L(\delta)})}{L} &\geq \frac{1}{L}\log(\beta_{L,A_L(\delta)}). \end{aligned}$$

It is known from the law of large numbers that as $L \to \infty$ then $\alpha_{L,A_L(\delta)} \to 0$ and $P_{\theta_1}^L(A_L(\delta)) \to 1$. Fixing $\epsilon = \delta > 0$, thus

$$\lim_{\epsilon \to 0} \lim_{L \to \infty} \frac{1}{L}\log(\beta_{L,A_L(\epsilon)}) = -\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}}).$$

Finally, for the converse, it is demonstrated that there is no other sequence of acceptance regions better than the previous one. Let B_T be a sequence of acceptance region with $\alpha_{L,B_L} = P_{\theta_1}^L(B_L) < \epsilon$.

$$\begin{aligned} \beta_{L,B_L} = P_{\theta_2}^L(B_L) &\geq P_{\theta_2}^L(A_L \cap B_L), \\ &= \sum_{A_L \cap B_L} f_{\theta_2}(D_{\mathcal{T}}(x)), \\ &\geq \left(\sum_{A_L \cap B_L} f_{\theta_1}(D_{\mathcal{T}}(x)) \right) 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})+\delta)}, \\ &= (\mu_{\theta_1}(A_L \cap B_L))\, 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})+\delta)}, \\ &\geq (1-\mu_{\theta_1}(A_L^c) - \mu_{\theta_1}(B_L^c))\, 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})+\delta)}, \\ &= (1-\alpha_{L,A_L} - \alpha_{L,B_L})\, 2^{-L(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}})+\delta)}. \end{aligned}$$

Repeating the previous procedure on the last equation,

$$\lim_{L \to \infty} \frac{1}{L}\log(\beta_{L,B_L}) \geq -\left(\sum_{t=1}^{T} \mathbf{w}_t D(f_{\theta_1^{((j_t,p_t))}} \| f_{\theta_2^{((j_t,p_t))}}) + \delta \right),$$

for a δ arbitrarily small. □

Bibliography

[1] J. C. Felipe, A. J. M. Traina, and C. Traina, "Retrieval by content of medical images using texture for tissue identification," in *16th IEEE Symposium Computer-Based Medical Systems, 2003. Proceedings.*, pp. 175–180, June 2003.

[2] F. Kayitakire, C. Hamel, and P. Defourny, "Retrieving forest structure variables based on image texture analysis and ikonos-2 imagery," *Remote Sensing of Environment*, vol. 102, no. 3, pp. 390 – 401, 2006.

[3] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian densities and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[4] P. de Rivaz and N. Kingsbury, "Complex wavelet features for fast texture image retrieval," in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, vol. 1, pp. 109–113 vol.1, 1999.

[5] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, pp. 1168–1178, Dec 2005.

[6] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Rotation-invariant texture image retrieval using rotated complex wavelet filters," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, pp. 1273–1282, Dec 2006.

[7] R. Kwitt and A. Uhl, "A joint model of complex wavelet coefficients for texture retrieval," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 1877–1880, Nov 2009.

[8] M. N. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models," *IEEE Transactions on Multimedia*, vol. 4, pp. 517–527, Dec 2002.

[9] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with gaussianized steerable pyramids," *IEEE Transactions on Image Processing*, vol. 15, pp. 2702–2718, Sept 2006.

[10] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture re-

trieval via signature alignment based on steerable sub-gaussian modeling," *IEEE Transactions on Image Processing*, vol. 17, pp. 1212–1225, July 2008.

[11] C. Li, G. Duan, and F. Zhong, "Rotation invariant texture retrieval considering the scale dependence of gabor wavelet," *IEEE Transactions on Image Processing*, vol. 24, pp. 2344–2354, Aug 2015.

[12] C. Li, Y. Huang, and L. Zhu, "Color texture image retrieval based on gaussian copula models of gabor wavelets," *Pattern Recognition*, vol. 64, pp. 118 – 129, 2017.

[13] Y. Wang, M. Shi, S. You, and C. Xu, "Dct inspired feature transform for image retrieval and reconstruction," *IEEE Transactions on Image Processing*, vol. 25, pp. 4406–4420, Sept 2016.

[14] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, pp. 2091–2106, Dec 2005.

[15] G. Kutyniok and D. Labate, *Shearlets: Multiscale Analysis for Multivariate Data*. Birkhäuser Basel, 2012.

[16] Y. Dong and J. Ma, *Statistical Contourlet Subband Characterization for Texture Image Retrieval*, pp. 495–502. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[17] M. S. Allili, N. Baaziz, and M. Mejri, "Texture modeling using contourlets and finite mixtures of generalized gaussian distributions and applications," *IEEE Transactions on Multimedia*, vol. 16, pp. 772–784, April 2014.

[18] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, "Texture classification and retrieval using shearlets and linear regression," *IEEE Transactions on Cybernetics*, vol. 45, pp. 358–369, March 2015.

[19] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval," *IEEE Transactions on Image Processing*, vol. 19, pp. 281–289, Feb 2010.

[20] M. S. Allili, "Wavelet modeling using finite mixtures of generalized gaussian distributions: Application to texture discrimination and retrieval," *IEEE Transactions on Image Processing*, vol. 21, pp. 1452–1464, April 2012.

[21] H. Rami, L. Belmerhnia, A. D. E. Maliani, and M. E. Hassouni, "Texture retrieval using mixtures of generalized gaussian distribution and cauchy–schwarz divergence in wavelet domain," *Signal Processing: Image Communication*, vol. 42, pp. 45 – 58, 2016.

[22] S. Sakji-Nsibi and A. Benazza-Benyahia, "Copula-based statistical models for multicomponent image retrieval in thewavelet transform domain," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 253–256, Nov 2009.

[23] R. Kwitt, P. Meerwald, and A. Uhl, "Efficient texture image retrieval using copulas in a

bayesian framework," *IEEE Transactions on Image Processing*, vol. 20, pp. 2063–2077, July 2011.

[24] N. E. Lasmar and Y. Berthoumieu, "Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms," *IEEE Transactions on Image Processing*, vol. 23, pp. 2246–2261, May 2014.

[25] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 21, pp. 2874–2886, May 2012.

[26] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural texture similarity metrics for image analysis and retrieval," *IEEE Transactions on Image Processing*, vol. 22, pp. 2545–2558, July 2013.

[27] K. G. Krishnan, V. P. T., and R. Abinaya, "Performance analysis of texture classification techniques using shearlet transform," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1408–1412, March 2016.

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

[29] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186–1191, 1993.

[30] X.-Y. Jiang and R.-C. Zhao, "Texture segmentation based on incomplete wavelet packet frame," in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, vol. 5, pp. 3172–3177 Vol.5, Nov 2003.

[31] 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 *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. iii–597–600 vol.3, May 2004.

[32] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture discrimination using multimodal wavelet packet subbands," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 3, pp. 1493–1496 Vol. 3, Oct 2004.

[33] K. Huang and S. Aviyente, "Information-theoretic wavelet packet subband selection for texture classification," *Signal Processing*, vol. 86, no. 7, pp. 1410 – 1420, 2006.

[34] S. C. Kim and T. J. Kang, "Texture classification and segmentation using wavelet packet frame and gaussian mixture model," *Pattern Recognition*, vol. 40, no. 4, pp. 1207 – 1221, 2007.

[35] M. Vetterli and J. Kovacevic, *Wavelet and Subband Coding*. Englewood Cliffs, NY: Prentice-Hall, 1995.

[36] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, third ed., 2009.

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

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

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

[40] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[41] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and their Applications* (B. Ruskai, ed.), pp. 153–178, Jones and Barlettt, 1992.

[42] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs:NY Prentice-Hall, 1993.

[43] R. Gray and L. D. Davisson, *Introduction to Statistical Signal Processing*. Cambridge Univ Press, 2004.

[44] L. Breiman, *Probability*. Addison-Wesley, 1968.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Interscience, New York, 1991.

[46] C. Scott, "Tree pruning with subadditive penalties," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4518–4525, 2005.

[47] G. Lugosi and K. Zeger, "Concept learning using complexity regularization," *IEEE Transactions on Information Theory*, vol. 42, pp. 48–54, Jan 1996.

[48] C. Scott and R. D. Nowak, "Minimax-optimal classification with dyadic decision trees," *IEEE Transactions on Information Theory*, vol. 52, pp. 1335–1353, April 2006.

[49] J. F. Silva and S. S. Narayanan, "On signal representations within the bayes decision framework," *Pattern Recognition*, vol. 45, pp. 1853–1865, May 2012.

[50] P. Chou, T. Lookabaugh, and R. Gray, "Optimal pruning with applications to tree-structure source coding and modeling," *IEEE Transactions on Information Theory*, vol. 35, no. 2, pp. 299–315, 1989.

[51] R. Pickard, C. Graszyk, S. Mann, J. Wachman, L. Pickard, and L. Campbell, "Vistex database," *Media Lab., MIT, Cambridge, Massachusetts*, 1995.

[52] P. Brodatz, *Textures: a photographic album for artists and designers*, vol. 66. Dover New York, 1966.

[53] R. Kwitt and P. Meerwald, "Salzburg texture image database," 2012.

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

[55] R. Kwitt and A. Uhl, "Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval," in *2008 15th IEEE International Conference on Image Processing*, pp. 933–936, Oct 2008.