Universidad Nacional Autónoma de México
Posgrado en Ciencia e Ingeniería de la Computación

Overcomplete Image Representations for Texture Analysis

TESIS

que para optar por el grado de:
Doctor en Ingeniería
(Computación)

PRESENTA

Uriel Rodrigo Nava Velazco

TUTOR
Dr. Boris Escalante Ramírez
Facultad de Ingeniería

COMITÉ TUTORAL
Dr. Gabriel Cristóbal Pérez
Consejo Superior de Investigaciones Científicas
Dr. Fernando Arámbula Cosío
Centro de Ciencias Aplicadas y Desarrollo Tecnológico

México D.F. 2013

OVERCOMPLETE IMAGE REPRESENTATIONS FOR TEXTURE ANALYSIS

by
URIEL RODRIGO NAVA VELAZCO

A DISSERTATION SUBMITTED
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF ENGINEERING

UNIVERSIDAD NACIONAL AUTÓNOMA
DE
MÉXICO

Advisor
DR. BORIS ESCALANTE RAMÍREZ

Supervising Committee
DR. GABRIEL CRISTÓBAL PÉREZ
DR. FERNANDO ARÁMBULA COSÍO

MEXICO CITY 2013

– A la memoria de Nina que al irse se llevó la mitad de mi corazón.

– A mi mamá, quien tiene la otra mitad.

ANÁLISIS DE TEXTURAS
MEDIANTE
REPRESENTACIONES SOBRECOMPLETAS

RESUMEN

En años recientes el procesamiento de imágenes y la visión por computadora han jugado un papel importante en diversas áreas tecnológicas y científicas debido principalmente a que la sociedad moderna destaca la visión sobre el resto de los sentidos. A lo largo del tiempo los requerimientos de las aplicaciones y la complejidad de los problemas han ido en aumento, por lo que cada vez es más difícil proponer un modelo general, ya que en muchos casos la solución dependerá de las características intrínsecas del problema. En paralelo, el avance en el conocimiento del sistema visual humano ha permitido proponer modelos de representación de imágenes más sofisticados. Dichos modelos incorporan fenómenos simples que ocurren en las primeras etapas del sistema visual y adaptan los estímulos visuales para un procesamiento posterior. La presente tesis tiene como objetivo investigar características de la visión como la sobrerrepresentación y la orientación de los campos receptivos para proponer nuevos algoritmos de inspiración biológica. A partir de los trabajos de Gabor, Daugman y Hubel proponemos un modelo sobrecompleto de representación de imágenes y lo comparamos con diversos modelos presentes en el estado del arte. Nuestro modelo basado en funciones de Gabor optimiza la redundancia de la información y la distribuye de manera uniforme en diferentes bandas de frecuencia y en orientaciones, con resultados prometedores. Es bien conocido que los modelos de Gabor generan representaciones multidimensionales por lo que proponemos una etapa de reducción de la dimensión basada en la teoría de Fisher y los métodos kernel, lo que conduce a una mejor caracterización de la escena visual. Para validar nuestra propuesta hemos realizado experimentos de segmentación y clasificación de texturas sintéticas, además, incluimos un caso de estudio sobre clasificación de imágenes médicas. En la última parte de la tesis sostenemos la hipótesis de que la combinación de descriptores globales y locales mejora la capacidad discriminante. Presentamos un estudio sobre descriptores locales, en específico sobre patrones binarios, y proponemos un esquema combinado para clasificar texturas. Los resultados de los experimentos son consistentes con la teoría y demuestran la efectividad de nuestra propuesta.

OVERCOMPLETE IMAGE REPRESENTATIONS
FOR
TEXTURE ANALYSIS

ABSTRACT

In recent years, image processing and computer vision have played an important role in many scientific and technological areas mainly because modern society highlights vision over other senses. Throughout time application requirements and complexity have been increasing. Due to the fact that in many cases solutions depend on intrinsic characteristics of problems it is difficult to propose a universal model. In parallel, advances in understanding the human visual system have allowed to use sophisticated image models that incorporate simple phenomena, which occur in early stages of the visual system. Such phenomena suit *visual stimuli* for further processing. This thesis aims to investigate characteristics of vision such as over-representation and orientation of receptive fields in order to propose bio-inspired image models. Starting from studies of Gabor, Daugman, and Hubel we present an overcomplete image model that takes advantage of redundant information. Furthermore, we performed a comparison with several models from the state-of-art. Our proposal is based on Gabor filters and optimizes redundant information; such an information is distributed uniformly onto frequency bands and orientations with promising results. It is well known that Gabor models generated high-dimensional representations, therefore, we included a step where data dimension is reduced using Fisher theory and kernel methods. This step leads to a better characterization of visual scenes. In order to validate our method, we performed several experiments of segmentation and classification using synthetic textures. The last part of this dissertation claims that the combination of global and local descriptors will provide robust features that lead to an improvement in the classification rate. We included a study of local descriptors, specifically based on local binary patterns, and introduced a combined scheme for classifying textures of lung emphysema. The results of the experiments are consistent with the theory and demonstrate the effectiveness of our proposal.

Contents

Page

List of Figures

List of Tables

List of Algorithms

Page

– Research is what I'm doing when I don't know what I'm doing.

Wernher von Braun

1

Introduction

1.1 Motivation

IT has been over 40 years since Martin Minsky from MIT linked a camera to a computer in order to achieve that "computers understand images and videos." [4] At that time there was great excitement about the use of algorithms to process digital data and retrieve three-dimensional structures from still images and use them to understand scenes.

The first steps in image analysis research consisted in developing interfaces between the visual world and computers. However, the visual input stage was regarded by researchers just as an easy issue prior to solve more difficult problems such as higher-level reasoning and planning. During this period, Rosenfeld and Pfaltz [5] proposed several algorithms for processing digitized pictorial information, such algorithms describe techniques for manipulation of intensity data defined on uniform grids. Early advances and efforts were encouraged partially by the space race. One of the first applications reported in the literature was geometrical rectification of digital images obtained from the first Earth Resources Technology Satellite launched by the United States National Aeronautics and Space Administration in 1972 [6].

However, digital image processing is hardly new; first applications date back to 1920 where newspapers were digitized and sent by a submarine cable between London and New York. Since then, research and development in visual information are moving forward at an extremely fast pace. We can say that image analysis is a research area that integrates several disciplines such as mathematics, optics, electronics, among others.

A good example of this convergence and an important contribution that speeded up the progress of image analysis was the development of multiresolution analysis and wavelet theory. In the late 70's Jean Morlet [7, 8] presented an alternative to the Short-time Fourier transform; the procedure consisted in applying a window to a signal and then compute the Fourier coefficients. This operation depends on two parameters: the location and the frequency of the window. In this direction, the French mathematical school has made significant contributions; Alex Grossmann [9] worked on a concept where a signal could be decomposed into wavelets and then back; he proposed an inverse Morlet transform. Later in 1985, Yves Meyer [10] applied in harmonic analysis both transforms in order to study certain singularities in the Hilbert space domain. Meyer's work led to the construction of a wavelet basis with excellent time-frequency properties. In 1986, Stépane Mallat [11] conceived a structure in multiple scales divided into frequency bands. Such a structure allows to analyze large image features using large windows whereas fine features are analyzed with small windows in the spatial domain. This model was inspired by the Laplacian pyramid originally proposed by Burt and Adelson [12]. Donoho et al. [13] extended this work by proposing complex wavelets.

Since the beginning, image analysis has been interconnected with the Human Visual System (HVS). A proof of this relationship is the fact that psychophysical and physiological experiments have shown that multi-scale transform, which is a mathematical tool, seems to appear in the visual cortex of mammals. Additionally, in many applications the final user is a human being; in fact, early attempts to understand the HVS were based on extracting edges [14].

One of the primary goals of low-level vision is to extract geometric information from visual scenes. This is possible because of two operations: **edge detection**, which identifies locations where image intensities change abruptly, John Canny [15] proposed an edge detector based on Gaussian derivatives in order to take advantage of some properties of the HVS; and **stereopsis**, which synthesizes a depth map from two or more images. Perhaps the most important book that summarizes the first advances in the area is **Vision** by David Marr [16] where vision is treated as an information processing system with three levels of analysis.

Due to the large amount of data that the HVS must process, it is reasonable to think that the visual cortex of mammals has evolved over the years and uses efficient coding strategies. Barlow [17] established the idea that the goal of photoreceptors in the retina is to transform visual input in such a manner that the HVS takes advantage of the redundancy [18]. Redundant or overcomplete representations are interesting in the sense that their inherent redundancy can be exploited to increase the robustness of the visual representation.

So far, overcomplete schemes based on Gabor filters have been useful in many image analysis applications. A Gabor filter in spatial domain is a Gaussian function multiplied by a complex exponential; its shape is Gaussian in both spatial and frequency domains. The filter is closely localized in space and frequency. Furthermore, simple cells of the visual cortex of mammals are best modeled as a family of self-similar 2D Gabor filters.

Nevertheless, Gabor filters present some drawbacks that can be summarized as follows: (i) the filter averaging is not null, therefore the DC component influences intermediate bands. Gabor filters overlap more at lower frequencies than at higher ones yielding a non-uniform coverage of the Fourier domain; (ii) the traditional arrangement in scales and orientations of Gabor filters does not cover uniformly the Fourier plane leading to an inaccurate reconstruction. In fact, they are band-pass filters by definition so that they cannot cover lowest and highest frequencies; and (iii) it is not possible to build a complete orthogonal basis of Gabor functions. Non-orthogonality implies that exact reconstruction using the same filters for analysis and synthesis will not be possible unless an overcomplete dictionary is considered. Despite previous drawbacks, Gabor-based techniques are considered the state-of-art in texture characterization, which is one of the most difficult problems in image analysis.

In this dissertation, new contributions are reported in the following six chapters. Initial work was focused on an image model derived from the visual cortex of mammals. In order to build such a model Daugman and Field's studies were considered and the filter bank parameters were optimized. Additionally, even scales were rotated by a constant factor consisting of the half a distance between filter centers. These modifications resulted in a logarithmic model that separates and distributes frequency bands in a better way. Furthermore, this approach reduces redundancy. Since Gabor filters have been successfully used in problems involving textures, we included a study of local texture operators. Also, the problem of rotation invariance was addressed. An analysis about the influence of local contrast was performed as well. This dissertation uses higher order moments to characterize the response of Gabor filters, we conducted several experiments to assess their performance with different versions of Gabor filters. The final part of this dissertation proposes a better characterization of textures by combining global and local descriptors. To validate the hypothesis experiments of segmentation and classification were performed on both synthetic textures and medical images.

1.2 Organization and outline

This dissertation is organized as follows: in Chapter 2 an introduction about the visual cortex is given; we reviewed essential concepts of bases and frames and introduced basic definitions required to build our proposal. In Chapter 3 the classical Gabor paradigm is presented; we also analyzed fundamental results involved in the development of Gabor functions and its limitations. A major improvement is shown, we computed optimal filters in order to match simple psychophysical aspects of the HVS. We used a novel scheme based on log-Gabor filters. In Chapter 4 we investigated the advantages of log-Gabor filters by comparing their performance with regular Gabor filters. In Chapter 5 a detailed overview of local binary patters is given, we analyzed the most significant extensions that have been written in the literature from theoretical to a practical perspective, whereas Chapter 6 is dedicated to classification of emphysematous patterns. Finally in Chapter 7 the conclusions and further research are drawn.

– A scientist is not a person who gives the right answers,
he's one who asks the right questions.

Claude Lévi-Strauss

2

From biology to linear algebra

2.1 Human visual system

THE HVS must process a huge amount of spatio-temporal information in real-time. Therefore, it has developed coding strategies to represent natural images. As a result, we are able to recognize and categorize thousands of objects and compute millions of data every day without any effort. This task is possible because the HVS focuses on important aspects of the visual field that are used as low-level visual features in a cognitive process based on experience. Although the optical system of the eye is complex, the mechanism of image formation can be simplified as follows: *visual stimuli* pass through the optics of the eye as inverted images until they reach the bottom of the retina. Curiously, the incoming stimuli must go through the bulk of the retina's neural apparatus to reach photoreceptors. This fact, according to evolutionists, causes a degradation of the formed image. This arrangement of the retina is said to be inverted [19] because visual cells are oriented so that their sensory ends are in opposite position to the incident light, see Fig. 2.1.

Two types of photoreceptors can be distinguished in the retina. On one hand, **rods** are responsible for monochromatic vision under low light conditions (scotopic conditions). They employ a sensitive photopigment called *rhodopsin*, which perceives and absorbs the green-blue light [1]. Around 120 millions of rods are concentrated in the peripheral region of the retina, thus, most people will find that they can see better if they focus their gaze just off to the point of view [20].

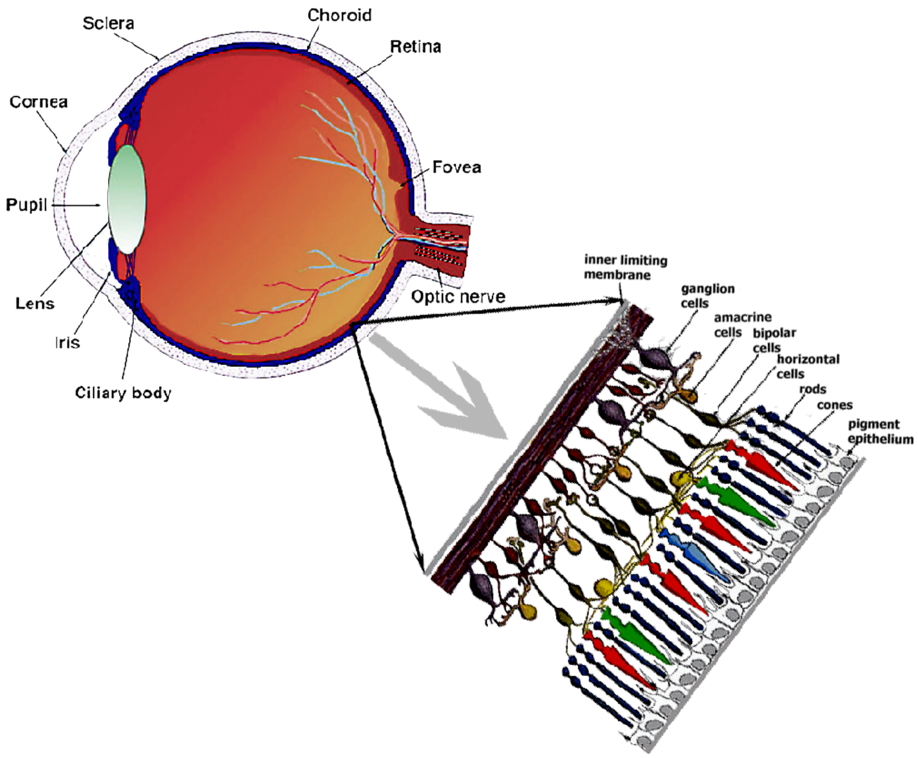


Figure 2.1: *A schematic section through the human eye with an enlargement of the retina; the gray arrow indicates the light direction. This picture was modified with permission of Prof. Dr. Helga Kolb [1].*

On the other hand, **cones** are in charge of photopic vision and high acuity tasks like reading and target detection –this may be thought as the ability of the eye to see fine details. Cones are densely concentrated in the fovea (around 4.5 millions of cones are in the central region of the retina) but rapidly reduce in number towards the periphery. They differ from rods in shape because they are shorter with a broad base and bulb-shaped. Despite the fact that most mammals are dichromatic, primates and humans are trichromatic because they have three types of cone cells according to their visual pigments [21]. In combination, three types of cones enable us to perceive color. Thus, while the visual resolution is superior with cones, rods are better for motion estimation.

The information from photoreceptors is processed by Retinal Ganglion Cells (RGC) via two intermediate neurons: bipolar and amacrine cells. The latter have diverse morphologies; up to 40 different types of amacrine cells allow us to adjust sensitivity or contrast changes [22]. RGC gather information and send their output toward the Lateral Geniculate Nucleus (LGN) through the optic nerve using synaptic connections. LGN is inside the thalamus and is the principal structure that connects the eye to the back of the brain where the primary visual cortex (V1) is located. However, LGN not only connects axons of RGC to V1 but it is a relay structure with feedback connections, which can be both excitatory or inhibitory. In fact, it was discovered that LGN may perform certain normalization of visual stimuli and is involved in visual quality assessment [23].

V1 or striate cortex is a layer of 2mm thick that encompass cells and is responsible for creating a representation of the space. Each cell has a receptive field, which is a region where the presence or absence of a stimulus will cause cellular activation.

According to Hubel and Wiesel [20, 24], there are three types of cells: **(i)** simple cells are known to be line detectors, they have narrow and elongated excitatory and inhibitory zones divided by a straight line. These cells are selective to orientation and spatial frequency of a stimulus; **(ii)** complex cells are the most common cells in the striate cortex. They have larger receptive fields than simple cells without clear excitatory or inhibitory zones. Many complex cells respond best to moving edges with a specific orientation and direction. These neurons are powerful motion detectors; and **(iii)** hypercomplex cells have very large receptive fields that may combine complex cells' signals. These types of cells respond to a specific orientation but are "end-stopped", namely the cell is inhibited if the oriented stimulus extends beyond a specific part of the corresponding receptive field. These neurons are powerful corner detectors.

V1 encodes information in terms of local contrast and is organized into orientation columns. This configuration allows to detect edges of objects in the visual world. Furthermore, V1 is connected to many other regions such as V2, which is responsible for the phenomenon of color constancy, while V5/MT plays an important role in motion processing. Nevertheless, these topics are out of scope of this dissertation, for further details see [25].

So far, there has been progress in understanding the HVS but there is still much work to be done. Researchers have claimed that we have understood only a small portion of the primary visual cortex of mammals [19, 26–28]. The majority of current theories about the HVS are based on responses to simple stimuli like checkerboards, sinusoidal functions, edges, and random patterns. These theories have been extended to natural environments without success due to natural image statistics vary from one scene to another. In fact, the original idea that the primary visual cortex performs global Fourier transforms has been ruled out because the bandwidths and the local spatial properties of cortical neurons restrict an adequate extraction of Fourier coefficients. Furthermore, the receptive fields of some types of ganglion cells cover visual space repeatedly; such an overlap does not guarantee that different cell types encode entirely independent information. Therefore, natural stimuli cannot be fully characterized by their spatial power spectrum [29]. The inference to be drawn from this fact is that we need to understand the process of vision in terms of **redundancy** (overcompleteness).

The formulation of a mathematical theory of the visual cortex should have many advantages, for instance, we can quantify the spatial computations performed by each cell. The first strategy was based on sending pulses to the retina for recording ganglion cell responses. This model is known as Population Coding (PC) [30] that is implemented by an arrangement of neurons where each one of them responds to a set of inputs $I(x)$. Simultaneously, the outputs of some neurons are combined to determine the information at a_i. This model takes advantage of information redundancy for improving the decision making process, see Fig. 2.2.

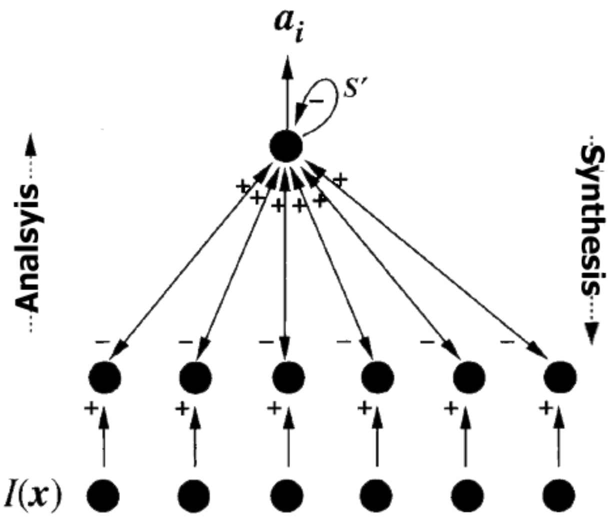


Figure 2.2: *Population coding model. a_i makes inferences about visual information by the analogous process of image analysis-synthesis.*

Nevertheless, one disadvantage of PC is that redundancy generates complex statistic dependence among neurons of the input stream. Consequently, one goal for researchers is to reduce that dependence, so that images can be approximated in terms of a collection of statistically independent events.

In 1961, Horace Barlow [17] discovered that one of the main goals of visual processing is, indeed, the reduction of the redundancy (sparseness). In addition, he found that if the brightness of the neighboring points in an image are similar, then the retina reduces the redundant information. Therefore, two questions arise: how large is the set of receptive fields that best describe all natural scenes (how overcomplete) and how many of them are active in a single scene (how sparse)?

We have learned that computations to enhance features of a visual scene can be made easier if we employ an orthogonal basis, where certain characteristics become obvious in the transform domain; allowing each element in the space to be written as a linear combination. However, the conditions to a basis are very restrictive and the representation is non-redundant. In such a case, corruption or loss of transform coefficients can be fatal. The redundant counterpart of orthogonal bases are called frames and since they are less constrained than bases, they are used when we need more flexibility. The main difference between bases and frames is that the former is a unique representation and uses a number of vectors equal to the dimension of the space. When this number is larger, we can still have a representative set of vectors, except that the vectors are no longer linearly independent and the resulting set is no longer called a basis but a frame. Intuitively, we can think about a frame as a basis with at least one element added.

In the following Section 2.2 we present the essential concepts of bases and frames from a Hilbert theory approach (named after David Hilbert).

2.2 Hilbert spaces

Since objects in nature may appear in a very large range of orientations and scales, vision systems must extract them from any background. Hence, space characterization is the first step to describe any object. Commonly, such a space is based on the Hilbert theory. The two basic concepts of linear algebra in the Hilbert theory are vector space and inner product. These concepts induce a norm, thus, a Hilbert space is a normed space.

In signal theory, a classical problem is to represent a signal as a convergent series of elementary functions of the same structure. This problem is generally addressed with linear combinations and concepts of bases. Initially, we assume that objects are vectors in a finite-dimensional vector space in $\mathbb{R}^n$ and $\mathbb{C}^n$. However, there are only two basic operations defined in a vector space: addition of two vectors and multiplication by a scalar. Formally, we can say:

Definition 2.1 By a vector space we mean a non-empty set $\mathcal{V}$ with two operations:

1 a mapping $(x, y) \to x + y$ from $\mathcal{V} \times \mathcal{V}$ into $\mathcal{V}$ called *addition*.

2 a mapping $(\lambda, x) \to \lambda x$ from $\mathbb{C} \times \mathcal{V}$ into $\mathcal{V}$ called *multiplication by a scalar*.

We must define operations that can measure the elements within the vector space. If the concept of norm between two vectors, defined as the absolute value of the difference between them is introduced, then a vector space becomes a metric space. This progression allows us to measure similarities by finding distances among vectors. Moreover, if a metric space is considered as a closed set, namely a representative subset of vectors that can describe any vector in the same space then the space becomes a **Hilbert space** denoted by $\mathcal{H}$, here, the norm is defined by an inner product.

However, we need more elements in order to characterize vectors in a common reference mark. Although bases are not limited to Hilbert spaces, we can use them to represent vectors in $\mathcal{H}$.

A basis is a set of linearly independent vectors used to represent every vector in a given vector space, formally we say:

Definition 2.2 A finite subset $\Phi = \{\phi_i\}_{i \in \mathbb{N}^+}$ of a vector space $\mathcal{V}$ is a basis of $\mathcal{V}$ if:

1 the vectors in Φ are linearly independent.

2 $\mathcal{V} = \operatorname{span}(\Phi)$.

This definition is very useful because we can describe an object using its approximations. For instance, an approximation of a natural image can be a low-pass version of it, which is a blurred version of the original one. Typically, a basis is used to highlight salient features such as edges, patterns, and textures; this operation often facilitates subsequent processing tasks.

Elements of the space $\mathcal{V}$ are generated by the subset Φ using linear combinations. Furthermore, we can say that the dimension of the space $\mathcal{V}$, $\dim(\mathcal{V})$, is n if $i = \{1, \ldots, n\}$.

We considered finite-dimensional vector spaces and sequences of finite energy or sequences of finite square sums, x, where $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$ are in general complex values. For such spaces, the inner product between two vectors x and y is defined as:

$$\langle x, y \rangle = \sum_{i \in J} \boldsymbol{W}_i x_i y_i^* \tag{2.1}$$

where * denotes the complex conjugate, J is some index set, and $\boldsymbol{W}$ is a diagonal matrix. Here, we considered $\boldsymbol{W} = \boldsymbol{I}$.

Eq. (2.1) associates each pair of vectors with a scalar quantity. The inclusion of a distance allows intuitive geometrical notions of length of a vector and the angle between two vectors.

Also the concept of an inner product induces an associated norm, as well. In general, the idea of a norm is an abstract generalization of the length of a vector in a vector space as follows:

Definition 2.3 A function $\|\bullet\|$ on a vector space $\mathcal{V}$ is called a norm if:

1 $\|x\| = 0 \iff x = 0$;

2 $\|\lambda x\| = |\lambda| \, \|x\| \; \forall x \in \mathcal{V}$ and $\lambda \in \mathbb{C}$;

3 $\|x + y\| \leq \|x\| + \|y\| \; \forall x, y \in \mathcal{V}$.

Norm in a vector space is a generalization of the distance between two points in $\mathbb{R}^2$. In Hilbert spaces, a norm is defined using Eq. (2.1) as follows:

$$\begin{aligned} \|x\| &= \sqrt{\langle x, x \rangle} \\ &= \sqrt{\sum_{i \in J} |x_i|^2} \end{aligned} \tag{2.2}$$

2.3 Orthonormal bases

Bases play a prominent role in the analysis of vector spaces. They are used in both finite-dimensional cases and infinite-dimensional cases.

Definition 2.4 A basis is called an orthonormal basis if:

$$\left\langle \phi_i, \phi_j \right\rangle = \delta_{i-j} \tag{2.3}$$

where $\delta_k = 1$ for $k = 0$, otherwise $\delta_k = 0$.

It is possible to represent the signal of interest through an orthogonal projection by using a subset of basis vectors as follows:

$$x = \sum_{i \in I} \langle x, \phi_i \rangle \phi_i \tag{2.4}$$

this means that the representation of the signal x is a sum of projections onto a space $\mathcal{V}$.

The terms $C_i = \langle x, \phi_i \rangle$ are the coefficients in the expansion of x over the base Φ. The consequence of this definition is that all $x \in \mathbb{V}$ have a unique representation in terms of the elements of the base.

Considering the previous Eq. (2.4), we can state that, indeed, there is a set of unique coefficients, $\{C_i\}_{i=1}^{n}$, that can be used to define the best approximation of x as follows:

Definition 2.5 Let $\hat{x}^{(k)}$ be the best approximation of x using the orthogonal set $\{\phi_0, \phi_1, \ldots \phi_{k-1}\}$ and $\hat{x}^{(0)} = 0$ then:

$$\hat{x}^{(k+1)} = \hat{x}^{(k)} + \langle x, \phi_{k+1} \rangle \phi_{k+1} \tag{2.5}$$

This means that the new best approximation is the sum of the previous best approximations. However, such representations are not redundant or are not based on the PC model. Therefore, problems of corruption or loss of information in the transform coefficients may appear. In [31] these limitations were addressed. One way to avoid them is with frames, originally proposed in 1952 by Duffin and Schaeffer [32] in order to reconstruct band-limited signals from irregularly spaced samples.

2.4 Frames

It took almost 50 years to the frame theory gain attraction from researchers. It is well known that the use of redundancy in engineering systems improves robustness and numerical stability. Motivated by this observation, redundant signal expansions have found widespread use in many different engineering disciplines. The main promoters to this effort are Daubechies, Grossman, and Meyer [31, 33].

The idea behind this theory relies on the fact that redundancy is generated with the expansion of the signal of interest. This approach has been used in coding [34], denoising [35], and restoration and enhancement [36].

The notion of bases in finite-dimensional spaces implies that the number of representative vectors is the same as the dimension of the space. For instance, the number of vectors in $\mathbb{R}^2$ is 2, whereas in $\mathbb{R}^3$ is 3, and so on. When this number is larger than the dimension of the space, the set is no longer linearly independent and the result is no longer a basis.

In the following paragraphs we will start from the scratch using a simple orthonormal basis in order to illustrate how to build a frame.

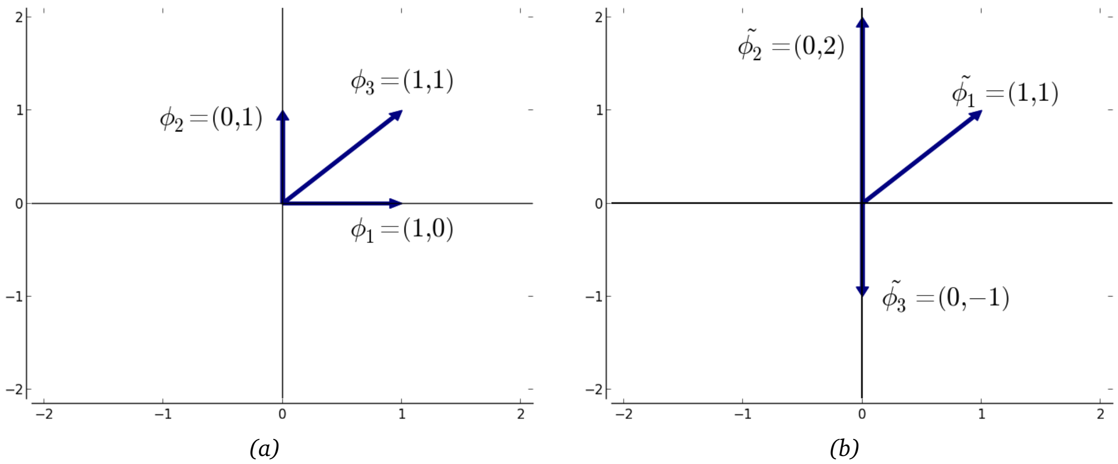


Figure 2.3: *Biorthogonal frames in* $\mathbb{R}^2$. ***(a)*** *Frame* $\Phi = \{\phi_1, \phi_2, \phi_3\}$ *and* ***(b)*** *its dual frame* $\tilde{\Phi} = \left\{\tilde{\phi}_1, \tilde{\phi}_2, \tilde{\phi}_3\right\}$.

Let $\Phi = \{\phi_1, \phi_2\} = \{(1, 0), (0, 1)\}$ be an orthonormal basis in $\mathbb{R}^2$. Then, using the Eq. (2.4), we can state that the following projection of x is true:

$$x = \langle x, \phi_1 \rangle \phi_1 + \langle x, \phi_2 \rangle \phi_2 \tag{2.6}$$

But what happen if a third vector (e.g., $\phi_3 = \phi_1 + \phi_2$) is added to the basis Φ? The answer is that the property of linear independence is lost. However, these vectors must be able to represent every vector in $\mathbb{R}^2$ since their subset is able to do so.

In order to build a frame, it is possible to add a zero vector to Eq. (2.6) as follows:

$$x = \langle x, \phi_1 \rangle \phi_1 + \langle x, \phi_2 \rangle \phi_2 + \underbrace{\left(\langle x, \phi_2 \rangle - \langle x, \phi_2 \rangle\right)(\phi_1 - \phi_2)}_{0} \tag{2.7}$$

The previous expression was rearranged, such that:

$$\begin{aligned} x &= \underbrace{\langle x, \phi_1 \rangle \phi_1}_{a} + \underbrace{\langle x, \phi_2 \rangle \phi_2}_{b} + \underbrace{\langle x, \phi_2 \rangle \phi_1}_{a} - \underbrace{\langle x, \phi_2 \rangle \phi_2}_{c} - \underbrace{\langle x, \phi_2 \rangle \phi_1}_{c} + \underbrace{\langle x, \phi_2 \rangle \phi_2}_{b} \\ &= \underbrace{\langle x, \phi_1 + \phi_2 \rangle \phi_1}_{a} + \underbrace{\langle x, 2\phi_2 \rangle \phi_2}_{b} + \underbrace{\langle x, -\phi_2 \rangle (\phi_1 + \phi_2)}_{c} \end{aligned} \tag{2.8}$$

the factor $(\phi_1 + \phi_2)$ can be recognized as ϕ_3. In addition, we defined the next three variables as follows:

$$\begin{aligned} \tilde{\phi}_1 &= \phi_1 + \phi_2 \\ \tilde{\phi}_2 &= 2\phi_2 \\ \tilde{\phi}_3 &= -\phi_2 \end{aligned} \tag{2.9}$$

If we replace Eq. (2.9) into Eq. (2.8), thus, the expansion can be rewritten as:

$$\begin{aligned} x &= \left\langle x, \tilde{\phi_1} \right\rangle \phi_1 + \left\langle x, \tilde{\phi_2} \right\rangle \phi_2 + \left\langle x, \tilde{\phi_3} \right\rangle \phi_3 \\ &= \sum_{i=1}^{3} \left\langle x, \tilde{\phi_i} \right\rangle \phi_i \end{aligned} \tag{2.10}$$

Herein, it is possible to use matrix notation: $\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ and $\boldsymbol{N} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & -1 \end{pmatrix}$ in order to rewrite the Eq. (2.10) as:

$$\begin{aligned} x &= \sum_{i=1}^{3} \left\langle x, \tilde{\phi_i} \right\rangle \phi_i \\ &= \boldsymbol{M}\boldsymbol{N}^* x \end{aligned} \tag{2.11}$$

The redundant set of vectors $\boldsymbol{M} = \{\phi_i\}_{i=1,2,3}$ is called **frame**, see Fig. 2.3(a), whereas the set $\boldsymbol{N} = \left\{\tilde{\phi_i}\right\}_{i=1,2,3}$ is known as a **dual-frame**, see Fig. 2.3(b). Moreover, $\boldsymbol{M}$ is orthogonal to $\boldsymbol{N}$, thus, ϕ_i and $\tilde{\phi_i}$ are interchangeable, namely that the filter used for analysis is the same filter used in synthesis: $\boldsymbol{M}\boldsymbol{N}^* x = \boldsymbol{N}\boldsymbol{M}^* x$.

This relationship of biorthogonality can be represented as:

$$\left\langle \tilde{\phi_i}, \phi_j \right\rangle = \delta_{i-j} \tag{2.12}$$

however, not necessarily such frames preserve the norm [37].

2.4.1 MERCEDES-BENZ FRAME

Consider $\boldsymbol{\Phi} = \{\phi_1, \phi_2, \phi_3\}$ in matrix form: $\boldsymbol{\Phi} = \begin{pmatrix} 0 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} \\ 11 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$ with its corresponding expansion:

$$\begin{aligned} x &= \sum_{i=1}^{3} \langle x, \phi_i \rangle \phi_i \\ &= \boldsymbol{\Phi}\boldsymbol{\Phi}^* x \end{aligned} \tag{2.13}$$

which is one of the most popular frames. It is known as Mercedes-Benz Frame (MBF), the reason for this naming becomes evident in Fig. 2.4. It was originally proposed by Peres-Wooters [2] in quantum information theory and is a collection of three vectors in $\mathbb{R}^2$.

Since Eq. (2.13) is used for analysis and reconstruction, Φ is known as a dual-frame –at first sight, you can think that represents a generalization of an orthogonal basis except that the vectors are not linearly independent. Frames of this type are called **tight frames**.

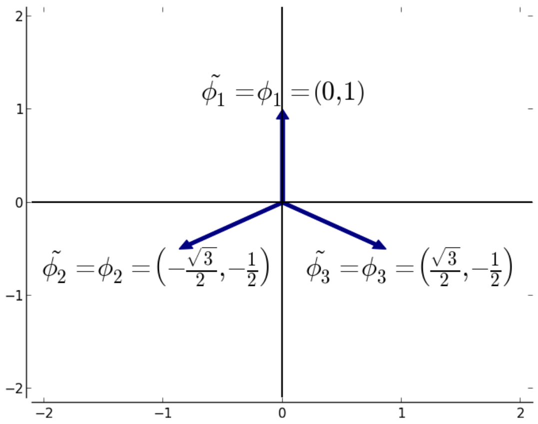


Figure 2.4: *The unit-norm MBF is an example of a dual-frame. This frame can be generated by rotations with steps of $\frac{2\pi}{3}$. MBF is used mainly in image restoration problems [2].*

In order to preserve the normalization, $\mathbf{\Phi\Phi}^* = \boldsymbol{I}$, the factor $\frac{2}{3}$ is introduced

$$\mathbf{\Phi\Phi}^* = \begin{pmatrix} 0 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{3}{2}} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -\sqrt{\frac{3}{2}} & -\frac{1}{2} \\ \sqrt{\frac{3}{2}} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix} \tag{2.14}$$

So that the unit-norm expansion is computed as:

$$\begin{aligned} x &= \frac{2}{3} \sum_{i=1}^{3} \langle x, \phi_i \rangle \phi_i \\ &= \frac{2}{3} \Phi\Phi^* x \end{aligned} \tag{2.15}$$

Let us now calculate the norm of the coefficients C_i using Eq. (2.2) as follows:

$$\begin{aligned} ||C||^2 &= \sum_{i=1}^{3} |\langle x, \phi_i \rangle| \\ &= \frac{3}{2} ||x||^2 \end{aligned} \tag{2.16}$$

note that the expansion equation of the MBF has a scaling factor of $\frac{2}{3}$.

If a frame is tight and all vectors have unit norm –as in this case– the inverse of the scaling factor represents the redundancy. In this case, about $\frac{3}{2}$ or 50% more vectors are needed to describe any vector in $\mathbb{R}^2$. This frame also minimizes the mean squared error [38].

2.4.2 TIGHT FRAMES

In Section 2.4.1 we introduced the concept of frames by examples. We now present the formal definition of a frame.

Definition 2.6 Let $\Phi = \{\phi_i\}_{i\in I}$ be a set in Hilbert space $\mathcal{H}$ then Φ is a frame if there exist two constants $0 < A \leq B < \infty$, such that for all x in $\mathcal{H}$:

$$A\,\|x\|^2 \leq \sum_{i\in I} |\langle x,\, \phi_i\rangle|^2 \leq B\,\|x\|^2 \tag{2.17}$$

where A and B are the frame boundaries.

Frame boundaries are related to stability. Kovačević [39] pointed out that in order to obtain a stable reconstruction, the transform coefficients C_i have to be bounded. In other words, reconstruction of any signal x is possible if and only if Eq. (2.17) is true and A and B are close.

When the condition $A = B$ is fulfilled, a frame becomes a **tight frame** and then we can rewrite Eq. (2.17) as:

$$\sum_{i\in I} |\langle x,\, \phi_i\rangle|^2 = A\,\|x\|^2 \tag{2.18}$$

Now, it is possible to place the factor A into the summation

$$\sum_{i\in I} \left|\left\langle \frac{1}{\sqrt{A}}x,\, \phi_i\right\rangle\right|^2 = \|x\|^2 \tag{2.19}$$

$\Phi = \left\{\left(\frac{1}{\sqrt{A}}\right)\phi_i\right\}_{i\in I}$ is called a “tight frame-1” and we can establish the following definition:

Definition 2.7 If $\Phi = \{\phi_i\}_{i\in\mathbb{N}^+}$ is a tight frame-A then the signal x can be represented as follows:

$$x = \frac{1}{A}\sum_{i\in I} \langle x,\, \phi_i\rangle\, \phi_i \tag{2.20}$$

Although Eq. (2.20) is similar to a basis expansion (except for the factor A) is not a “general basis” because its elements are not linearly independent (i.e., the mutual inner product is not zero). Therefore, the expansion is not unique.

– I learned very early the difference between knowing the name of something and knowing something.

Richard P. Feynman

3

Overcomplete image models

3.1 Introduction

In the previous chapter we have seen that a common approach to represent images is by linear superposition of basis functions. In fact, decomposition of an image into its structural components is a common task in image processing. State-of-the-art methods analyze local derivatives for distinguishing image variations due to shading or reflectance. Fourier analysis is the most common way to characterize images in the space-frequency domain using sine and cosine functions, however, this type of expansion is limited [40]. As an alternative, we used a more general and powerful methodology based on the so-called overcomplete methods (also called overcomplete dictionaries) [41–43].

Commonly, we say that such methods are biologically-inspired because the V1 area is strikingly overcomplete [44]. In other words, there are many more cells than are needed to represent the visual information. There are around 300 neurons per LGN in human beings and consequently such a configuration introduces redundancy.

A few years ago, it was believed that receptive fields of simple cells performed quasilinear transformations to the light that is perceived by the retina [45]. However, latter works have discovered that receptive fields also perform some nonlinear operations such as normalization of the contrast, rectification, and filtering to extract texture boundaries. In fact, there is evidence that quasilinear and nonlinear texture-boundary processes, each with an appropriate contrast normalization, may operate in parallel to provide a higher visual analysis.

Unlike the basis, the decomposition of an image using overcomplete methods is not unique, however, this may provide some advantages such as greater flexibility in approximations (for instance, an image can be decomposed onto multiple bases) and the increased stability of the representation namely less sensitive to noise. Another advantage is that we can explain high-dimensional data images in terms of a concise set of primitive features [46].

Many computational vision models have incorporated some biological properties of vision. They have shown the functioning of V1 by representing an image with an overcomplete set of visual neurons [47]. Thus, the motivation in this chapter is to investigate the advantages of biologically-inspired overcomplete models.

This chapter is organized as follows: in Section 3.2 we reviewed the classical Gabor paradigm, which is optimal in terms of the uncertainty principle. We discussed the fundamental results involved in the development of Gabor functions and its limitations. Here, a major improvement was proposed: we computed optimal filter parameters in order to match simple psychophysical aspects of the HVS. In Section 3.3 we present a novel scheme based on log-Gabor filters and summarized its properties. We also computed optimal parameters to match simple phenomena of the HVS in this approach. Finally, in Sections 3.4 a methodology for extracting features based on Gabor filters is presented.

3.2 Gabor filters

The work initiated by Hubel and Wiesel [20] shows that every simple cell is connected to its neighbors by the mechanism of inhibitory or excitatory responses that may occur at an *axoaxonal synapse*, which is a synapse between the axon of one cell and the axon of another, and responds to visual features located around a particular position with a specific frequency band and orientation. Such a spatial structure have a remarkable resemblance to Gabor filters, see Fig. 3.1. This fact suggests that simple receptive fields may provide the best possible simultaneous description of the spatial position and spectral content of visual stimuli as well.

In [48], Goesta Granlund proposed a general operator, which corresponds to a 2D Gabor filter, addressing the octave spacing of frequencies as well. John Daugman later showed a equivalence between the experimental findings on orientation selectivity of visual cortical neurons and the structure based on 2D Gabor functions. So, we can say that Gabor functions are a good choice for obtaining localized frequency information. Since their conjoint area has a lower limit analogous to the uncertainty principle proposed by Heisenberg, they offer the best simultaneous localization of spatial and frequency information. It means that the product of the uncertainties in frequency and time must be greater that a fixed constant; therefore the accuracy in one domain limits the best possible accuracy in the other one:

$$(\Delta x \Delta \omega) \geq \frac{1}{4\pi} \tag{3.1}$$

where Δx is the interval where the signal is not zero and $\Delta \omega$ is the bandwidth.

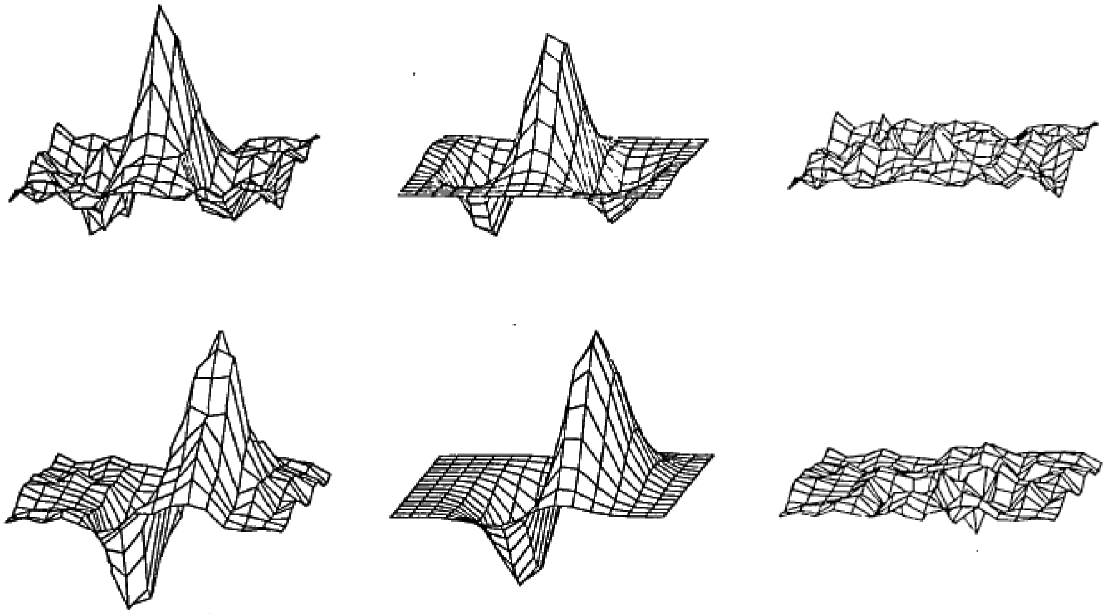


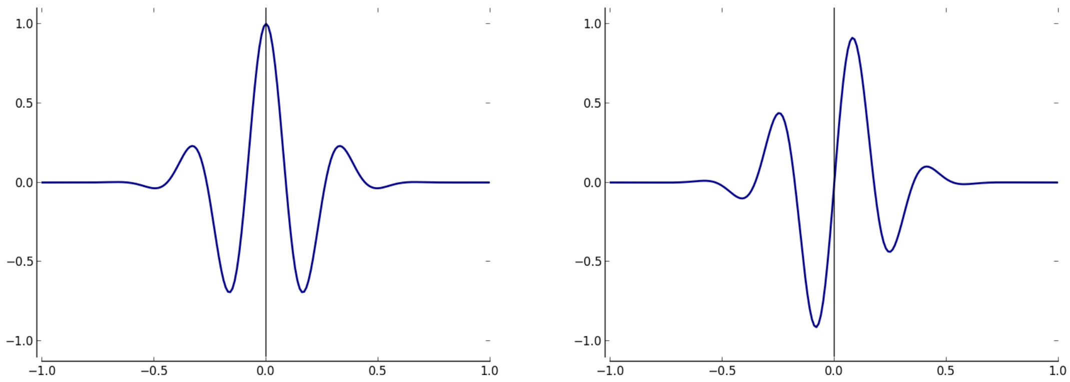
Figure 3.1: *Receptive field profiles of simple cells in a cat visual cortex. A comparison between receptive field profiles* ***(first column)*** *and Gabor filters* ***(second column)****. Residual values obtained by subtracting the best-fitting 2D Gabor filter and a measured profile* ***(third column)****. In 97% of cell studies the residual values can be neglected. Images were borrowed and modified from [3].*

John Daugman's work focused on proving that the conjoint time-frequency properties of one-dimensional Gabor function are still satisfied for the two-dimensional case. In 1987, Jones and Palmer [49] confirmed that Gabor filters provided good fits to the receptive field profiles that they measured in cat visual cortex, see Fig. 3.1. However, there are other functions with additional parameters that fit to biological profiles such as Differences of Gaussians (DoG) or Hermite functions [50, 51].

Daugman [52] also found that such psychophysical features could be modeled with **Gabor functions**, which are defined as the product of a complex sinusoid and a Gaussian function that acts as a modulator or signal envelope. Starting from the definition in 1D, we can use Euler's identity and prove that a Gabor function consists of two filters, as follows:

$$\begin{aligned} g_{\omega_0}(x) &= e^{-\pi\alpha^2 x^2 + i\omega_0 x} \\ &= e^{-\pi\alpha^2 x^2} e^{i\omega_0 x} \\ &= e^{-\pi\alpha^2 x^2}\left[\cos(\omega_0 x) + i\sin(\omega_0 x)\right] \\ &= e^{-\pi\alpha^2 x^2}\cos(\omega_0 x) + i e^{-\pi\alpha^2 x^2}\sin(\omega_0 x) \\ &= g_e(x) + i g_o(x) \end{aligned} \tag{3.2}$$

where g_e represents an even-symmetric filter, whereas g_o is an odd-symmetric filter. Note that the Gaussian function is a particular case when $\omega = 0$.



(a) *(b)*

Figure 3.2: *Gabor functions in spatial domain.* ***(a)*** *Even-symmetric Gabor function, $g_e(t)$, that it is able to detect salient edges.* ***(b)*** *Odd-symmetric Gabor function, $g_o(t)$; this function is ideal for detecting step-like discontinuities.*

Moreover, it has been found that certain pairs of cells have their receptive fields located in the same position but with a relative separation in phase of $\Delta\phi = 90°$, which means that they are in quadrature phase [1]. As we showed in Eq. (3.2), both real and imaginary parts can be separate in order to check that they are in quadrature phase, see Fig. 3.2.

2D Gabor functions are a generalization of 1D elementary functions or **logons**, originally proposed by Dennis Gabor [53] in 1946 for signal expansion. Gabor's work is a continuation and partly parallel to the works of Harry Nyquist [54] and Claude Shannon [55] who found the theory of communication. Gabor functions are band-pass filters and occupy the smallest possible volume in the time-frequency space, where orthogonal axes correspond to spatial (x, y) and frequency (u, v) variables. The volume meets the theoretical lower bound of the uncertainty principle as follows: $(\Delta x)(\Delta y)(\Delta u)(\Delta v) \geq \frac{1}{16\pi^2}$. The canonical 2D Gabor filter is defined in spatial domain as:

$$g(x, y) = Ke^{-\frac{1}{2}\left[\frac{(x-x_0)^2+\gamma^2(y-y_0)^2}{\alpha^2}\right]+i\left(2\pi\left[u_0(x-x_0)+v_0(y-y_0)\right]+\phi\right)} \tag{3.3}$$

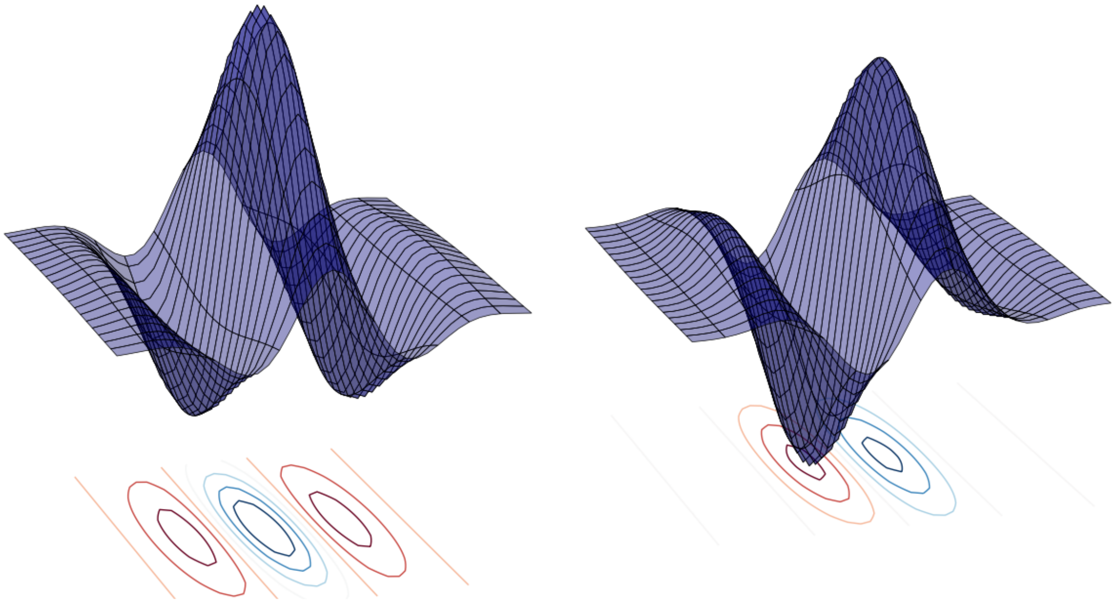
where $K = \frac{\gamma}{2\pi\alpha^2}$ is the normalizing constant; (x_0, y_0) are the center of the filter, (u_0, v_0); and ϕ represent the radial frequency and the phase of the sinusoidal signal, respectively. (α, γ) are the space constants of the Gaussian envelope along x- and y-axes respectively and they control the filter bandwidth.

In order to simplify Eq. (3.3) we only considered filters centered at the origin ($x_0 = 0$, $y_0 = 0$). Although the phase is implicitly encoded, it was ruled out because it can take very different values even if it is sampled at image locations only a few pixels apart. This fact makes it difficult to extract reliable and discriminative features from the phase responses. So, the equation is reduced as:

$$g(x, y) = Ke^{-\frac{1}{2}\left(\frac{x^2+\gamma^2y^2}{\alpha^2}\right)}\left[\cos(2\pi u_0 x) + i\sin(2\pi u_0 x)\right] \tag{3.4}$$

---

[1] Two signals are in quadrature phase, ϕ, if the difference between phases is constant and equal to $\pm\frac{\pi}{2}$



(a) (b)

Figure 3.3: *2D Gabor functions in the spatial domain.* ***(a)*** *Real part, which is an even-symmetric function and* ***(b)*** *imaginary part that corresponds to an anti-symmetric function.*

As in the previous Eq. (3.2), the Euler's identity was also applied to Eq. (3.3) in order to separate real and imaginary parts, $g\left(x,y\right) = g_e\left(x,y\right) + ig_o\left(x,y\right)$ where:

$$g_e\left(x,y\right) = Ke^{-\frac{1}{2}\left(\frac{x^2+\gamma^2y^2}{\alpha^2}\right)}\cos\left(2\pi u_0 x\right) \tag{3.5}$$

is an even-symmetric function, which responds with a maximum in zero, see Fig. 3.3(a). This property provides suitable functions for detecting salient edges. On the other hand,

$$g_o\left(x,y\right) = Ke^{-\frac{1}{2}\left(\frac{x^2+\gamma^2y^2}{\alpha^2}\right)}\sin\left(2\pi u_0 x\right) \tag{3.6}$$

is an odd-symmetric function, which responds to zero-crossing, see Fig. 3.3(b). This function is ideal for detecting step-like discontinuities [56].

Traditionally, a rotation transform is applied to Gabor filters to move them onto a specific angle and take advantage of their directionality, $g_\theta^*\left(x,y\right) = Rg\left(x,y\right)$. Thus, using the rotation matrix

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{3.7}$$

and applaying to Eq. (3.5) yields a 2D polar Gabor function as follows:

$$g_\theta^*\left(x,y\right) = e^{-\frac{1}{2}\left(\frac{\widetilde{x}^2+\gamma^2\widetilde{y}^2}{\alpha^2}\right)}\cos\left(2\pi u_0\widetilde{x}\right) \tag{3.8}$$

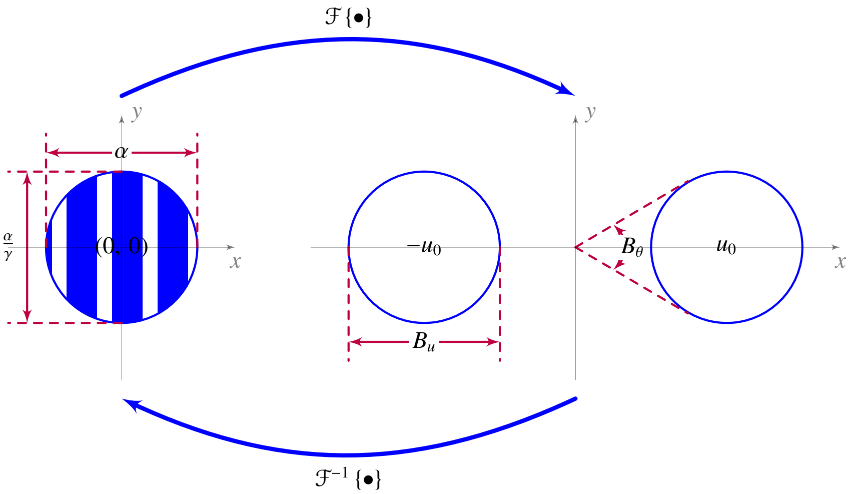


Figure 3.4: *Fourier transform pair of a 2D Gabor filter. The figure illustrates the dependence between spatial and frequency bandwidths.*

with

$$\begin{aligned}\widetilde{x} &= x\cos\theta - y\sin\theta \\ \widetilde{y} &= x\sin\theta + y\cos\theta\end{aligned} \tag{3.9}$$

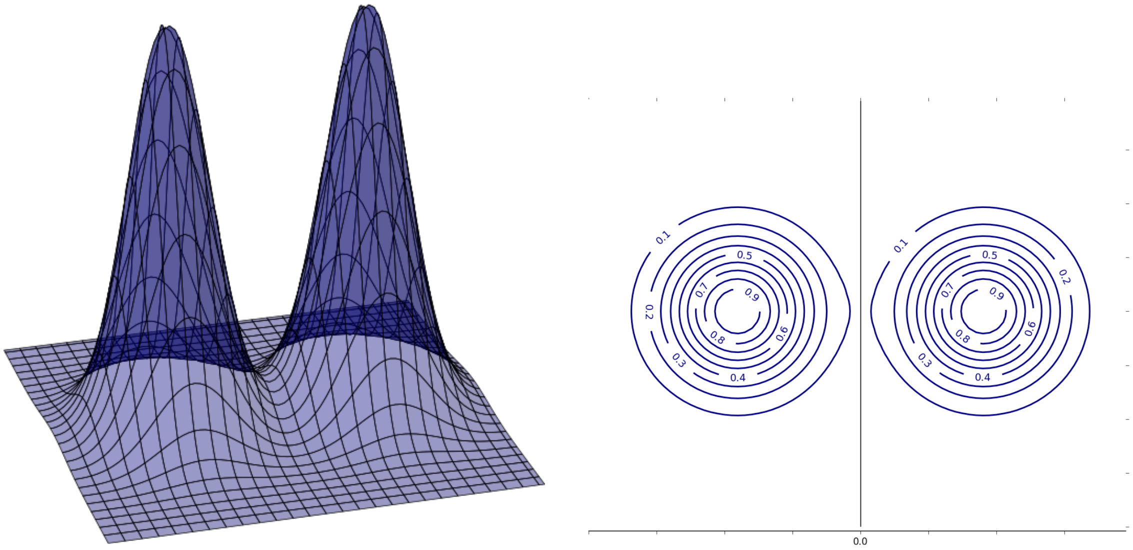
this procedure is applicable in a similar manner to the odd-symmetric filter to obtain the imaginary part.

In the spatial domain, 2D Gabor filters are complex sinusoids modulated by 2D Gaussian functions, while in the frequency domain they are shifted Gaussians. Heisenberg's principle states that there is a trade-off between these representations. The filter's spatial-frequency and orientation bandwidths are linked and a circular filter envelope in the space domain is supported by the sum of two circular regions in the frequency domain. Centers of those regions correspond to the filter's modulation frequency. Additionally, we can say that spatial-frequency and orientation bandwidths are inversely related to the space-domain envelope diameter, see Fig. 3.4.

The frequency and orientation selectivity properties of Gabor functions can be more explicit in the Fourier domain. If we consider that g_e is a pure-real even-symmetric function, then its Fourier transform, $\mathcal{F}\{g_e\}$, is given by $\frac{1}{2}\left[\hat{G}(u,v)+\hat{G}(-u,-v)\right]$, which is a symmetric function about the y-axis, see Fig. 3.5(a), whereas g_o is a pure-real odd-symmetric function and its Fourier transform, $\mathcal{F}\{g_o\}$, is $\frac{1}{2}\left[-i\hat{G}(u,v)+i\hat{G}(-u,-v)\right]$, which is a symmetric function around the origin. Note that if both even and odd parts are used, they closely approximate to a Hilbert transform pair namely negative frequencies become zero. Therefore, the Fourier transform of $g(x,y)$ is given by

$$\hat{G}(u,v) = e^{-2\pi^2\alpha^2\left[(\tilde{u}-u_0\cos\theta)^2+\frac{1}{\gamma^2}(\tilde{v}+u_0\sin\theta)^2\right]} \tag{3.10}$$

$\hat{G}(u,v)$ represents a rotated Gaussian function by an angle θ with u_0 frequency units shifted along the x-axis and $(\tilde{u},\tilde{v}) = (u\cos\theta, -u\sin\theta)$.



(a) (b)

Figure 3.5: *Fourier transform of a 2D real Gabor function.* ***(a)*** *In the frequency domain the even-symmetric filter is represented by two real-valued Gaussian functions symmetrically placed respect to the origin.* ***(b)*** *Contour lines of a Gabor filter in the frequency domain. Note that Gaussian filters are deformed as they approach to the origin due to the average of all DC components.*

The selection of the optimal Gabor parameters is an open issue because they depend mainly on the application problem, therefore is very difficult to propose a general scheme. There are many possibilities, one of them is to experiment with the width of the Gaussian envelope by changing (α, γ) values, which determine resolution in both spatial and frequency domains.

$G(u, v)$ is a band-pass filter and its bandwidth is controlled by α and γ; if the aspect ratio $\lambda = \frac{\alpha}{\gamma} = 1$, then Eq. (3.10) represents an axisymmetric filter. However, the main issue lies on choosing radial-frequencies (u_i) that define the degree of overlapping between two adjacent filters. The higher the central frequency of the Gabor sinusoidal carrier, the smaller area the Gaussian envelope will cover in the spatial domain [57].

A class of self-similar functions can be obtained by appropriate translations namely changing radial-frequency values and rotations (θ) of a Gabor function to form an array. If we assume that the number of translations is T and the number of rotations is R then the filter bank consists of $T \times R$ filters.

Psychophysical experiments have shown that frequency bandwidths of simple receptive fields in the HVS are about one octave[2] apart [52, 58]. To fulfill this condition, the half-amplitude bandwidth (B_u) of each filter was linked to its central frequency as follows:

$$\alpha = \frac{\sqrt{\log(2)}\left(2^{B_u} + 1\right)}{\sqrt{2}\pi u_0 \left(2^{B_u} - 1\right)} \tag{3.11}$$

---

[2] An octave is the bandwidth between two frequencies, one of which is double the other.

Note that the maximum possible discrete frequency of an image along x- or y- axes is $\frac{1}{2}$ cycles per pixel. Beyond this limit, the radial bandwidth will be very large, thus, the filtering may cause artifacts [59]. Therefore, the maximum number of frequency bands must be computed by dividing the central frequency by the size of the image. So, we selected the four following dyadic values $u = \{\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}\}$ to build an optimal filter bank with 4 scales, the central frequency should be given in cycles/image-width. On the other hand, if the central frequency is very small, then filters will behave as low-pass filters rather than band-pass filters; this fact usually leads to a loss of information in classification problems.

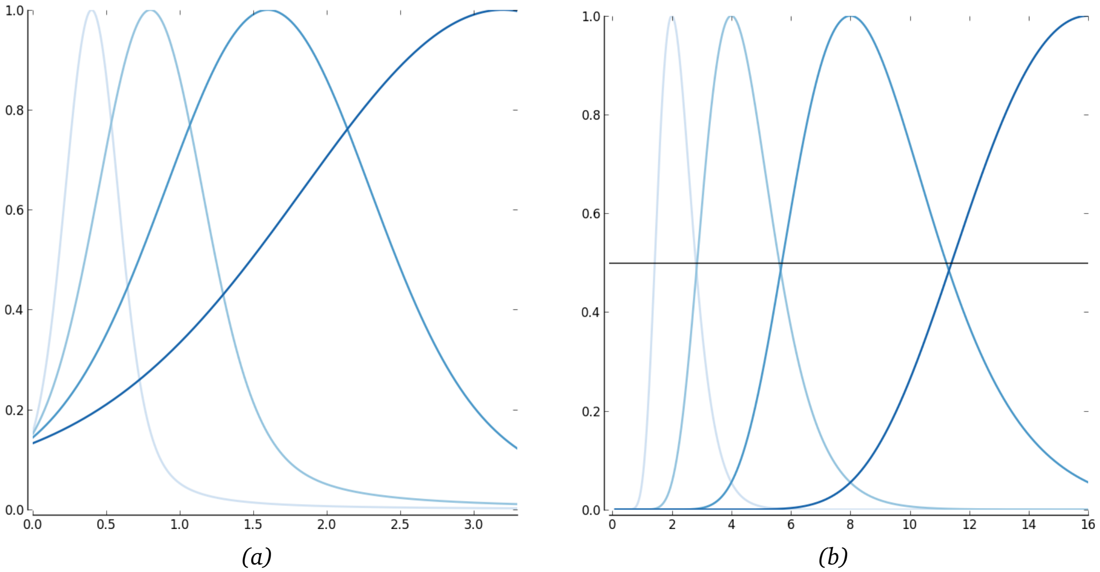
In order to determine the optimal angular bandwidth (B_θ), we considered axisymmetric filters and set $\gamma = 1$ in Eq. (3.12). In this way, $B_\theta \approx 36^\circ$ but for computational efficiency $B_\theta = \frac{\pi}{6}$ was chosen. This setting resulted in a filter bank with 6 orientations.

$$\frac{\alpha}{\gamma} = \frac{\sqrt{\log(2)}}{\sqrt{2}\pi u_0 \tan\left(\frac{B_\theta}{2}\right)} \tag{3.12}$$

3.3 Log-Gabor filters

Gabor filters possess a number of interesting mathematical properties: first, they have a smooth and indefinitely differentiable shape; and second, they do not have side lobes neither in space nor frequency domain. Nevertheless, they present some important drawbacks such as the maximum bandwidth limited to approximately one octave and Gabor filters are not optimal if one is seeking broad spectral information with maximal spatial localization [60]. In the following list we reviewed three important drawbacks.

- The Gabor filter averaging is not null, hence, the DC component influences the intermediate bands. Filters overlap more at lower frequencies than at higher ones yielding a non-uniform coverage of the Fourier domain, see Fig. 3.6(a). In many image-based applications, variation of DC components generally deteriorates system performance. To deal with this problem, researchers commonly use zero DC Gabor filters by setting $(x = 0, y = 0)$ to zero. Additionally, to eliminate the influence of the power of Gabor filters, a normalized version is used, $g_n(x, y) = \frac{g(x,y)}{\|g(x,y)\|}$.
- Gabor filters have been successfully used for image analysis and applications where exact reconstruction is not required because the traditional arrangement in scales and orientations does not cover uniformly the Fourier plane, therefore, the reconstruction is not exact. In fact, Gabor filters are band-pass filters by definition; they can not cover lowest and highest frequencies. In parallel, different methods for reconstruction improvement have been proposed: Gross et al. presented a method to cover more uniformly the Fourier domain [61], in [62], Nestares et al. introduced a method to recover the highest frequencies, and in [63], Lee showed an improvement for the reconstruction stage.



(a) *(b)*

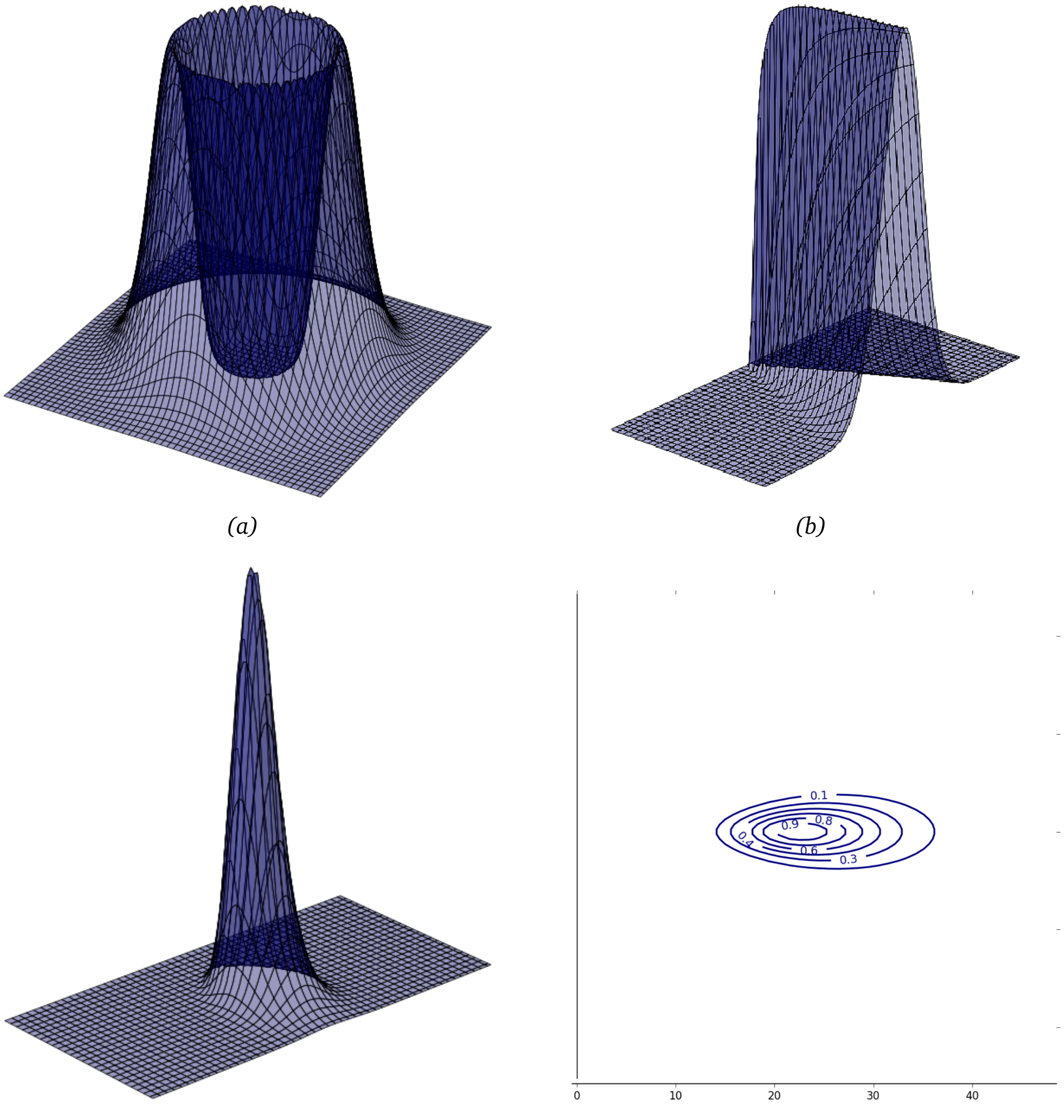
Figure 3.6: *Frequency response profiles of* ***(a)*** *Gabor filters and* ***(b)*** *log-Gabor filters. Note that DC component is minimized by the introduction of the logarithm.*

- It is not possible to build a complete orthogonal basis of Gabor functions. Non-orthogonality implies that exact reconstruction using the same filters for analysis and synthesis will not be possible unless an overcomplete dictionary is considered [56]. However, several sparse coding algorithms have been used to address this issue such as Matching Pursuit, which chooses one by one the highest coefficients in all the dictionary and Basis Pursuit, which solves the linear system minimizing a penalizing function corresponding to the sum of the amplitude of all coefficients. Both of these algorithms perform iteratively and globally through all the dictionary [47].

In order to avoid these drawbacks, in 1987, David Field [64] proposed Log-Gabor filters, which are defined in the frequency domain as Gaussian functions shifted from the origin because of the singularity of the $\log$ function, see Fig. 3.6(b).

By definition they always have a null-DC component and can be optimized to produce filters with minimal spatial extent in an octave scale multi-resolution scheme. Additionally, this filtering scheme not only achieves important mathematical properties, it also follows the knowledge on the receptive field properties of simple cells in the V1 area.

Compare to the state of art, log-Gabor filters show excellent ability to segregate image information. We used the biological knowledge about V1 as a useful guide for choosing the best parameters in order to achieve an optimal filtering scheme. It is important to provide good image quality from the human perceptual point of view because human beings are the ultimate receivers in most applications.



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

Figure 3.7: *Example of a Log-Gabor Filter.* ***(a)*** *Concentric rings of the radial component of the filter.* ***(b)*** *The angular component is only defined for the half of the plane in order to reduce redundancy.* ***(c)*** *The product of radial and angular components.* ***(d)*** *Contour lines of a log-Gabor coefficient shifted from the origin.*

Log-Gabor filters can be split into two components: radial component, see Fig. 3.7(a), and angular component, see Fig. 3.7(b). The former controls the frequency band and the latter controls the orientation. The two components are multiplied together to construct the overall filter, $\hat{G}(\rho, \theta) = \hat{G}_\rho \hat{G}_\theta$, see Fig. 3.7(c) and Fig. 3.7(d).

$$\hat{G}(\rho, \theta) = e^{-\frac{1}{2}\left[\frac{\log\left(\frac{\rho}{u_0}\right)}{\log\left(\frac{\alpha_\rho}{u_0}\right)}\right]^2} e^{-\frac{1}{2}\left[\frac{(\theta-\theta_0)}{\alpha_\theta}\right]^2} \tag{3.13}$$

where (ρ, θ) represent polar coordinates; u_0 is the central frequency; θ_0 is the orientation angle; α_ρ and α_θ determine the scale and the angular bandwidth respectively.

In order to better cover the Fourier plane, we introduced a modified multi-scale decomposition scheme. The even scales were rotated by a constant factor consisting of the half a distance between filter centers. Based on experiments, we set $\alpha_\rho = 0.75$ that results in a minimal overlap among scales one octave apart and $\alpha_{theta} = \frac{pi}{6}$ as it was mentioned in Section 3.2.

This implementation fulfills the following constraints: **(i)** an optimal localization in space, frequency, and orientation through the use of logarithmic filters; **(ii)** a resemblance to receptive fields of V1 simple cells; **(iii)** this configuration of filters is particularly suited for coding local contrast changes in natural images; and **(iv)** when log-Gabor filters are compared with other schemes, such as biorthogonal wavelets, translation invariant wavelets, and steerable pyramids, the log-Gabor filters have shown a significant improvement in the noise removal capability by using simple threshold operations [47].

3.4 Gabor jets

The main drawback when using Gabor or log-Gabor filters is the high degree of redundant information. For each real-value pixel in the image I there are $S \times D$ complex values corresponding to the number of scales, S, and orientations, D, of the filter bank. This means that neighboring pixels share about the same information and, hence, can be approximated for only salient Gabor coefficients namely important coefficients in terms of energy, statistics, or even visual aspects.

Buhmann et al. [65] introduced the concept of **Gabor jets** by concatenating the responses of all Gabor coefficients at a specific position into a vector, see Fig. 3.8. Although they used a different Gabor family, the fundamentals of Gabor filters not change.

The Gabor jet, $J^I(x, y)$, of an image I at the pixel position (x, y) is define by

$$J^I(x, y) = \{J_k(x, y) | k = 1, \ldots, S \times D\} \tag{3.14}$$

where

$$J_k(x, y) = \{I \star g_{(s,\theta)}\}_{(x,y)} \tag{3.15}$$

J_K stands for the k-th coefficient of the feature vector extracted from the convolution of the image I

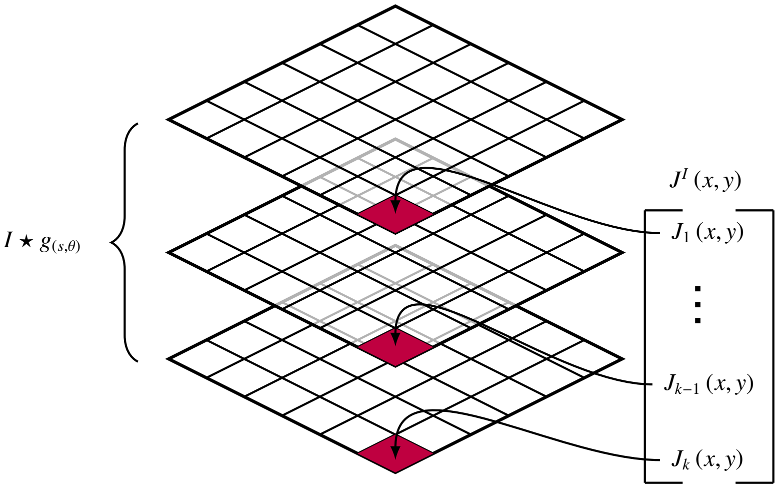


Figure 3.8: *A Gabor jet vector, $J^I(x,y)$, is built by concatenating the responses of all Gabor coefficients at a specific position into a single vector.*

and the k-th Gabor filter at pixel position (x,y). Here, $\left|J^I\right|$, can be seen as the response of a different complex cell. Thus, we can establish the following idea:

The absolute values around an offset point are relatively stables for a small displacements d

$$\left|J^I(x,y)\right| \approx \left|J^I(x+d_x, y+dy)\right| \tag{3.16}$$

This allows us to perform a downsampling procedure (↓) by a factor of N in order to minimize redundant information. For instance, let A be an image of 31×31 pixels and let B be a Gabor filter bank with 6 orientations and 4 scales. Therefore, the length of the Gabor jet $J^A(x,y)$ at pixel position (x,y) will be 24 complex values and a possible feature vector will have 23064 bins. If a $\downarrow 3$ procedure is applied, then the feature vector will be 15384 bins.

– Some scientists claim that hydrogen, because it is so plentiful, is the basic building block of the universe. I dispute that. I say there is more stupidity than hydrogen, and that is the basic building block of the universe.

Frank Zappa

4

Texture segmentation

4.1 Introduction

For human beings, an image is not just a group of pixels but a collection of objects and regions that are somehow related to each other, and in spite of the large variation of visual scenes, the HVS has no problems interpreting them. In the real world, images often do not exhibit uniform intensities but variations in brightness, color, texture, or other attributes which form a certain pattern.

Since an image is made up of pixels, texture can be defined as an entity consisting of mutually related pixels and group of pixels called texture primitives or texture elements (**texels**). It is generally believed that the HVS uses them as effective features to discriminate objects because the brain is able to decipher important variations in data at smaller scales. However, the issue of texture characterization lies on the fact that some spatial patterns can be quite simple as stripes while others can exhibit complex behaviors. From a mathematical point of view, it is usual to analyze spatial distributions of pixels as intensity variations from regularity, where texture contains periodic patterns, to randomness, where texture looks like unstructured noise.

Texture plays an important role in distinguishing one region from another because intensity variations may reflect different areas, shapes, or objects. According to Dixit et al. [66], there are three fundamental features that human beings use to interpret images: spectral, textural, and contextual information. However, only texture features provide information about the structural arrangement of a surface or region and the relationship with its surrounding environment.

The first steps towards computational texture characterization were made by Julesz [67] who investigated texture perception and conjectured that two textures are not unconsciously distinguishable if their second order statistics are identical. Afterward, he proved that his own idea was false [68] but he established that texture might be modeled by low-order statistics.

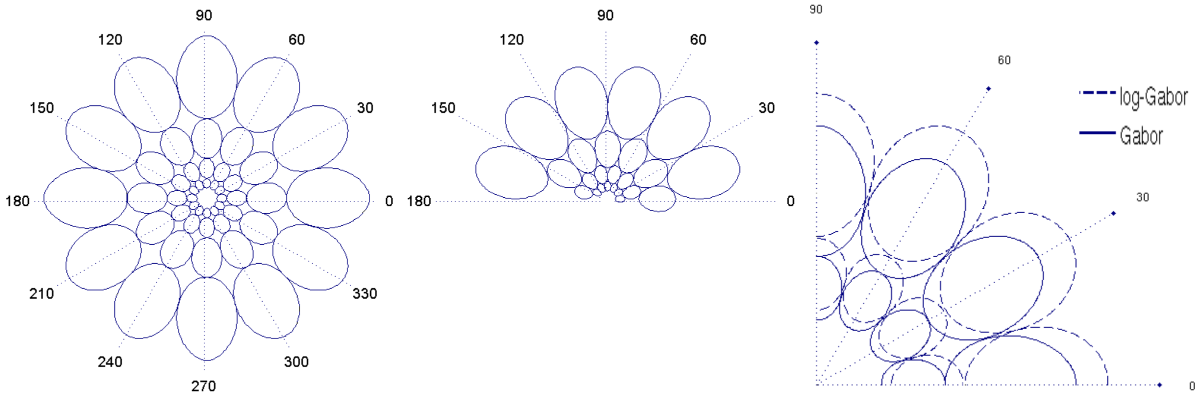
Based on the fact that texture possesses spatial continuity at both local and global scales and can exhibit a large number of features, it has been widely used to perform segmentation. The occurrence of texture in an image is useful because when two or more regions meet, an edge is formed. We used this fact and defined texture segmentation as a division of the whole image into homogeneous regions characterized by the same texture.

Therefore, texture segmentation can be performed in two different ways: as gray level segmentation or as feature segmentation.

In [69], the following four approaches for texture segmentation have been recognized: **(i)** statistical methods analyze spatial distributions of pixels by computing local features at each point of the image and deriving a set of statistics. These statistics may describe properties such as mean, variance, skewness, kurtosis, and so on. They are based on the assumption that intensity variations are more or less constant within a texture region and take a greater value outside their boundary [70]. Latest approaches in this category use Markov fields to model relationships among pixels within regions and over time [71]; **(ii)** spectral methods collect a distribution of filter responses for a further classification. For instance, researchers have used Gabor filters to discriminate between different kind of textures. These methods emphasize the extraction of appropriate features for discriminating among specific textures. A comparative study can be found in [72]. Many algorithms in this category are focused on face recognition and classification as in [73]; **(iii)** structural methods characterize texture by a set of primitives, which are organized according to a certain placement rule that defines the spatial relationship among primitives and may be expressed in terms of adjacency [74]; and **(iv)** stochastic methods assume that textures are the realization of a stochastic process and estimate parameter associated with the process, e.g., [75] uses a Bayesian approach as a texture descriptor.

The aim of texture segmentation is to produce homogeneous regions with respect to texture primitives that correspond to objects in the real scene under study. However, it is still an open and complicated problem, which is strongly influenced by the quality of data. For this reason, we must consider two fundamental issues: what is the feature that best discriminates textures? and what is the underlying model that defines region homogeneity and, thus, specifies what a good segmentation should be?

Based on the fact that the HVS can segment textures robustly, many segmentation schemes use biological models. There is an evidence that perceptually texture differences correspond to differences in local spatial frequency content [76]. This suggests that a good algorithm must decompose a texture image into a joint space/spatial frequency representation and then use this information to locate regions of similar local spatial frequency content.



(a) *(b)* *(c)*

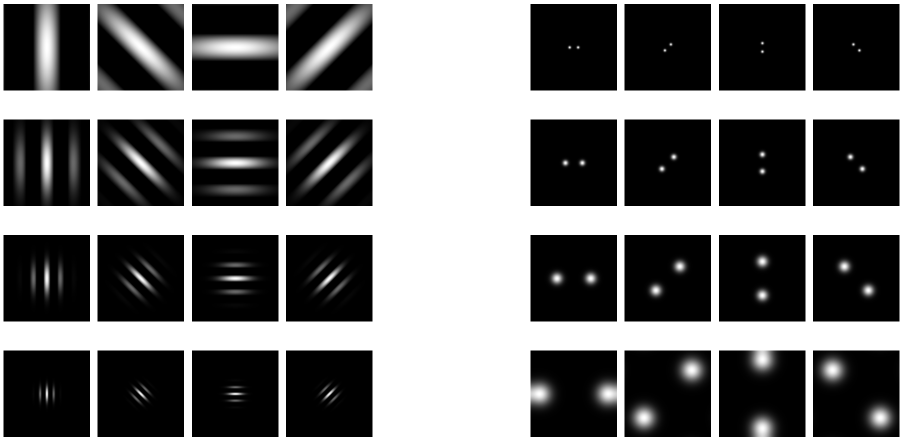
Figure 4.1: *Half-amplitude bandwidth of **(a)** an ensemble of 2D Gabor filters in the frequency domain; **(b)** 2D Log-Gabor filters; and **(c)** contour comparison between Gabor filters and log-Gabor filters before rotating log-Gabor even bands.*

In particular, Gabor filters, see Section 3.2, provide the ability to perform multiresolution decomposition; they offer the best simultaneous localization of spatial and frequency information and respond to a limited range of signals which form a repeating structure in some direction and in some frequencies [77]. In the literature we can find quite a few papers devoted to texture segmentation using Gabor filters [58, 76, 78, 79]. Moreover, texture segmentation techniques can be either supervised or unsupervised. The latter stand out because it does not need any prior knowledge concerning textures. This is very useful due to the fact that no prior information is available in most practical applications [80, 81].

Despite the fact that Gabor filters optimize the theoretical limit of joint resolution between space and frequency domain, they do not have zero-mean, which induces a DC component in frequency bands. In addition, they do not have a uniform coverage of the frequency domain. These drawbacks may cause errors in the extraction of the appropriate texture features.

In this Chapter, the Jain and Farrokhnia's work [79] is visited; the motivation here is to investigate the advantages of log-Gabor filters, Section 3.3, by comparing their performance with Gabor filters, Section 3.2. Log-Gabor filters allow to eliminate DC component; they can yield a fairly uniform coverage of the frequency domain in an octave scale scheme and preserve redundancy at the same time, see Fig 4.1. This proposal is partially based on some recommendations given by Clausi and Jernigan [82] and Bovik et al. [58].

In the next Section 4.2, the Jain and Farrokhnia's algorithm is presented along with a modification to prevent that the algorithm uses random seeds in k-means. In Section 4.3 experimental results are presented. Conclusions are drawn in Section 4.4.



(a) *(b)*

Figure 4.2: *Ensemble of a Gabor filter bank in* ***(a)*** *the spatial and* ***(b)*** *Fourier domain. Four radial frequencies* $u = \left\{8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2}, 64\sqrt{2}\right\}$ *cycles/image-width and four orientations* $\theta = \left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$.

4.2 Jain and Farrokhnia's unsupervised method

This method uses a multi-channel filtering approach, which involves two processing steps: **linear filtering**, where an image is filtered by a set of channels, see Fig. 4.2. The objective in this step is to estimate the energy in the filter output within a local region. Previous studies pointed out that Gaussian-like filters are, by far, the better choice [72, 83]; and **post-processing step**, this process typically involves a non-linear point operation followed by computation of some local statistics.

Jain and Farrokhnia proposed a bank of even-symmetric Gabor filters (4 orientations and 7 scales) to characterize the channels; each filter response is subjected to a bounded nonlinear transformation that behaves as a blob detector. They included a statistical stage to capture image features. Finally, the feature vectors are classified using k-means. The complete algorithm is summarized in Algorithm 4.1.

Algorithm 4.1: Jain and Farrokhnia's unsupervised method.

```
Input: I := texture image
Input: NofT := number of textures
Output: SI := segmented image
begin
    FI := Gabor filter bank(I);
    BI := non-linearity (FI);
    FVI := local energy computation (BI);
    SI := $k$-means(FVI, NofT);
end
```

4.2.1 FEATURE EXTRACTION

The frequency and angular bandwidths, (B_u, B_θ) respectively, can be set to constant values to match psychovisual data. We used Eq. (3.11) to determine the value of the Gaussian envelope along the x-axis (α). Psychophysical experiments showed that frequency bandwidths of simple cells in the visual cortex are one octave apart namely $B_u = 1$. By setting the frequency cut-off to -6 dB, we determined the ratio $\alpha = \frac{0.56}{u_0}$ namely the half-amplitude bandwidth of each filter is linked to its central frequency.

We determined the optimal B_θ by considering Eq. (3.12); we assumed that the filters are isotropic, so we set $\alpha = \gamma = 1$. In this way, $B_\theta \approx 36^\circ$ but for computational efficiency $B_\theta = \frac{\pi}{6}$ was chosen. This setting suggest to use a filter bank consisting of 6 orientations while the number of scales is constrained by the size of the images.

The following values of radial frequency were used: $u = \left\{2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, 16\sqrt{2}, 32\sqrt{2}, 64\sqrt{2}\right\}$ cycles/image-width for an image of size 256×256. In this way, we guarantee that the pass-band filter with the highest radial frequency falls inside the image array. At this point, we rule out low-pass filter and filters with low radial frequencies because these filters capture spatial variations that are too large for textures [82].

Each filtered image, f_i, was subjected to a sigmoid function that can be interpreted as a blob detector. Jain used an empirical value of $\gamma = 0.25$. We reached the best performance using $\gamma = 0.1$.

$$\phi(f_i) = \tanh(\gamma f_i) = \frac{1 - e^{-2\gamma f_i}}{1 + e^{-2\gamma f_i}} \tag{4.1}$$

Jain and Farrokhnia also suggested to apply a Gaussian window function to each filtered image. The bandwidth of each Gaussian function is proportional to the average size of intensity variations in the image.

4.2.2 CLUSTERING

The last step consists in integrating the features corresponding to different filters to produce a good segmentation. If feature vectors are capable of discriminating patterns belonging to different textures, each patterns will form a cluster. Each feature vector was built using the 36 Gabor filters by concatenating the responses of all coefficients at a specific position, "Gabor-jets-like style", see Section 3.4. Furthermore, we included the spatial coordinates of the pixels, the mean, and the variance as additional features in every vector.

The clustering algorithm used in this approach is k-means with initial random seeds. The issue of clustering or grouping n objects into k groups arises in many scientific disciplines and is not always easy to obtain an optimal partition. The k-means algorithm is very popular, however, the algorithm is prone to the initializing values which greatly influence the determination of an optimal solution.

We used an initialization procedure proposed in [84] where a bootstrap method for finding seeds was used. This method consists in clustering several subsamples of the dataset using k-means. Each cluster produces a different set of candidates from which the initializers are chosen. The complete algorithm is summarized in Algorithm 4.2.

Algorithm 4.2: K-means bootstrap algorithm.

```
Result: Image pixels are assigned to K classes (labeled pixels)
Input: DS := dataset
Input: K := number of classes
begin
    for $i \leftarrow 1$ to 10 do
        subDS := a random subset of DS;
        randSeeds := extractRandomSeeds(subDS);
        [centroids[i], labels of subDS] := $k$-means(randSeeds, subDS, K);
    end
    initializers = average of centroids;
    [centers, labels of DS] := $k$-means(initializers, DS K);
end
```

4.3 Experimental results

The algorithm previously described was tested using textures extracted from the Brodatz album [85]. We used both Gabor and log-Gabor filters into the pipeline in order to compare their performance. All tests were assessed using image mosaics from two to five textures of size 256×256, see Fig. 4.3 **(first row)**. The final segmentation with Gabor filters is shown in Fig. 4.3 **(second row)** and Fig. 4.3 **(third row)** illustrates results using log-Gabor filters.

We used confusion matrices and computed the accuracy rate (AR), which is the number of correct guesses(elements in the diagonal), using the next equation:

$$AR = \left(\frac{\sum_i^k a_{i,i}}{\sum_{i,j}^k a_{i,j}}\right)100\% \tag{4.2}$$

where (i, j) are the matrix indexes and k is the number of textures in the image.

Due to complexity the five-texture case is the most interesting one. Note that the gap between Gabor and log-Gabor filters is wider, see Fig. 4.4, because segmentation using Gabor filters observed more artifacts than log-Gabor filters near the circular border. Furthermore, qualitative assessment suggests that log-Gabor misclassification is caused by the circular convolution. Next, we present the final texture segmentation using confusion matrices. Table 4.1 shows the Gabor filter performance while Table 4.2 summarizes the result using log-Gabor filters.

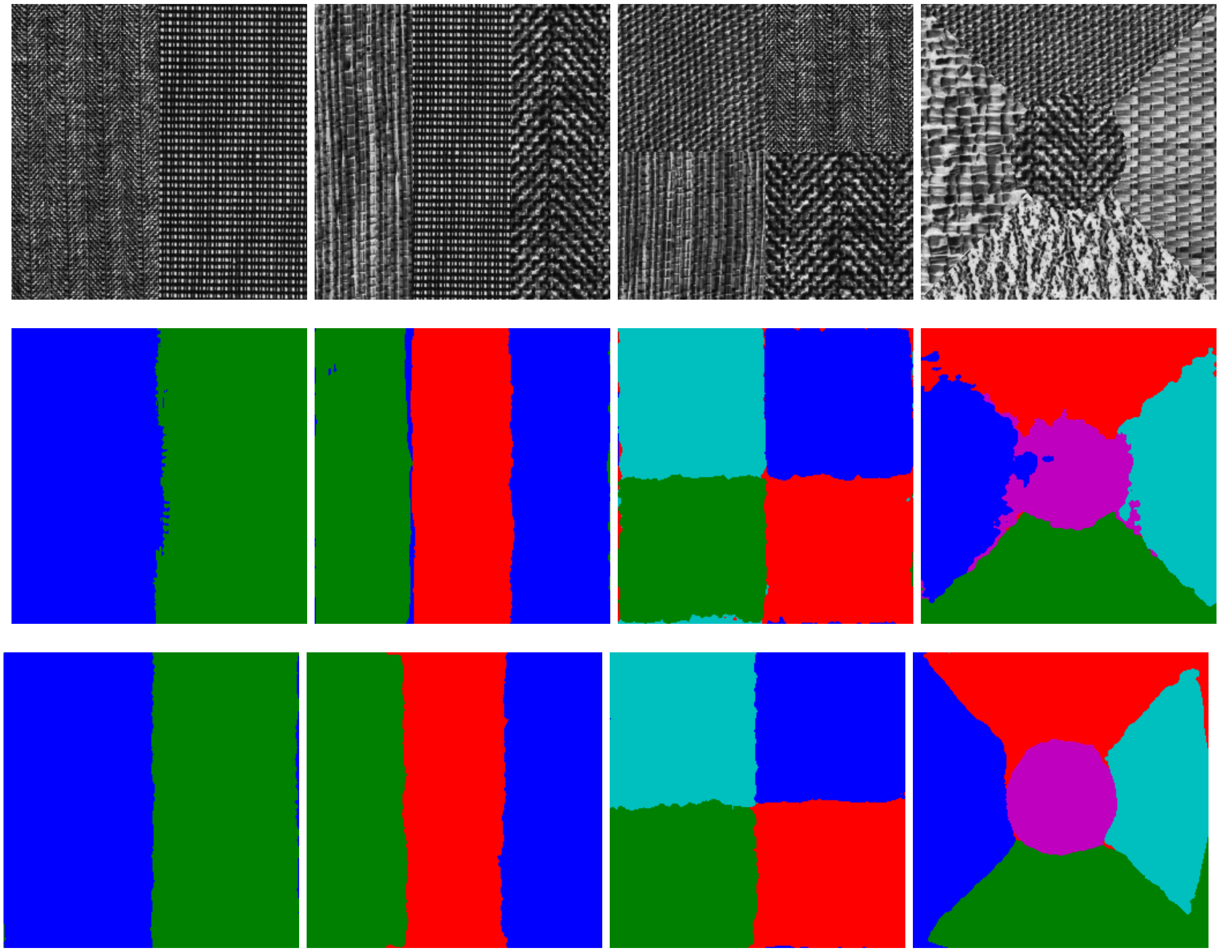


Figure 4.3: ***(First row)*** *Brodatz images used for segmentation. two mosaics: D16 and D21; three mosaics: D79, D21, and D17; four mosaics: in clockwise direction D77, D16, D17, and D79; five mosaics: D55, D77, D84, D24, and D17. The number indicates the page of the album where the texture was taken. Segmentation results obtained using a total of 36 filters. The feature vectors were built using the spatial coordinates of the pixels, the mean, and the variance* ***(second row)*** *with even-symmetric Gabor filters and* ***(third row)*** *log-Gabor filters.*

Table 4.1: *Confusion matrix for the experiment with five textures and Gabor filters. We obtained an AR = 91.48%. Jain and Farrokhnia reported 80% of pixels correctly segmented. The precision of the algorithm decreases because the algorithm cannot distinguish the border between D55 and D17 patterns.*

		Predicted values				
		1	**2**	**3**	**4**	**5**
Actual values	**1**	11895	1167	0	561	757
	2	0	14164	0	0	126
	3	0	736	12385	613	666
	4	0	0	0	14496	0
	5	498	131	0	327	7014

Table 4.2: *Confusion matrix for the experiment with five textures and log-Gabor filters. We obtained an AR = 94.93%. This means that our proposal achieved over a 4% of pixels correctly classified.*

		Predicted values				
		1	**2**	**3**	**4**	**5**
Actual values	**1**	14106	242	0	32	0
	2	108	13810	224	0	148
	3	0	501	12510	1384	5
	4	209	9	0	14262	16
	5	110	99	81	156	7524

We computed precision and sensibility for both schemes using the confusion matrices. From Table 4.1, precision = 91.19% (±7.08%) and sensibility = 91.17 (±7.89%), while from Table 4.2 precision = 95.35% (±3.28%) and sensibility = 94.88% (±4.74%). This represents an increase over 4% of pixels correctly classified with our proposal, namely more than 2000 pixels. Note that precision differs from AR because it only measures the number of true positives, which are the proportion of positive cases that were correctly identified, while AR measures true positives and false negatives.

The final accuracies for all the cases are shown in Fig. 4.4; note that in the case of two textures, both Gabor and log-Gabor filters have almost the same performance but by increasing the number of classes, Gabor filters classify more pixels incorrectly, which lead us to think that in case of more complex patterns log-Gabor filters can do a better segmentation with fewer errors. Furthermore, in all cases log-Gabor filters overcame Gabor filters.

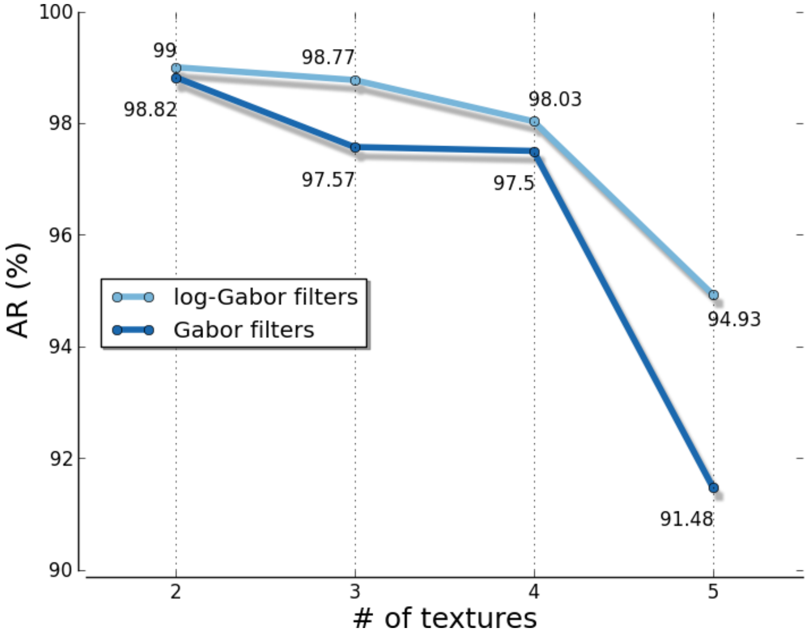


Figure 4.4: *The final accuracies for both Gabor and log-Gabor filters. Note that as increasing the number of classes the accuracy goes down and the gap between Gabor and log-Gabor filters is wider.*

4.4 Conclusions

Since Gabor filters have been successfully used in texture segmentation, we are interested in compare our proposal with the state-of-art method under the same conditions. We selected the optimal parameters for Gabor filters and kept the equivalent values in the log-Gabor scheme. Although Jain and Farrokhnia's algorithm is already proven, we modified some stages and added an initialization of centroids for *k*-means classifier. Based on biological studies we considered six scales and six orientations –Jain and Farrokhnia used only 4 scales. Results showed that log-Gabor filters overcame Gabor approach in complex textures. For the test of five textures log-Gabor provides 94.92% of accuracy while Gabor scheme scored 91.48% of pixels correctly classified. Both schemes outperform Jain and Farrokhnia's report. The major differences occur at the borders between textures. The confusion matrices, in Table 4.1 and Table 4.2, show that major mistakes were found between textures D55 and D17 for Gabor scheme and between D77 and D84 for log-Gabor filters respectively.

– Science never solves a problem without creating ten more.

George Bernard Shaw

5

Texture analysis and retrieval

5.1 Introduction

TEXTURE is a fundamental property of images. It is composed of repetitive patterns and represents perceptually homogeneous regions. It has been studied in the field of visual perception and computer vision. We can find many methods of analysis in the literature that describe texture in terms of its intrinsic features. However, texture may appear in many different ways, therefore, the algorithms should also take into account the purpose for which an image is used. Commonly, texture is employed in early stages of visual information processes such as segmentation (see Chapter 4) and classification [86].

The goal of texture analysis is to provide a mathematical description of the spatial behavior of intensity values in any given region. The first step in texture analysis is the **feature extraction** where salient characteristics are computed for further computer vision tasks. To facilitate this step, good descriptors are needed; a key property of a desired descriptor is its robustness to environmental changes including both geometrical and photometric transformations.

There have been studies about robust texture descriptors that are invariants to geometrical transformations such as planar rotation, translation, and scaling. Such descriptors are based mainly on a statistical analysis of texels in the spatial domain [87, 88]. Other approaches have been proposed in the spectral domain, nevertheless, the majority of such methods have not been capable to perform well for real-world textures, they are computationally too complex to meet real-time requirements.

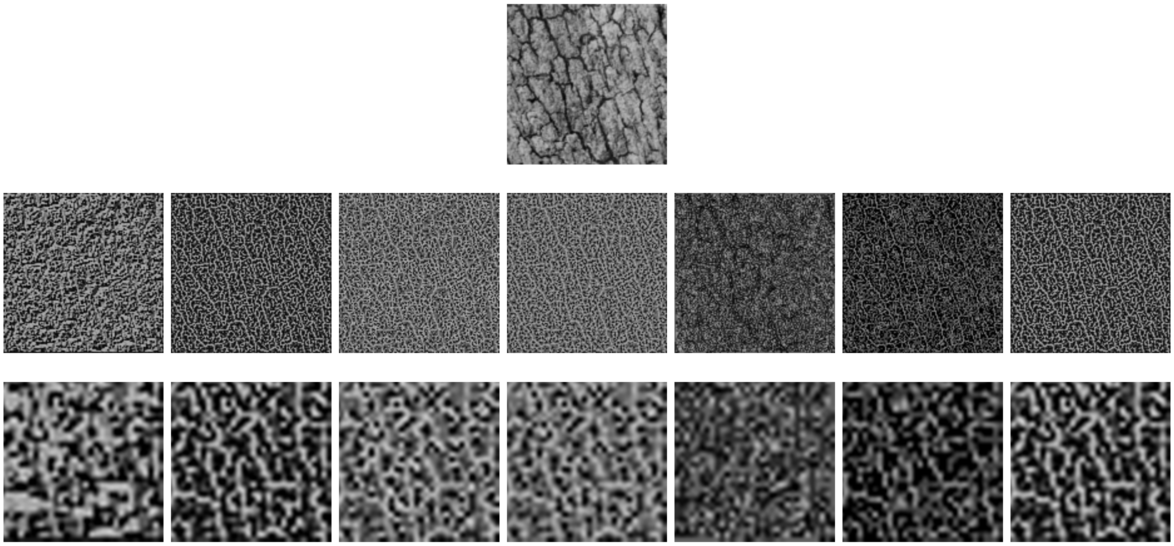


Figure 5.1: *Many non-parametric local transformations have emerged in the literature, most of them are based on the paper of Ojala et al. We present seven examples using the bark texture (D12)* ***(first row)****. LBP, $LBP^{min}_{P,R}$, $LBP^{dom}_{P,R}$, $LBP^{uni}_{P,R}$, $LBP^{num}_{P,R}$, $LBP^{ni}_{P,R}$, and $LBP^{med}_{P,R}$ labeled textures* ***(second row)****. The results were magnified by a factor of 4x for a better visualization* ***(third row)****.*

Recent trends in texture analysis attempt to unify the concepts of statistical and structural approaches, see Section 4.1. Ojala et al. [74] have addressed the problem correctly and observed that these two concepts have complementary characteristics that allow to model texture as a distribution. They recovered the Wang and He's work [89] and proposed a two-level local operator, which compares values within a square mask against their central pixel. This operator belongs to a group of non-parametric local transformations, which are distinguished by the use of ordered information among data. They rely on the relative order of pixel values and transform an image into an array of integer labels, see Fig. (5.1). The pixel-wise information of textures is encoded as a histogram that can be interpreted as the fingerprint of the analyzed object.

In this direction, Zabih and Woodfill [90] proposed two non-parametric local transforms. The first transform, called Rank Transform , is defined as the number of pixels within a local square region that are lesser than their central pixel value. The second one, named Census Transform , maps a local square neighborhood into a bit string that represents the set of neighbor pixels that are lesser than their central pixel value. Both RT and CT depend solely on a set of pixel comparisons. Nevertheless, a limitation of these type of methods is that the amount of information associated to a pixel is not very large, which induces noise sensitivity. Another limitation is that local measures rely heavily upon the intensity of a central pixel. However, the last drawback can be avoided by doing comparisons using local means or median values instead of central pixel intensities.

In this chapter we present a detailed overview of the original work of Ojala et al, Section 5.2. We analyzed the most significant extensions that have been proposed in the literature from theoretical to a practical perspective, for a recent review see [91]. In Section 5.2.3 we performed the assessment of seven algorithms in presence of rotational changes, noise degradations, contrast information, and different sizes of masks using the USC-SIPI database [92]. This study included classification tests using the Kullback-Leibler distance. The second part of this chapter is dedicated to texture retrieval applications. Finally, the results are shown in Section 5.3.1.

5.2 Local binary patterns

The Local Binary Pattern (LBP) descriptor [74] was proposed by Ojala et al. It is based on the idea that textural properties within an homogeneous region can be mapped into patterns that represent micro-features. The original method uses a fixed rectangular 3 × 3 mask called **texture spectrum**, which represents a square neighborhood around a central pixel. The values within the rectangular mask are compared with their central pixel; those ones lesser than the central value are labeled with “0” otherwise with “1.” The labeled pixels are multiplied by a weighting function according with their positions to form a pattern chain. Afterward, the sum of the eight pixels replaces the value of the central pixel, see Fig. 5.2.

This method describes differences of intensities and produces 2^8 labels. After this process is completed for the whole image, a histogram is computed so that can be interpreted as a fingerprint of the analyzed object. One advantage is that this method has very low-computational and spatial complexity. Ojala et al. have claimed that this type of threshold operations provide a robust way for describing local texture patterns.

Although this method provides information about local spatial structures, it is not invariant to rotational changes and does not include contrast information, which has been demonstrated to be crucial to improve the discrimination of some textures. Tan et al. [93] have revisited the approach and demonstrated that a generalization of the LBPs called Local Ternary Patterns is more discriminant and less sensitive to noise for texture analysis. After the initial LBP proposal, many modifications and improvements have emerged in the literature, most of them are related to face analysis where it is assumed that input faces are registered. For a thorough description of LBP operators see two recent surveys and a book monograph [73, 94, 95].

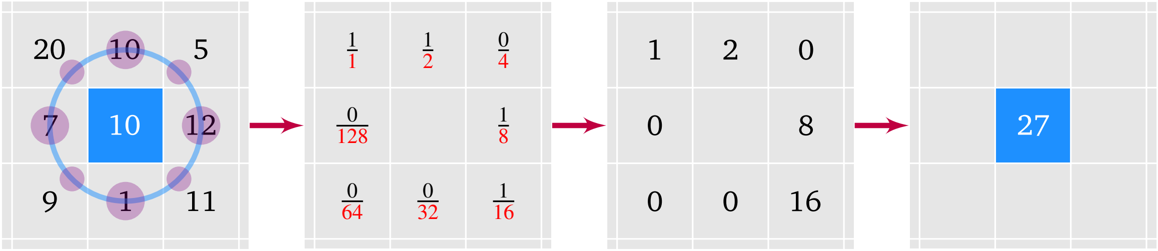


Figure 5.2: *Based on a 3 × 3 rectangular mask, the LBP algorithm computes comparisons between a central pixel and its surrounding neighbors. In this example, the central value is* $p_c = 10$ *and the final label is* 27*.*

5.2.1 MODIFICATIONS OF THE ORIGINAL LOCAL BINARY PATTERN IMPLEMENTATION

In this section, a large set of LBP methods was analyzed. We summarized the main characteristics of the selected algorithms. In particular, we focused our study on schemes that provide rotational invariance because their direct implication in real-world textures, which can occur at arbitrary orientations and they may be subjected to varying illumination conditions.

- In [96], Ojala et al. proposed a generalization of their own operator called **Circular-LBP**. Such generalization uses a circular neighborhood instead of a fixed rectangular region. The sampling coordinates of the neighbors are calculated using the expression: $(x_p, y_p) = (x_c + R\cos[\frac{2\pi p}{P}], y_c - R\sin[\frac{2\pi p}{P}])$; if a coordinate does not fall at an integer position, then the intensity values are bilinearly interpolated. This improvement is denoted by the subscript (P, R) and allows to choose the spatial resolution (R) and the number of sampling points (P) as follows:

  $$LBP_{P,R}(g_c) = \sum_{p=0}^{P-1} s\left(g_p - g_c\right) 2^p \tag{5.1}$$

  where g_c, is the central pixel at (x_c, y_c) and $\left\{g_p \middle| p = 0, \ldots, P-1\right\}$ are the values of the neighbors.

  The Heaviside function is defined as:

  $$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

  Eq. (5.1) represents a “texture unit” composed of $P + 1$ elements (central pixel included). In total, there are 2^P possible texture units that describe spatial patterns in a neighborhood of P points.

  $LBP_{P,R}$ achieves invariance against any monotonic transformations by considering the sign of the differences in $s\left(g_p - g_c\right)$. Note that $LBP_{8,1}$ is approximately equivalent to the original 3×3 square LBP.

- In order to minimize the effects of rotation, Pietikäinen et al. [97] proposed a modification called **Rotational-Invariant-LBP**($LBP^{min}_{P,R}$). The main idea is to apply a circular shift to a pattern chain to find the minimum integer value that it may represents. An arbitrary number of binary shifts is made until the value matches one of the 36 different patterns when using a neighborhood size of $P = 8$.

  $$LBP^{min}_{P,R}(g_c) = min\left\{ROR\left(LBP_{P,R}(g_c), i\right) \middle| i = 0, \ldots, P-1\right\} \tag{5.3}$$

  where $ROR(x,\ i)$ performs a circular bit-wise right shift operation i times. Nevertheless, this method achieves rotational invariance only when a rotation of 90° is applied.

- The authors of the original LBP observed that over 90% of texture patterns can be described with a few spatial transitions, which are the changes (0/1) in the labeled LBP [96]. Therefore, they introduced a measure of uniformity, $U\left(LBP_{P,R}\left(g_c\right)\right)$, that describes fundamental features such as bright spots, flat areas, and edges. For instance, when $P = 8$, there are 9 uniform patterns out of 36 unique rotational invariant patterns.

  $$U\left(LBP_{P,R}\left(g_c\right)\right) = \left|s\left(g_{P-1} - g_c\right) - s\left(g_0 - g_c\right)\right| + \sum_{p=1}^{P-1} \left|s\left(g_p - g_c\right) - s\left(g_{p-1} - g_c\right)\right| \tag{5.4}$$

  which corresponds to the number of spatial transitions. So that, the **Uniform-LBP** ($LBP_{P,R}^{uni}$) can be obtained as:

  $$LBP_{P,R}^{uni}\left(g_c\right) = \begin{cases} \sum_{p=0}^{P-1} s\left(g_p - g_c\right) & \text{if } U\left(LBP_{P,R}\left(g_c\right)\right) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \tag{5.5}$$

  Note that non-unifom patterns are grouped in a unique bin and are consider as noise.

- Yan Ma [98] proposed the **Number-LBP** ($LBP_{P,R}^{num}$) as an extension of the Eq. (5.5) by dividing non-uniform patterns into groups based on the number of “1” or “0” bits as follows:

  $$LBP_{P,R}^{num}\left(g_c\right) = \begin{cases} \sum_{p=0}^{P-1} s\left(g_p - g_c\right) & \text{if } \; U\left(LBP_{P,R}\left(g_c\right)\right) \leq 2 \\ Num_1\left\{LBP_{P,R}\left(g_c\right)\right\} & \text{if } \; \begin{array}{l} U\left(LBP_{P,R}\right) > 2 \quad \text{and} \\ Num_1\left\{LBP_{P,R}\left(g_c\right)\right\} \geq Num_0\left\{LBP_{P,R}\left(g_c\right)\right\} \end{array} \\ Num_0\left\{LBP_{P,R}\left(g_c\right)\right\} & \text{if } \; \begin{array}{l} U\left(LBP_{P,R}\right) > 2 \quad \text{and} \\ Num_1\left\{LBP_{P,R}\left(g_c\right)\right\} < Num_0\left\{LBP_{P,R}\left(g_c\right)\right\} \end{array} \end{cases} \tag{5.6}$$

  where $Num_1\{\bullet\}$ is the number of “1” and $Num_0\{\bullet\}$ is the number of “0” in the non-uniform pattern.

- Liu et al. [99] stated that the likelihood of a central pixel only depends on its neighbors. Hence, they proposed the **Neighbor-intensity-LBP** ($LBP_{P,R}^{ni}$), which can be obtained by replacing the value of the central pixel with the average of its neighbors as follows:

  $$LBP_{P,R}^{ni}\left(g_c\right) = \sum_{p=0}^{P-1} s\left(g_p - \mu\right) 2^p \tag{5.7}$$

  where

  $$\mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \tag{5.8}$$

  is the average of the P neighbors.

- The presence of noise can seriously impair the performance of LBPs. The Zabih and Woodfill's proposal [90] named **Median-LBP**, replaces the central pixel with the median of itself and its P neighbors as follows:

  $$LBP_{P,R}^{med}\left(g_c\right) = \sum_{p=0}^{P-1} s\left(g_p - \widetilde{g}\right) \tag{5.9}$$

  where $\widetilde{g}$ represents the median of the P neighbors and the central pixel.

  This modification is still invariant to rotation but less sensitive to noise. It is also invariant to monotonic illumination changes.

- Fu and Wei [100] addressed the problem of noise by considering that in most cases a central pixel provides more information than their neighbor counterparts. Thus, they redefined Eq. (5.2):

  $$s\left(x\right) = \begin{cases} 1 & |x| \geq c \\ 0 & |x| < c \end{cases} \tag{5.10}$$

  where c is a fixed threshold.

  With this modification, they proposed the **Centralized-LBP**($LBP_{P,R}^{cen}$) as follows:

  $$LBP_{P,R}^{cen}\left(g_c\right) = \sum_{p=0}^{\frac{P}{2}-1} s\left(g_p - g_{p+\frac{P}{2}}\right) 2^p + s\left(g_c - g_{tot}\right) 2^{\frac{P}{2}} \tag{5.11}$$

  and g_{tot} is defined as:

  $$g_{tot} = \frac{1}{P+1}\left(g_c + \sum_{p=0}^{P-1} g_p\right) \tag{5.12}$$

  The algorithm considers correlations between opposite pixel points, therefore, it is not invariant to rotational changes.

- In [101], Tan et al. proposed an extension called **Extended-LBP** ($LBP_{P,R}^{ext}$) by using the value of a central pixel plus a tolerance interval t as a local threshold. t is a user-specific value, usually set at "1".

  Each pixel in the interval $g_c \pm t$ is quantized as zero. Pixels above the tolerance interval are labeled with "1" and those ones below the interval are labeled with "–1" as follows:

  $$s\left(x\right) = \begin{cases} 1 & \text{if } x > t \\ 0 & \text{if } |x| \leq t \\ -1 & \text{if } x < -t \end{cases} \tag{5.13}$$

  x is the difference between the P neighbors and their central pixel.

  Each ternary pattern is split into upper and lower patterns and each pattern is encoded as a separate LBP. Finally, the histograms are concatenated.

- Referring to Fig. 5.2, it is simple to compute the differences between a given central pixel, g_c, and its P neighbors as $d_p = \left\{ g_p - g_c \middle| p = 0, 1, \ldots, P-1 \right\}$. Guo et al. [102] suggested to consider both the sign and magnitude of d_p to form the **Completed-LBP** ($LBP^{com}_{P,R}$). d_p is split into two components as follows:

  $$d_p = s_p * m_p = \begin{cases} s_p & = sign\left(d_p\right) \\ m_p & = \left|d_p\right| \end{cases} \tag{5.14}$$

  where s_p and m_p are the sign and magnitude of d_p respectively.

- Liao et al. [103] proposed the **Dominant-LBP** ($LBP^{dom}_{P,R}$), which is a modification of Eq. (5.5). The authors are based on the fact that in practice $LBP^{uni}_{P,R}$ is not well suited to encode some complicated pattern textures such as curvature edges and crossing boundaries of corners. A possible explanation is that the extracted uniform patterns do not have a dominant proportion to represent an object (or image).

  Liao et al. have shown that given a set of training images, the required number of patterns that better represent textures corresponds to at least 80% of the pattern occurrences. This procedure guarantees a suitable framework for representing textures, see Liao et al. pseudo-algorithm in Algorithm 5.1.

Algorithm 5.1: Liao's algorithm for calculating the Dominant-LBP

```
Input: A := An image
Output: DOMH := $LBP^{dom}_{P,R}$ histogram
begin
    AH := compute $LBP^{uni}_{P,R}$ histogram of A;
    SAH := sort AH in descending order;
    DOMH := extract 80% of pattern occurrences;
end
```

5.2.2 ADDING CONTRAST INFORMATION TO LBPS

Previously in [104], Ojala et al. studied the use of a joint representation $LBP_{P,R}/VAR_{P,R}$, where $VAR_{P,R}$ represents the local variance. However, $VAR_{P,R}$ is a continuous signal, hence, it has to be quantized. On the other hand, Guo et al. [102] included complementary information of local contrast in a new scheme called **Local Binary Pattern Variance** ($LBPV_{P,R}$).

So, a rotation invariant measure of local variance can be defined as:

$$VAR_{P,R}\left(g_c\right) = \frac{1}{P} \sum_{p=0}^{P-1} \left(g_p - u\right)^2 \tag{5.15}$$

where $\left\{ g_p \middle| p = 0, \ldots, P-1 \right\}$ are the neighbors of g_c and $u = \frac{1}{P} \sum_{p=0}^{P-1} g_p$.

Since LBP histograms do not include information of variance, $VAR_{P,R}$ can be used as an adaptive weight to adjust the contribution of the LBP code in the histogram calculation. The $LBPV_{P,R}$ descriptor offers a solution for that as follows:

$$LBPV_{P,R}(k) = \sum_{i=1}^{N} \sum_{j=1}^{M} \omega\left(LBP_{P,R}(i,j), k\right), k \in [0, K] \tag{5.16}$$

and

$$\omega\left(LBP_{P,R}(i,j), k\right) = \begin{cases} VAR_{P,R}(i,j) & LBP_{P,R}(i,j) = k \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

The advantage of a conjoint scheme, $LBP_{P,R}/VAR_{P,R}$, is the use of both local spatial and contrast information simultaneously. Variance quantization can be done by distributing all the quantized variance values into a histogram with the same number of bins that the LBP histogram. It is important to consider that few bins will fail to provide enough discriminative information while too many bins may lead to sparse and unstable histograms.

5.2.3 EXPERIMENTAL RESULTS

We conducted experimental tests using a set of rotated texture images from the USC-SIPI image database, available at [92]. This set consists of thirteen grayscale textures of 512 × 512 pixels, however, for purposes of efficiency, the images were rescaled at 64 × 64 and 128 × 128 pixels with 8 bits/pixel using bicubic interpolation.

The textures bark (D12), brick (D94), bubbles (D112), grass (D9), leather (D24), pigskin (D92), raffia (D84), sand (D29), straw (D15), water (D38), weave (D16), wood (D68), and wool (D19) –the number between parenthesis is the identification number in the Brodatz texture book [85]– were digitized at seven different rotation angles: {0, 30, 60, 90, 120, 150, 200} degrees. As the main advantage the USC-SIPI dataset provides a hardware-rotated subset of textures avoiding in this way the introduction of artifacts.

It is possible to distinguish two types of methods: Rotational-Invariant-LBP, Neighbor-intensity-LBP, Median-LBP, and Dominant-LBP are modifications or extensions based on the Circular-LBP. Such methods compute labels based on pattern chain values. On the other hand, Number-LBP is a refined model of the Uniform-LBP and computes labels based on the number of transitions in a pattern chain.

The first experiment consisted in computed the Normalized Measure of Dispersion C_v of each LBP image, then we compared the C_v of the reference images (non-rotated image) against its corresponding six rotated versions.

$$C_v = \frac{\sigma}{\mu} \tag{5.18}$$

where σ and μ represent the standard deviation and the mean respectively.

Table 5.1: *The slight numerical variations of the C_v indicates the robustness of the extracted features. The lower the coefficient of variation the higher the robustness to rotational changes.*

	LBP^{min}	$LBP^{min}_{P,R}$	$LBP^{ni}_{P,R}$	$LBP^{med}_{P,R}$	$LBP^{dom}_{P,R}$	$LBP^{uni}_{P,R}$	$LBP^{num}_{P,R}$
μ_{64}	0.0104	0.0057	0.0162	0.0100	0.0100	0.0071	**0.0049**
σ^2_{64}	**0.0094**	0.0175	0.0321	0.0350	0.0350	0.0303	0.0329
μ_{128}	0.0089	0.0053	0.0089	0.0050	0.0050	0.0046	**0.0033**
σ^2_{128}	**0.0086**	0.0122	0.0222	0.0299	0.0299	0.0275	0.0241

Table 5.1 summarizes the statistical variations of C_v among all the orientations and operators. μ_{64} and σ^2_{64} represent the mean and the variance of C_v in images of 64×64 pixels, while μ_{128} and σ^2_{128} represent the mean and the variance of C_v in images of 128×128 pixels, respectively. A robust rotational invariant representation should provide a C_v with no variations namely the best descriptor is the one that generates the lowest variance. We should note that LBP^{min} and $LBP^{min}_{P,R}$ differ in that the former does not use interpolated neighbors but the latter does. In this experiment the interpolated neighbors minimized the mean μ but increased the variance σ^2.

5.2.4 KULLBACK-LEIBLER DISTANCE

Since a LBP histogram acts as a fingerprint of a texture, it is possible to consider a similarity measures to assess distances among all different textures. Although the Kullback-Leibler divergence (KL) –a generalization of the Shannon's entropy– is not a true metric rather it is a relative entropy measure, it can be used as a suitable descriptor for measuring distances between histograms or feature vectors.

The distance between two images with their corresponding histograms is defined as follows:

$$D_{KL}(A, B) = \sum_{i=0}^{b-1} P_i(B) \log \frac{P_i(B)}{P_i(A)} \tag{5.19}$$

where A and B are histograms with b-bins length each and P_i denotes the probability of the i-bin.

The classification procedure setup consisted in comparing the histogram of each reference pattern against all the rotated images. The distances were sorted in an increasing order and the closest one was retrieved. This experiment was performed for images of 64×64 and 128×128 pixels, (see Table 5.2).

The classification rates are consistent with those reported in the literature. Pietikäinen [97] reported an error rate of 38.5% for the LBP^{min} algorithm. Here, we reached 38.46%. The best rate was achieved with the $LBP^{dom}_{P,R}$, one possible reason is that this approach rules out 20% of patterns that in many cases can be interpreted as noise.

Table 5.2: *Comparison of seven descriptors using the KL distance. # textures is the number of images correctly classified from a total of* 91 *textures.*

Scheme	64 × 64		128 × 128		References
	# textures	Accuracy (%)	# textures	Accuracy (%)	
LBP^{min}	56	61.54	66	72.53	[97]
$LBP^{min}_{P,R}$	53	58.24	63	69.23	[97]
$LBP^{ni}_{P,R}$	57	62.64	65	71.43	[99]
$LBP^{med}_{P,R}$	54	59.34	71	78.02	[90]
$LBP^{dom}_{P,R}$	**68**	**74.73**	**73**	**80.22**	[103]
$LBP^{uni}_{P,R}$	67	73.63	**73**	**80.22**	[96]
$LBP^{num}_{P,R}$	63	69.23	**73**	**80.22**	[98]

5.2.5 NOISE SENSITIVITY

LBP is very sensitive to noise, specially when a small neighborhood is used. Since the amount of information associated to a pixel is not very large, even a small change in any pixel could lead to a different label. Table 5.3 shows the classification performance using noisy images of 128 × 128 pixels. Seven algorithms were evaluated with the addition of Gaussian noise with mean $\mu = 0$ and standard deviation $\sigma = 0.1$ and under the effects of Poisson noise. Noise was added using the Matlab *imnoise* function. According to Table 5.3 and the images analyzed, $LBP^{dom}_{P,R}$ performed best for Gaussian noise. This operator discards 20% of random patterns that in many cases are considered noise. On the other hand, $LBP^{med}_{P,R}$ reached the higher rate for textures with Poisson noise. $LBP^{med}_{P,R}$ replaces central pixels with the median of the P neighbors and the central pixel. This procedure reduces the effects of noise such as Salt and Pepper and Poisson.

Table 5.3: *Performance of seven LBP descriptors under additive Gaussian noise with media* $\mu = 0$ *and* $\sigma^2 = 0.1$ *and Poisson noise. # textures is the number of images correctly classified from a total of* 91 *textures.*

Scheme	Gaussian noise		Poisson noise	
	# textures	Accuracy (%)	# textures	Accuracy (%)
LBP^{min}	55	60.44	71	78.02
$LBP^{min}_{P,R}$	52	57.14	66	72.53
$LBP^{ni}_{P,R}$	51	56.04	73	80.22
$LBP^{med}_{P,R}$	53	58.24	**74**	**81.32**
$LBP^{dom}_{P,R}$	**61**	**67.03**	69	75.82
$LBP^{uni}_{P,R}$	60	65.93	69	75.82
$LBP^{num}_{P,R}$	51	56.04	51	56.04

5.2.6 ADDING VARIANCE

Illumination variation is one of the most important challenges for the current descriptors. Tan et al. [93] claim that LBP performance decreases almost exponentially under extreme conditions. The LBP by itself is only invariant to monotonic illumination changes and does not entail the contrast information of textures, which is important in the discrimination process.

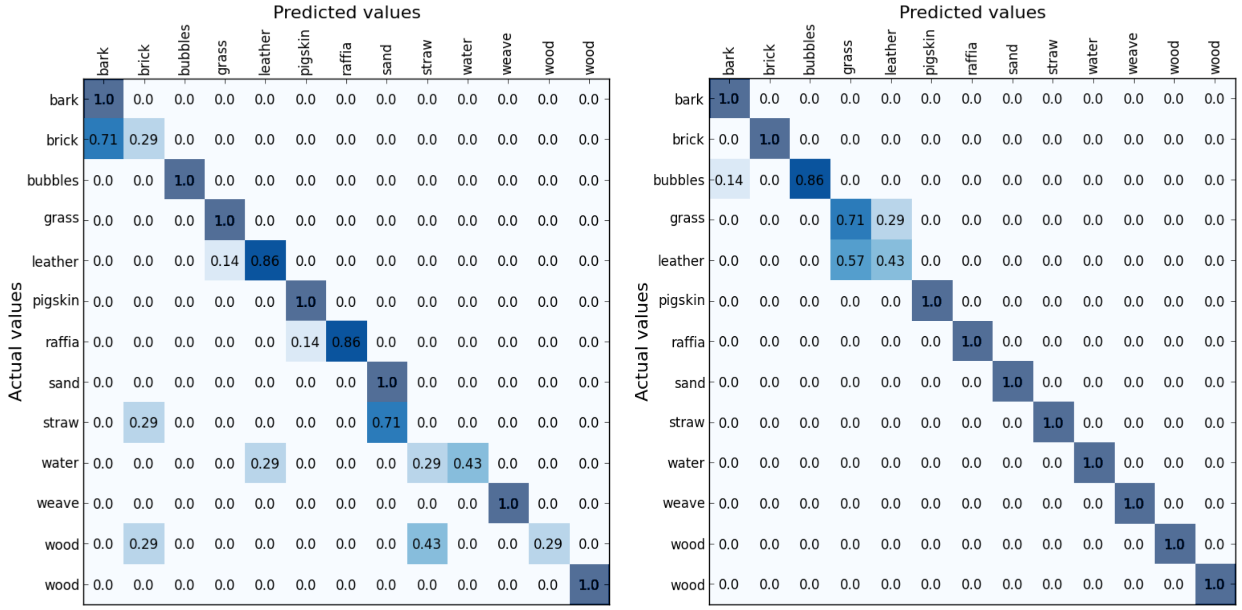
We used Eq. (5.16) to compute the local variance of the test images. We are interested in combined information of LBPs and local variance. However, $VAR_{P,R}$ produces continuous values that should be quantized. Ojala et al. [104] proposed to quantize variance values so that all bins in the histogram have an equal number of elements. So far, establishing the number of bins is still an open issue. LBP and $VAR_{P,R}$ histograms could be combined in two ways, jointly or mixed [105]. In the former, similar to 2D joint histograms, we can build a 3D joint histogram, while in the latter, a large histogram is built by concatenating both LBP and $VAR_{P,R}$ histograms to form the so-called "pseudo-joint histogram."

Table 5.4 shows classification rates of seven descriptor with and without $VAR_{P,R}$ information on images of size 128×128. We used the pseudo-joint histogram approach. As we expected, the results obtained with joint pairs of features provided the best performance with error rates around 7.70%. This fact emphasizes the importance of using other features besides LBP information. In fact, if we consider only the local variance as a feature descriptor, the result reached 86.81% of textures correctly classified.

In Fig. 5.3 we present a comparison between $LBP^{uni}_{P,R}$ and $LBP^{uni}_{P,R} + VAR_{P,R}$ confusion matrices. There is a strong indication that local variance is more discriminant than LBP features themselves in the classification process. A Fisher discriminant score [106] will allow to select the most informative features by rejecting those noisy features.

Table 5.4: *Comparison of texture classification including local variance information. # textures is the number of images correctly classified from a total of 91 textures.*

Scheme	LBP		$LBP + VAR_{P,R}$	
	# textures	Accuracy (%)	# textures	Accuracy (%)
LBP^{min}	66	72.53	76	83.51
$LBP^{min}_{P,R}$	63	69.23	81	89.01
$LBP^{ni}_{P,R}$	65	71.43	81	89.01
$LBP^{med}_{P,R}$	71	78.02	80	87.91
$LBP^{dom}_{P,R}$	**73**	**80.22**	81	89.01
$LBP^{uni}_{P,R}$	**73**	**80.22**	**84**	**92.30**
$LBP^{num}_{P,R}$	**73**	**80.22**	82	90.10



	bark	brick	bubbles	grass	leather	pigskin	raffia	sand	straw	water	weave	wood	wood
bark	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
brick	0.71	0.29	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
bubbles	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
grass	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
leather	0.0	0.0	0.0	0.14	0.86	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
pigskin	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
raffia	0.0	0.0	0.0	0.0	0.0	0.14	0.86	0.0	0.0	0.0	0.0	0.0	0.0
sand	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
straw	0.0	0.29	0.0	0.0	0.0	0.0	0.0	0.71	0.0	0.0	0.0	0.0	0.0
water	0.0	0.0	0.0	0.0	0.29	0.0	0.0	0.0	0.29	0.43	0.0	0.0	0.0
weave	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
wood	0.0	0.29	0.0	0.0	0.0	0.0	0.0	0.0	0.43	0.0	0.0	0.29	0.0
wood	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

(a)

	bark	brick	bubbles	grass	leather	pigskin	raffia	sand	straw	water	weave	wood	wood
bark	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
brick	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
bubbles	0.14	0.0	0.86	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
grass	0.0	0.0	0.0	0.71	0.29	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
leather	0.0	0.0	0.0	0.57	0.43	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
pigskin	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
raffia	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
sand	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
straw	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
water	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
weave	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
wood	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
wood	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0

(b)

Figure 5.3: *Performance comparison between **(a)** $LBP^{uni}_{P,R}$ and **(b)** $LBP^{uni}_{P,R} + VAR_{P,R}$. In most of the cases, the variance information minimized the error rate except for the "grass" class. In general, the variance information increases up to 12% the classification rate.*

5.2.7 NEIGHBORHOOD SIZE

Another important issue of the LBPs is the neighborhood size. The next experiment was aimed to assess the radius size influence in texture classification. Table 5.5 presents the classification performance of the seven LBP approaches with radius $R = \{1, 2, 3\}$ on images of size 128×128. The highest classification rate was achieved with $LBP^{dom}_{P,R}$ and $R = 2$, while higher radius size caused poor classification rates.

Table 5.5: *Comparison of texture classification using different neighborhood size from a total of* 91 *textures:* $R = \{1, 2, 3\}$.

Scheme	$R = 1$	Accuracy (%)	$R = 2$	Accuracy (%)	$R = 3$	Accuracy (%)
$LBP^{min}_{P,R}$	63	69.23	**71**	**78.02**	56	61.54
$LBP^{ni}_{P,R}$	65	71.43	**71**	**78.02**	70	76.92
$LBP^{med}_{P,R}$	**71**	**78.02**	64	70.33	57	62.64
$LBP^{dom}_{P,R}$	73	80.22	**74**	**81.32**	67	73.63
$LBP^{uni}_{P,R}$	**73**	**80.22**	72	79.12	70	76.92
$LBP^{num}_{P,R}$	**73**	**80.22**	68	74.73	63	69.23

5.3 Image retrieval

In the previous Section 5.2 we presented several texture classification experiments based on a procedure consisted in comparing distances between histograms; such distances are sorted in increasing order and the closest one is retrieved. This is the simplest case of the Content-based Image Retrieval (CBIR) approach, which has become an active research area due to the massive amount of digital image collections. CBIR extracts visual information of an image such as color [107] or texture [108] and its goal is to retrieve the closest images from a data bank using features that best describe the scene in an **image query** [109].

CBIR uses image features to catch similarities among images, therefore, feature extraction is a crucial stage. Theoretically, having more features implies a greater ability to discriminate images. However, this is not always true because not all features are important for understanding or representing a visual scene [110].

Although spectral methods are computationally complex to meet real-time requirements, they have proven to be powerful tools for characterizing textures, (see Chapter 3). Such methods collect a distribution of filter responses and extract features from the first and second order statistics [111]. Especially, the use of Gabor filters in texture analysis was motivated because the studies of Daugman on visual modeling of simple cells. He found that the experimental findings on orientation selectivity of visual cortical neurons were previously observed by Hubel and Wiesel in cats [25, 52, 112].

In [113], Manjunath and Ma proposed a method for texture analysis where input images are filtered using a set of Gabor filters and the mean and the standard deviation are taken to build a feature vector or image query. This method is generally accepted as a benchmark for **texture retrieval**. Nevertheless, as we mentioned earlier, Gabor filters have not zero mean, this fact leads to non-uniform coverage of the Fourier domain; the distortion may cause fairly poor pattern retrieval accuracy.

We propose a simple yet efficient image retrieval approach, see Fig. (5.4), based on a novel log-Gabor filter scheme, see Section 3.3. The first step is to compute the image coefficients, $C_{(S,\Theta)}$, as follows:

$$C_{(S,\Theta)}(x,y) = \mathcal{F}^{-1}\left\{\hat{I}(u,v) \bullet \hat{G}_{S,\Theta}(\rho,\theta)\right\} \tag{5.20}$$

where $\hat{I}(u,v)$ is the given image in the Fourier domain and $\hat{G}_{S,\Theta}(\rho,\theta)$ is the log-Gabor filter, see Eq. (3.13), at the scale S and orientation Θ.

The coefficients represent texture characteristics in a particular scale and orientation, thus, energy signatures such as mean, $\mu_{(s,\theta)}$, and variance, $\sigma^2_{(s,\theta)}$, that belong to the coefficient $C_{(S,\Theta)}$ can be used as texture features for constructing an image query as follows:

$$\bar{t} = \left[\mu_{(0,0)}, \sigma^2_{(0,0)}, \ldots, \mu_{(S-1,\Theta-1)}, \sigma^2_{(S-1,\Theta-1)}\right] \tag{5.21}$$

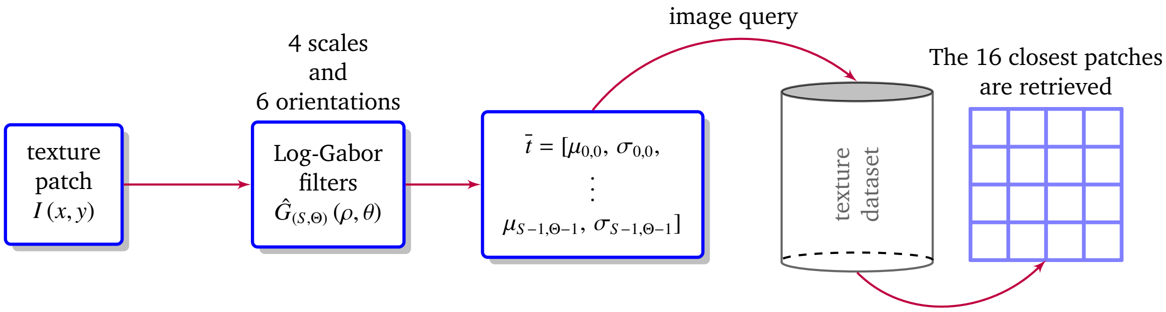


Figure 5.4: *Block diagram of our proposal. After the extraction of the mean and variance from each log-Gabor coefficient, we can build an image query for retrieving the 16 closest patches in the dataset.*

We used, in Section 5.2.4, the KL divergence as a similarity measure. In addition, we propose the Jensen-Shannon (JS) divergence [114] for evaluating distances between two textures A and B. The advantage of JS against KL divergence is that JS is a true metric. The JS divergence, denoted by ψ is defined as follows:

$$\psi = \sqrt{2D_{JS}(A,B)} \tag{5.22}$$

where

$$D_{JS}(A,B) = \frac{1}{2}D_{KL}\left(A, \frac{A+B}{2}\right) + \frac{1}{2}D_{KL}\left(B, \frac{A+B}{2}\right) \tag{5.23}$$

5.3.1 EXPERIMENTAL RESULTS

To validate our proposal, we used a set of twenty grayscale textures of 512 × 512 pixels from the USC-SIPI dataset [92] and provided an experimental evaluation composed of the following stages: (i) each texture was divided into sixteen non-overlapping patches of 128 × 128 pixels in order to build a database of 320 images; (ii) each patch was processed with a log-Gabor filter bank (4 scales and 6 orientations); the additional parameters that fit biological profiles were borrowed from [59]; (iii) for each coefficient, the mean and the variance were calculated and concatenated in order to build the corresponding feature vector. We must note that Manjunath and Ma used the mean and the standard deviation, however, we used the mean and the variance because they improved the retrieval performance.

In this way, we got 320 feature vectors of 48-bins length each. Each feature vector was used as a query pattern to calculate distances among the patches. The distances were sorted in increasing order and the closest sixteen patches were retrieved. Since each texture was divided into 16 patches, in the best-case scenario, a single query should return all the 16 patches that belong to the same texture; furthermore, by repeating this operation with the rest of the patches, the algorithm should retrieved 256 patches per texture. In order to measure the performance of our proposal we followed the recommendations in [113, 115] and computed the Retrieval Accuracy (RA), which is the standard metric for evaluating CBIR systems and is listed in Table 5.6.

Table 5.6: *The RA was computed for both Gabor and log-Gabor filters with D_{KL} and D_{JS} distances. D* means the Brodatz texture and # indicates the number of patches correctly retrieved.*

	Gabor filters				Log-Gabor filters			
	D_{KL}		D_{JS}		D_{KL}		D_{JS}	
	#	RA (%)	#	RA (%)	#	RA (%)	#	RA (%)
D1	256	100.00	256	100.00	256	100.00	256	100.00
D3	186	72.65	181	70.70	177	69.14	175	68.35
D4	256	100.00	256	100.00	256	100.00	256	100.00
D5	220	85.93	224	87.50	227	88.67	229	89.45
D6	256	100.00	256	100.00	256	100.00	256	100.00
D9	231	90.23	233	91.01	254	99.21	254	99.21
D10	191	74.60	194	75.78	217	84.76	221	86.32
D11	256	100.00	256	100.00	256	100.00	256	100.00
D15	183	71.48	179	69.92	194	75.78	187	73.04
D20	256	100.00	256	100.00	256	100.00	256	100.00
D24	240	93.75	238	92.96	243	92.92	242	94.53
D26	256	100.00	256	100.00	256	100.00	256	100.00
D56	256	100.00	256	100.00	253	98.82	256	100.00
D66	178	69.53	178	69.53	200	78.12	202	78.90
D93	231	90.23	234	91.40	242	94.53	244	95.31
D104	256	100.00	256	100.00	256	100.00	256	100.00
D105	136	53.12	136	53.12	205	80.07	203	79.29
D106	134	52.34	136	53.12	173	67.57	177	69.14
D109	163	63.67	163	53.67	227	88.67	227	88.67
D112	200	78.12	198	77.34	190	74.21	191	74.60
μ	–	**84.79**	–	**84.80**	–	**89.73**	–	**89.84**

Note that the lowest rate was achieved with the $D106$ texture and Gabor filters; the RA corresponds to 52.34% and 53.12% using D_{KL} and D_{JS} metrics respectively. On the contrary, with the same texture, the log-Gabor scheme achieved 67.57% and 69.14% of accuracy with D_{KL} and D_{JS} metrics respectively. These rates resulted in 39 and 41 more patches correctly classified. In general, our proposal achieved an average RA of 89.84%, which represents an increase of 259 patches correctly classified regarding Manjunath and Ma's proposal.

In the ideal case, given a single query, all the sixteen patches that belong to the same texture should be retrieved. An important metric that assesses this specific case is called Full Retrieval Rate (FRR). Our proposal achieves 52.18% of query patterns fully retrieved, it means a 6.56% higher rate compare to the Gabor scheme with 45.62%, (see Table 5.7).

Table 5.7: *FRR for Gabor and log-Gabor filters. Given a single query (patch), all the sixteen patches that belong to the same texture are retrieved. # indicates the number of patches fully retrieved.*

	Gabor filters		Log-Gabor filters	
	#	FRR (%)	#	FRR (%)
D_{KL}	146	45.62	**165**	**51.56**
D_{JS}	147	45.93	**167**	**52.18**

The Overall Retrieval Rate (ORR) measures the total number of patches correctly retrieved and is presented in Table 5.8. The Gabor scheme achieved 84.78% and 84.80% of patches retrieved correctly with D_{KL} and D_{JS} metrics respectively. On the other hand, our proposal achieved 89.72% and 89.84% of patches retrieved correctly with D_{KL} and D_{JS} metrics respectively. These rates represent increases up to 4.94% using KL divergence and 5.04% using Jensen-Shannon divergence.

Table 5.8: *Total number of patches correctly retrieved for Gabor and log-Gabor schemes.*

	Gabor filters	Log-Gabor filters
	ORR (%)	ORR (%)
D_{KL}	84.78	**89.72**
D_{JS}	84.80	**89.84**

– *Nothing shocks me. I'm a scientist!*

Harrison Ford as Indiana Jones

6

Applications in medical imaging

6.1 Introduction

The continuous development of texture analysis techniques has allowed to increase the information that can be obtained from images. Its goal is to improve visual skills of radiologists by extracting features that may be relevant to more accurate diagnoses. It is noticeable the use of texture patterns in medical imaging to distinguish between pathological and healthy tissue [116]. In recent years, Computed Tomography (CT) have made available a large collection of two-dimensional and three-dimensional images that capture the structure and function of different types of anatomical structures. Therefore, there has been an increased interest to describe, classify, and segment anatomical parts based on texture characteristics.

Along this chapter we addressed the problem of characterization and classification of Chronic Obstructive Pulmonary Disease (COPD), which is a progressive and irreversible lung condition typically related to emphysema. It hinders air from passing through airpaths and causes that alveolar sacs lose their elastic quality, increasing the risk of death. Findings of COPD may be manifested in a variety of CT studies. Nevertheless, visual assessment of CT images is time-consuming and depends on trained observers. Hence, a reliable computer-aided diagnosis system would be useful to reduce time and inter-evaluator variability. COPD describes a collection of lung diseases that are characterized by parenchymal destruction and gradual limitation of airflow. Although it can manifest as either emphysema, chronic bronchitis, or both; the former is the most common pathophysiological manifestation and is mainly attributable to tobacco smoking [117, 118].

Studies of the World Health Organization reports that 65 million people have COPD worldwide and predicts that COPD will be responsible for 10% of the world's mortality by 2030 [119]. Therefore, in order to prevent other health complications such as pneumothorax and respiratory infections, accurate characterization of emphysema is required for the development of efficient treatment options.

Literature [120] recognizes three types of emphysema: (i) **Paraseptal** (PS), also known as distal acinar emphysema, is characterized by destruction of distal airway structures, alveolar ducts, and alveolar sacs. The process is localized around the pleura; (ii) **Panlobular** (PL) or panacinar emphysema destroys uniformly alveoli and prevails in the lower half of the lungs; and (iii) **Centrilobular** (CL) or centriacinar emphysema is the most common type of pulmonary emphysema. It begins in the respiratory bronchioli and spreads peripherally. Most of the damage is usually contained to the upper half of the lungs.

Spirometry is the gold standard criterion to establish a diagnosis of emphysema. It measures the volume of air that a patient is able to expel from lungs after a maximal inspiration. Nevertheless, this method does not allow to discriminate pathological subphenotypes of emphysema. Attenuation values of CT images, which are expressed in the Hounsfield Unit (HU) scale, have been used for identification of pathological changes in lung parenchyma because they are linked to physical density of lung tissue [121]. Hayhurst et al. [122] showed that attenuation values in patients who had CL differed from healthy patients with Normal Tissue (NT). Density mask is another method to quantify emphysema. This technique describes the amount of air presented in a CT image and consists of computing the percentage of pixels with attenuation values lesser than a previously selected threshold. Routinely, –910 HU is taken as the standard threshold but recently Madani et al. [123] observed that if the threshold lies somewhere between -960 to -980 HU then the correlation with emphysema is greater. Mean lung density [124] defined as the percentage of lung below -950 HU is another objective measure of the extent of macroscopic emphysema. All these metrics consider that emphysema causes an abnormal enlargement of air spaces, thus, the air-tissue ratio in an emphysematous lung should increase, whereas density should decrease proportionally to the amount of emphysema. Nevertheless, they are too sensitive to scanner calibration and noise. In addition, they cannot distinguish emphysema patterns due to averaging effect. On the other hand, texture analysis in lung CT images may provide new insights towards the construction of a reliable computer-aided diagnosis system because it is capable of identify changes in lung parenchyma and abnormalities associated with emphysema.

Sørensen et al. [125] combined textural features using LBPs to classify three classes of emphysema and achieved an accuracy above 90%. In [126], fractal analysis was proposed to classify 3258 emphysema patches of size 64 × 64 pixels. A simpler alternative based on kernel density estimation of local histograms was introduced in [127]. A different approach was presented in [128] where the authors used meta-data to label lung samples. In [129] a technique based on the embedded probabilistic PCA was used to classify interstitial lung abnormalities, while in [130], the Riesz transform was presented to obtain features of lung abnormalities.

We propose a novel approach that exploits the advantages of Complex Gabor Filters (CGF), such as the strong correlation with the HVS, and simultaneously encodes local intensity information provided by LBPs (see Chapter 5).

Since low-attenuation areas in lung CT images describe different emphysema patterns, the discrimination problem was focused on the characterization of local intensities and global spatial variations. Our proposal considers these aspects and provides a robust representation for each type of emphysema. Therefore, an improvement in the classification rate can be attained.

This chapter is organized as follows: we described our proposal and the construction of the feature vectors in Section 6.2. In Section 6.3 a set of global and local descriptors is presented. In Section 6.4 we briefly explained the discriminant analysis theory for reducing dimensionality. The data are describe in Section 6.5, while the experiments and results are detailed in Section 6.6. Finally, our work is summarized in Section 6.7.

6.2 A bio-inspired model for feature extraction

We propose the combination of complex Gabor filters and local binary patters for a better characterization of emphysema; the former are global descriptors, whereas the latter are local operators. In order to assign a given patch to one of several patterns, we used a methodology composed of three stages: (i) feature extraction; (ii) dimensionality reduction using Kernel-Fisher Discriminant Analysis (KFDA); and (iii) classification using k-Nearest Neighbor classifier (k-NN) . In the following paragraphs we detailed our proposal.

In the mid-eighties, Daugman found out that the shape of Gabor functions and the psychophysical properties of simple receptive fields have a close match [25, 52, 53]. Furthermore, he proved that the conjoint time-frequency properties of 1D Gabor functions are still satisfied for the two-dimensional case.

We presented in the previous Section 3.2 Gabor filters. They are defined as the product of Gaussian functions and complex sinusoids, (see Fig. 3.2). Since Gabor filters can be divided into two parts: $g_e(x, y)$ and $g_o(x, y)$, we can use them to build CGF. Note that in Section 4.2 we only used the even-symmetric part to build Real Gabor Filters (RGF).

In this chapter we constructed a filter bank of 24 complex filters distributed in 4 scales and 6 orientations. The goal of feature extraction is to identify similar characteristics or patterns that are common to a specific class. Such patterns may vary slightly within the class but they must be sensitive enough to discriminate elements from different classes. Gabor filters perform an analysis in a specific orientation and a frequency band and extract characteristics called complex Gabor coefficients denoted by $C_{(s,\theta)}$.

Each coefficient can be computed as follows:

$$C_{(s,\theta)}(x, y) = \sqrt{E^2_{(s,\theta)}(x, y) + O^2_{(s,\theta)}(x, y)} \tag{6.1}$$

with

$$\begin{aligned} E_{(s,\theta)} &= I \star g_{e(s,\theta)} \\ O_{(s,\theta)} &= I \star g_{o(s,\theta)} \end{aligned} \tag{6.2}$$

where I is the given image and $\star$ indicates the convolution. $g_{o(s,\theta)}$ and $g_{e(s,\theta)}$ are the odd-symmetric and even-symmetric filters at the scale s and orientation θ, respectively.

Theoretically, the more the features, the greater the ability to discriminate images. Nevertheless, this statement is not always true because not all features are important for understanding or representing visual scenes [110]. Our study not only was focused on energy signatures such as the mean and the standard deviation but on higher-order statistics to increase the ability to extract characteristics.

Since Gabor coefficients can be considered as probability density functions, the mean, the standard deviation, the skewness, and the kurtosis are enough to provide a good approximation to them [131]. We investigated the following set of statistics, where M and N are the size of the coefficient.

- Mean

$$\mu_{(s,\theta)} = \frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} C_{(s,\theta)}(x, y) \tag{6.3}$$

- Standard deviation

$$\sigma_{(s,\theta)} = \sqrt{\frac{1}{NM} \sum_{x=1}^{N} \sum_{y=1}^{M} (C_{(s,\theta)}(x, y) - \mu_{(s,\theta)})^2} \tag{6.4}$$

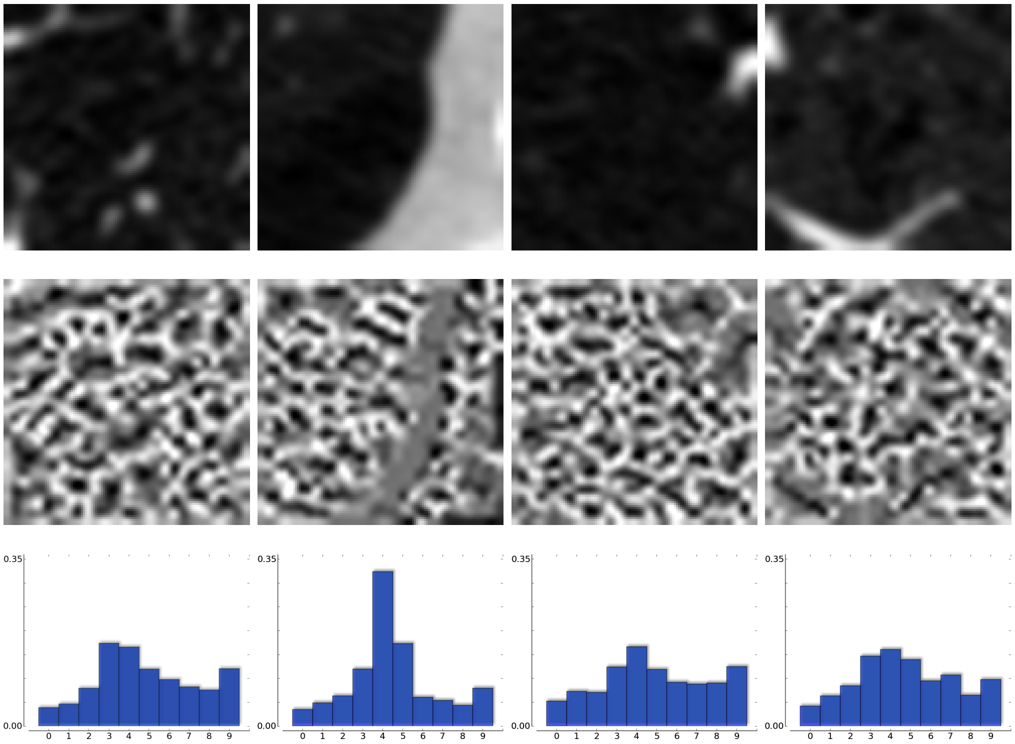
- Skewness (Υ) is a measure of asymmetry; it can be positive, which means that the distribution tends to the right, negative when the distribution tends to the left, or even zero, which typically implies a symmetric distribution:

$$\Upsilon_{(s,\theta)} = \frac{\mu^3_{(s,\theta)}}{\sigma^3_{(s,\theta)}} \tag{6.5}$$

- We also included a measure of contrast (Ψ) using kurtosis (K)

$$\Psi_{(s,\theta)} = \frac{\sigma_{(s,\theta)}}{K^{0.25}_{(s,\theta)}} \tag{6.6}$$

  where $K_{(s,\theta)} = \frac{\mu^4_{(s,\theta)}}{\sigma^4_{(s,\theta)}}$ represents the degree of peakedness of a distribution. We followed the recommendations in [132] and used 0.25 to reduce the contrast value when it comes distributions with biased peaks and to increase it with polarized distributions.



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

Figure 6.1: *Example of the $LBP_{8,1}^{uni}$ operator applied to COPD. (first row) emphysema patches in the window $[-1000, -500]$ HU, (second row) the labeled images; and (third row) shows their histograms. All the images were magnified by a factor of 4 for a better visualization.* ***(a)*** *NT,* ***(b)*** *PS,* ***(c)*** *PL, and* ***(d)*** *CL*

We used the previous descriptors to characterize emphysema patterns and to construct Complex Gabor Feature Vectors ($\overline{CGFV}$) as follows:

$$\overline{CGFV} = [\mu_{(0,0)},\ \sigma_{(0,0)},\ \Upsilon_{(0,0)},\ \Psi_{(0,0)}, \ldots, \\ \mu_{(s-1,\theta-1)},\ \sigma_{(s-1,\theta-1)},\ \Upsilon_{(s-1,\theta-1)},\ \Psi_{(s-1,\theta-1)}] \tag{6.7}$$

6.2.1 EXTENDED COMPLEX GABOR FEATURE VECTORS WITH LOCAL BINARY PATTERNS

Along with Gabor functions, LBPs have been successfully applied to texture classification, (see Section 5.2) . This approach is based on the idea that textural properties within homogeneous regions can be mapped into histograms that represent micro-features, (see Fig. 6.1).

We propose to concatenate a $\overline{CGFV}$ and its corresponding $LBP_{P,R}^{uni}$ histogram, see Eq. (5.5), into a single sequence to build a mixture descriptor called Extended-complex Gabor Feature Vector ($\overline{ECGFV}_{P,R}$) that represents any given texture patch:

$$\overline{ECGFV}_{P,R} = \left[\overline{CGFV},\ H_i\right] \tag{6.8}$$

where the subscript (P, R) indicates what $LBP_{P,R}^{uni}$ was used and

$$H_i = \sum_{x=1}^{N} \sum_{y=1}^{M} C\left\{L(x, y) == i \,|\, i = 0, \ldots, P+1\right\} \tag{6.9}$$

with

$$C(A) = \begin{cases} 1 & \text{if A is true} \\ 0 & \text{otherwise} \end{cases} \tag{6.10}$$

In this way, $\overline{ECGFV}_{P,R}$ simultaneously encodes global texture characteristics extracted by Gabor filters and local information provided by LBPs.

6.3 Other methods

- **Log-Gabor Functions** (LGF) [64]. Already presented in Section 3.3, they are defined in the frequency domain as Gaussian functions shifted from the origin, (see Fig. 3.7). They have a null DC component and can be split into two components, radial and angular filters. We also constructed a filter bank of 24 filters distributed in 4 scales (even scales were rotated by a constant factor consisting of the half a distance between filter centers) and 6 orientations and computed the log-Gabor coefficients $LG_{(S,\Theta)}$ as follows:

  $$LG_{(S,\Theta)}(x, y) = \mathcal{F}^{-1}\left\{\hat{I}(u, v) \bullet \hat{G}_{(S,\Theta)}(\rho, \theta)\right\} \tag{6.11}$$

  with $\rho = \sqrt{u^2 + v^2}$, $\theta = \arctan\left(\frac{v}{u}\right)$, and $\theta_0 = \Theta$. Thus, the feature vectors, $\overline{LGFV}$, were built in the same way as $\overline{CGFV}_{P,R}$ vectors using the set of four statistics previously described in Section 6.2.

- **Sparse Gabor Coding** (SGC[1]). Gabor filters provide redundant representations, which may hamper classification tasks. As proposed first by [133], this problem may be solved using a greedy algorithm. This approach corresponds to first choosing the single filter Φ_i that best fits the image, $I(x, y)$, along with a suitable coefficient a_i, such that the *single* source $a_i\Phi_i$ is a good match to the image:

  $$i = \arg\max_j \left(\left\langle \frac{I(x, y)}{\|I(x, y)\|}, \frac{\Phi_j}{\|\Phi_j\|} \right\rangle\right) \tag{6.12}$$

  where $\langle \cdot, \cdot \rangle$ represents the inner product.

  The associated coefficient is the scalar projection: $a_i = \langle I(x, y), \frac{\Phi_i}{\|\Phi_i\|^2} \rangle$. Knowing this choice, the image can be decomposed as: $I(x, y) = a_i\Phi_i + \mathbf{R}$ where $\mathbf{R}$ is the residual image. We then repeat this 2-step process on the residual until some stopping criterion is met.

---

[1] In collaboration with Dr. Laurent Perrinet from the Institut de Neurosciences Cognitives de la Méditerranée, Aix-Marseille University, France

This procedure is known as the Matching Pursuit algorithm, which has proven to be a good approximation for natural images [134]. Measuring the ratio of extracted energy in the images, $N = 256$ edges were on average enough to extract 90% of the energy of whitened images on all sets of images. Thus, We used this set of sparse coefficients as the input vector for the classification framework.

- **Gray-level Co-occurrence Matrices** (GLCM[2]) were proposed by Haralick [135]. This descriptor evaluates spatial relationship among gray levels. Each pixel in an image $I(x, y)$ is assigned to one of N_g gray levels. The GLCM matrix is composed of a set of $P_{ij}|i, j = 1, \ldots, N_g$ values. Here, P_{ij} represents the number of occurrences of two pixels with gray levels i and j separated by a distance d in the direction of the angle θ. The GLCM's elements are normalized, providing the relative frequency of occurrence for a pair of gray levels. The element $p(i, j)$ denotes the probability of finding the pair of levels (i, j) in the image, which is obtained as:

$$p(i, j) = P_{ij} \left(\sum_{i,j}^{N_g} P_{ij} \right)^{-1} \tag{6.13}$$

  10 features were chosen to capture texture properties: energy, contrast, correlation, homogeneity, entropy, autocorrelation, dissimilarity, cluster shade, cluster prominence, and maximum probability. In our study, N_g was set to 8 according previous works focused on texture analysis [72]. The distance parameter, d, was set to 1 while four different angle values were assessed: $0, 45, 90$, and 135 degrees. Thus, a total of 40 descriptors (10 statistical features for each of the four orientations) were obtained for each texture.

- **Discrete Tchebichef Moments** (DTM) [136] are computed by projecting the image $I(x, y)$ onto the set of Tchebichef polynomial kernels, see Fig. 6.2. DTM provides a unique representation of the image in the spanned Tchebichef space. The moment $T_{pq}\,(p, q = 0, 1, \ldots, N-1)$ of order $s = p + q$ is defined as:

$$T_{pq} = \frac{1}{\tilde{\rho}(p, N)\,\tilde{\rho}(p, N)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \tilde{t}_p(x)\,\tilde{t}_q(y)\,I(x, y) \tag{6.14}$$

  $\tilde{t}_p(x)$ and $\tilde{t}_q(x)$ are scaled Tchebichef polynomials and $\rho(n, N)$ is its squared norm.

  T_{pq} quantifies the correlation between the image $I(x, y)$ and the kernel $\tilde{t}_p(x)\,\tilde{t}_q(y)$. Hence, this magnitude will be higher for images characterized by repetitive patterns occurring at a similar rate to the kernel.

  The following feature evaluates the similarity between the image and the varying patterns implemented by s-order Tchebichef kernels: $T(s) = \sum_{p+q=s} |T_{pq}|$, $(s = 0, 1, \ldots, 2N-2)$. The analysis based on DTM yields a vector of length $2N - 1$ to describe the texture attributes.

---

[2] In collaboration with Dr. Victor Marcos from the Instituto de Óptica, CSIC, Spain.

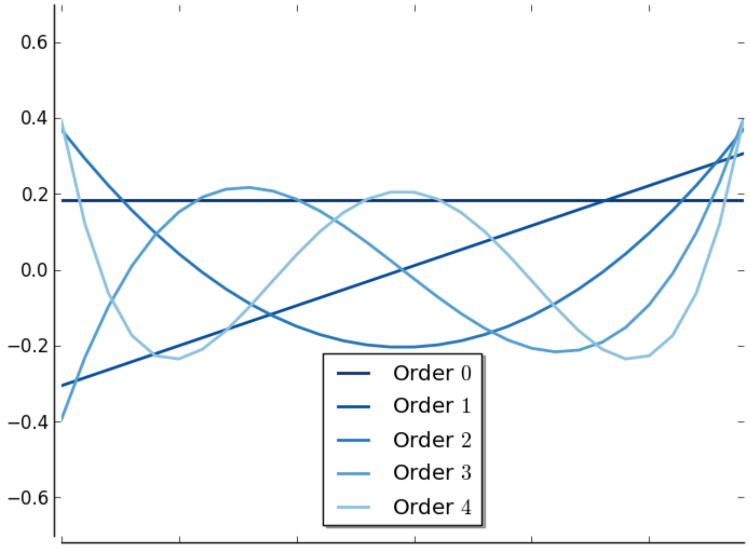


Figure 6.2: *Tchebichef polynomials of the first kind from zero to four order.*

6.4 Multi-class kernel Fisher discriminant analysis

It must be considered that the size of a training set should be exponentially increased with the dimensionality of the input space. Such problem, sometimes called "curse of dimensionality" has been widely investigated, e.g. in [137, 138]. Since the previous methods generate high dimensional feature vectors and a limited dataset is available in our problem, we added a discriminant analysis step to prevent overfitting. Another motivation for reducing feature vector dimension is that psychophysical findings indicate that perceptual tasks such as similarity judgment tend to be performed on a low-dimensional representation [139].

Discriminant analysis was firstly introduced by Ronald Fisher for two class problems (Fisher Discriminant Analysis, FDA) and remains to be one of the most popular methods for dimensionality reduction [140]. Contrary to PCA, FDA projects feature vectors onto a line which preserves direction useful for data classification. Below, we briefly present the general Fisher approach to C-classes (NFDA).

Let $\boldsymbol{X}^1 = \left\{x_1^1, x_2^1, \ldots, x_{l_1}^1\right\}, \ldots, \boldsymbol{X}^C = \left\{x_1^C, x_2^C, \ldots, x_{l_C}^C\right\}$ be samples from C classes, then Fisher's projection is given by the vector w which maximizes:

$$J(w) = \frac{w^T \boldsymbol{S_B} w}{w^T \boldsymbol{S_W} w} \tag{6.15}$$

S_B is the between scatter matrix defined by:

$$\boldsymbol{S_B} = \sum_{i=1}^{C} l_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{6.16}$$

where $\mu_i = \frac{1}{l_i} \sum_{j=1}^{l_i} x_j^i$ and μ is the overall mean.

On the other hand, S_W is the within class scatter matrix defined by:

$$\boldsymbol{S_W} = \sum_{i=1}^{C} S_i \tag{6.17}$$

where $S_i = \sum_{j=1}^{l_i} \left(x_j^i - \mu_i\right)\left(x_j^i - \mu_i\right)^T$.

Thus, the optimal projection matrix w^* are the $C-1$ largest eigenvalues of $\boldsymbol{S_W^{-1}} \boldsymbol{S_B}$.

However, NFDA has an important limitation because it assumes Gaussian likelihoods. To overcome it, Sebastian Mika ety al. proposed a non-linear generalization by mapping the original data into some feature space and computing FDA there [141]. Thus, the goal is to find $w^* \in \zeta$ that maximize:

$$J(w) = \frac{w^T \boldsymbol{S_B^\Phi} w}{w^T \boldsymbol{S_W^\Phi} w} \tag{6.18}$$

where $\boldsymbol{S_B^\Phi}$ and $\boldsymbol{S_W^\Phi}$ are the corresponding matrices in ζ and Φ is a mapping function from input data into a higher dimensional (possibly infinite) inner product space (ζ).

A drawback of using Eq. (6.18) is that the mapping function Φ must be calculated but in some cases is not possible. Mika et al. reformulated the algorithm in terms of inner products and proposed the kernel Fisher discriminant analysis (KFDA) for two class problems. The idea is to solve FDA limitations by calculating dot products of mapped data points without a mapping function Φ. First, it is necessary to define a kernel matrix of an inner product, $k(x,y) = \langle \Phi(x), \Phi(y) \rangle$, instead of computing the explicit mapping function. Thus, $\boldsymbol{K}(m,n) = k(X_m, X_n)$ where $\boldsymbol{X} = \bigcup_{i=1}^{C} \boldsymbol{X}^i$. There are three popular kernels: (i) polynomial kernel, $k(x,y) = (xy + a)^b$; (ii) sigmoidal kernel, $k(x,y) = \tanh(axy + b)$; and (iii) Gaussian kernel or Radial Basis Function (RBF), $k(x,y) = e^{-\frac{1}{2}\frac{\|x-y\|^2}{a^2}}$. For all kernels $a, b \in \mathbb{R}^+$. It follows that:

$$\begin{aligned} w^T \boldsymbol{S_B^\Phi} w &= \alpha \boldsymbol{P} \alpha^T \\ w^T \boldsymbol{S_W^\Phi} w &= \alpha \boldsymbol{Q} \alpha^T \end{aligned} \tag{6.19}$$

and the between scatter matrix is defined by

$$\boldsymbol{P} = \sum_{j=1}^{C} l_j \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T \tag{6.20}$$

where $\mu_j = \frac{1}{l_j} \sum_{\forall n \in X^j} \boldsymbol{K}(m,n)$ and $\mu = \frac{1}{l} \sum_{\forall n} \boldsymbol{K}(m,n)$.

$\boldsymbol{Q}$ is the within class scatter matrix defined by:

$$\boldsymbol{Q} = \boldsymbol{K}\boldsymbol{K}^T - \sum_{j=1}^{C} l_j \mu_j \mu_j^T \tag{6.21}$$

we used $\boldsymbol{Q} = \boldsymbol{Q} + r\boldsymbol{I}$ to guarantee that $\boldsymbol{Q}$ is positive definite.

Finally, α^* is built with the $C-1$ largest eigenvalues of $\boldsymbol{Q}^{-1}\boldsymbol{P}$ and the projection can be computed as:

$$y = \boldsymbol{K}\alpha^* \tag{6.22}$$

This method reduces $\overline{ECGFV}_{P,R}$ length to $C-1$ bins.

6.5 Material

We used two datasets labeled by experienced pulmonologists: the **Bruijne and Sørensen dataset** (BS), which was provided by Prof. Dr. Bruijne and Dr. Sørensen [125]. It consists of 168 non-overlapping patches of size 61×61 pixels manually annotated in 25 subject which were previously divided in three groups: healthy non-smokers, smokers without COPD, and smokers with moderate or severe COPD. These patches belong to three types of patterns: NT (59 patches from 8 subjects), CL (50 patches from 7 subjects), and PS (59 patches from 10 subjects); so that the NT patches were annotated in healthy non-smokers, while the CL and PS patches were annotated in both smokers with and without COPD; and **Brigham and Women's Hospital dataset** (BWH). This dataset was provided by researchers from the Brigham and Women's Hospital using a subset of the COPDGene study [127, 129]. The COPDGene study uses 342 CT scanners located in 16 sites. In total, 1337 patches, which belong to 353 subjects, were randomly selected. The distribution per pattern is: NT (370 patches from 74 subjects), PS (184 patches from 52 subjects), and PL (148 patches from 39 subjects). In addition, BWH includes three subtypes of CL (mild, moderate, and severe): CL1 (170 patches from 5 subjects), CL2 (287 patches from 84 subjects), and CL3 (178 patches from 49 subjects), respectively. The size of the samples was chosen to fit the physical extent of emphysema within the secondary lobule corresponding to 31×31 pixels. Prior to the application of our method, the data were normalized by the global mean and the standard deviation. Neither BS nor BHW contains private information of patients.

6.6 Experiments and results

Parameter selection is a fundamental step in any classification problem, its goal is to find a global optimum to achieve the best results in terms of accuracy and bias. 10-fold cross-validation is a simple and yet widely employed technique for model validation that randomly splits up data into 10 disjoint subsets of approximately equal size. For each fold the remaining 9 subsets are used to train the model, the average of all folds should provide an estimate of the model.

However, in order to reduce bias, Varma et al. [142] recommend a nested procedure that consists in splitting up the data into 10 folds, for each fold the remaining 9 subsets are again splitting up into 10 subsets and used to train the model. In this dissertation, during the training stage, we used the proposal of Varma et al. and chose the parameters found in the inner circle of the cross-validation that maximized the average accuracy.

The validation stage consisted in evaluating the full datasets with 10-fold cross-validation and with leave-one-patient-out cross-validation. We used k-NN classifier and the Euclidean distance as metric; such a distance was computed from input samples to every training data, so that we classified samples using the majority rule among the k-closest vectors.

6.6.1 BS DATASET

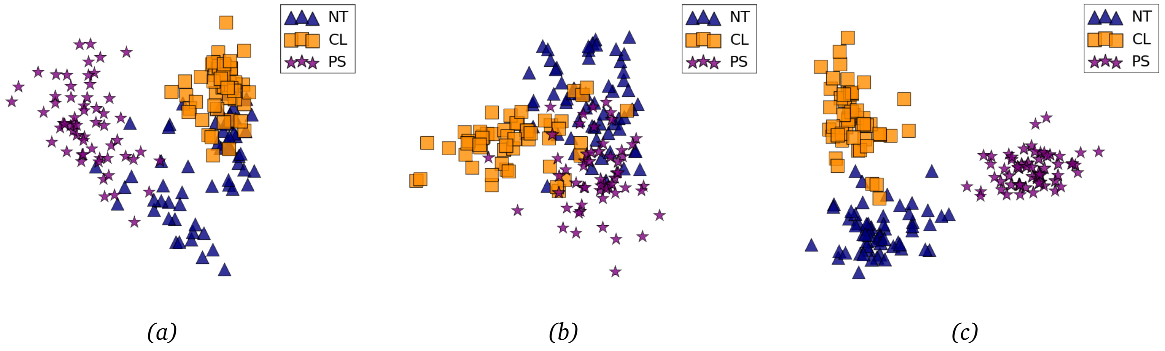
The first experiments were performed on the BS dataset; we completed several tests varying $k = \{1, 2, \ldots, 25\}$. In [125], Sørensen et al. used $k = 1$, here the best rate was achieved with $k = 20$. Since KFDA projects data onto a new space where class separation is maximized, the variations in classification rates due to changes in k were minimized, (see Fig. 6.3). We used the RBF kernel with $a = 543$,;the classification rates in the range $a < 450$ and $a > 550$ decreased dramatically due to the variance of the kernel.

In order to assessed our proposal, we tested three possible combinations of $\overline{CGFV}$ and $LBP^{uni}_{P,R}$ by varying the number of neighbors and radius length: $\{8, 1\}$, $\{16, 2\}$, and $\{24, 3\}$; these values are recommended in the literature for testing purposes [96]. We borrowed the Precision (Pr) and the Sensitivity (Se) from the confusion matrices. Furthermore, we computed the F_1-Score, which measures the accuracy of a test: $F_1\text{-Score} = 2 * \frac{\text{Pr} * \text{Se}}{\text{Pr} + \text{Se}}$.

The results using the BS dataset were summarized in Table 6.1. The best accuracy, 93.51%, was achieved using $\overline{ECGFV}_{16,2}$, which is a combination of $\overline{CGFV}$ and $LBP^{uni}_{16,2}$. We set $R = 2$ that led to a higher accuracy of about 6%.

On the contrary, $R > 2$ caused lower accuracies. This suggest that local variations, which can be interpreted as edges, may be useful for characterizing emphysema patterns. We also increased the number of neighbors, $P = 16$, that made our proposal less sensitive to noise.

We made an assessment of the methods previously described in Section 6.3, we evaluated each



(a) (b) (c)

Figure 6.3: *KFDA on the BS three-class dataset. The data were reduced into a 2D space. The vectors were produced by* **(a)** $\overline{CGFV}$ *(96-dimensional space);* **(b)** $LBP^{uni}_{8,1}$ *(10-dimensional space); and* **(c)** $\overline{ECGFV}_{8,1}$ *(109-dimensional space). The final space depends on the number of classes*

Table 6.1: *Classification rates of $\overline{ECGFV}_{8,1}$, $\overline{ECGFV}_{16,2}$, and $\overline{ECGFV}_{24,3}$ in the BS dataset. All the data are expressed in (%)*

	$\overline{ECGFV}_{8,1}$				$\overline{ECGFV}_{16,2}$				$\overline{ECGFV}_{24,3}$			
	NT	CL	PS	mean(±std)	NT	CL	PS	mean(±std)	NT	CL	PS	mean(±std)
Pr	81.03	88.00	90.00	86.34(±4.71)	88.71	95.83	96.55	**93.70**(±4.33)	78.69	89.36	88.33	85.46(±5.89)
Se	79.66	88.00	91.53	86.40(±6.10)	93.22	92.00	94.92	**93.38**(±1.47)	81.36	84.00	89.83	85.06(±4.34)
F_1**-Score**	80.34	88.00	90.76	86.37(±5.40)	90.91	93.88	95.73	**93.51**(±2.43)	80.00	86.60	89.07	85.22(±4.69)

Table 6.2: *F_1-Score of several methods on the BS dataset. The extended approach was built by concatenating $LBP^{uni}_{8,1}$ to the corresponding method. All data are in (%).*

F_1**-Score**	**Methods**		
	LGF	**GLCM**	**DTM**
single	72.98	73.76	60.96
extended	**79.88**	78.96	76.97

approach and its extended version, which was built by concatenating a $LBP^{uni}_{8,1}$ histogram to the corresponding method to form a single sequence. Results are shown in Table 6.2, note that SGC is not presented because the results are not statistically significant.

Furthermore, we carried out a comparison of $\overline{CGFV}$, $LBP^{uni}_{8,1}$, $LBP^{uni}_{16,2}$, and $LBP^{uni}_{24,3}$. We also computed the performance of Gabor filters using only the real part, $R_{(s,\theta)}$, see Eq. (6.1). The estimated classification accuracies are shown in Table 6.3.

We are interested in that our proposal generalizes to unseen patients and events. Therefore, we assessed our algorithm using leave-one-patient-out cross-validation. Table 6.4 shows classification rates and comparisons among $\overline{CGFV}$, $LBP^{uni}_{P,R}$, and $R_{(s,\theta)}$ which are distinguished as "single descriptors" and $\overline{ECGFV}_{P,R}$, distinguished as "extended descriptor." Although Sørensen et al. used the same dataset, a straightforward comparison is not possible because they did not report classification rates for patches of 51 × 51 pixels.

Table 6.3: *Comparison rates of $\overline{CGFV}$, $LBP^{uni}_{P,R}$, and $\overline{R_{(s,\theta)}}$ in the BS dataset. $\overline{R_{(s,\theta)}}$ denotes Gabor feature vectors using only real filters. All the data are expressed in (%)*

Descriptor	**Pr**				**Se**				F_1**-Score**
	NT	CL	PS	mean(±std)	NT	CL	PS	mean(±std)	
$LBP^{uni}_{8,1}$	68.85	79.59	74.14	74.19(±5.37)	71.19	78.00	72.88	74.02(±3.54)	74.11
$LBP^{uni}_{16,2}$	74.60	79.59	82.14	**78.78**(±3.84)	79.66	78.00	77.97	**78.54**(±0.97)	**78.66**
$LBP^{uni}_{24,3}$	73.33	85.11	77.05	78.49(±6.02)	74.58	80.00	79.66	78.08(±3.04)	78.28
$\overline{CGFV}$	70.37	75.00	87.10	77.49(±8.64)	64.41	78.00	91.53	77.98(±13.56)	77.73
$\overline{R_{(s,\theta)}}$	65.38	71.15	84.38	73.64(±9.74)	57.63	74.00	91.53	74.39(±16.95)	74.01

Table 6.4: *Comparison rates in the BS dataset using leave-one-patient-out cross validation. All the data are expressed in (%)*

	Descriptor	**Pr**				**Se**				F_1**-Score**
		NT	CL	PS	mean(±std)	NT	CL	PS	mean(±std)	
single	$LBP^{uni}_{8,1}$	61.40	76.60	67.19	68.40(±7.67)	59.32	72.00	72.88	68.07(±7.58)	68.24
	$LBP^{uni}_{16,2}$	74.58	76.60	75.81	75.66(±1.09)	74.58	72.00	79.66	75.41(±3.90)	75.54
	$LBP^{uni}_{24,3}$	69.84	82.61	77.97	**76.80**(±6.46)	74.58	76.00	77.97	**76.18**(±1.70)	**76.49**
	$\overline{CGFV}$	53.33	64.44	82.54	66.77(±14.74)	54.24	58.00	88.14	66.79(±18.58)	66.78
extended	$\overline{R_{(s,\theta)}}$	57.41	66.00	82.81	68.74(±12.92)	52.54	66.00	89.83	69.46(±18.88)	69.10
	$\overline{ECGFV}_{8,1}$	77.19	87.76	85.48	83.48(±5.56)	74.58	86.00	89.83	83.47(±7.93)	83.48
	$\overline{ECGFV}_{16,2}$	83.87	93.48	91.67	**89.67**(±5.10)	88.14	86.00	93.22	**89.12**(±3.70)	**89.39**
	$\overline{ECGFV}_{24,3}$	80.00	87.23	90.16	85.89(±5.23)	81.36	82.00	93.22	85.53(±6.67)	85.71

6.6.2 BWH DATASET

The BWH dataset was previously used in [127] where the authors achieved an accuracy around 66%. Also in [129], the dataset was used with a technique based on the embedded probabilistic PCA; the authors achieved a final rate of 69%.

Note that this dataset includes three subtypes of CL (mild, moderate, and severe), which increases the complexity of the classification tasks. The classification rates of the BWH dataset are summarized in Table 6.5. We kept the same configuration, $k = 20$ and $a = 543$, and also assessed three possible combinations of $\overline{CGFV}$ and $LBP^{uni}_{P,R}$. The best F_1-Score was achieved with $\overline{ECGFV}_{16,2} = 71\%$. AS in the previous Table 6.1, we set $R = 2$ that led to a higher accuracy of about 5%, while $R > 2$ caused lower accuracies.

We also performed comparison among all methods and the BWH dataset, (see Table 6.6). Note that Low Attenuation Areas (LAA) often exhibit a variety of shapes in patients with CL; some patients show LAA with well-defined borders, while others do not. CL is not a single morphological feature, therefore, the lower precision and sensitivity rates are among the three severity levels of CL.

We carried out a comparison among the methods: $\overline{CGFV}$, $LBP^{uni}_{8,1}$, $LBP^{uni}_{16,2}$, $LBP^{uni}_{24,3}$, and $\overline{R_{(s,\theta)}}$. The estimated classification accuracies are shown in Table 6.7. Note that all the single descriptor rates were lower than those obtained by our proposal.

Table 6.5: *Classification rates of* $\overline{ECGFV}_{8,1}$, $\overline{ECGFV}_{16,2}$, *and* $\overline{ECGFV}_{24,3}$ *in the BWH dataset. All the data are expressed in (%)*

	$\overline{ECGFV}_{8,1}$							$\overline{ECGFV}_{16,2}$							$\overline{ECGFV}_{24,3}$						
	NT	PS	PL	CL1	CL2	CL3	mean(±std)	NT	PS	PL	CL1	CL2	CL3	mean(±std)	NT	PS	PL	CL1	CL2	CL3	mean(±std)
Pr	77.43	83.42	73.03	58.20	63.11	57.69	68.81(±10.72)	80.79	85.28	78.23	65.25	62.54	61.64	**72.29**(±10.34)	79.27	86.63	78.01	56.72	60.62	57.93	69.86(±12.94)
Se	86.22	87.50	75.00	41.76	72.13	42.13	67.46(±20.66)	88.65	91.30	77.70	45.29	70.38	50.56	**70.65**(±19.22)	87.84	88.04	74.32	44.71	67.60	47.19	68.28(±19.03)
F_1**-Score**	81.91	85.41	74.00	48.63	67.32	48.70	67.66(±16.00)	84.54	88.19	77.96	53.47	66.23	55.55	**71.00**(±14.81)	83.34	87.33	76.12	50.00	63.92	52.01	68.79(±15.92)

Table 6.6: *F_1-Score of several approaches on the BWH dataset. The extended vectors were built by concatenating the $LBP^{uni}_{8,1}$ histogram to the corresponding method. All data are in (%).*

F_1-Score	**Methods**			
	LGF	**GLCM**	**DTM**	**SGC**
single	56.79	45.13	45.27	32.96
extended	**61.00**	57.05	55.77	51.58

Table 6.7: *Comparison rates of $\overline{CGFV}$, $LBP^{uni}_{P,R}$, and $\overline{R_{(s,\theta)}}$ in the BWH dataset. The latter denotes Gabor feature vectors using only the real filters. All the data are expressed in (%)*

Descriptor	**Pr**							**Se**							**F_1-Score**
	NT	PS	PL	CL1	CL2	CL3	mean(±std)	NT	PS	PL	CL1	CL2	CL3	mean(±std)	
$LBP^{uni}_{8,1}$	57.46	70.56	61.54	29.85	54.67	33.01	51.18(±16.25)	77.03	82.07	43.24	11.76	67.25	19.10	50.08(±30.07)	50.62
$LBP^{uni}_{16,2}$	58.17	80.09	57.47	36.26	55.26	29.81	52.84(±17.91)	78.92	91.85	33.78	19.41	65.85	17.42	51.21(±31.91)	52.01
$LBP^{uni}_{24,3}$	59.74	84.74	58.25	30.65	52.87	29.09	52.56(±20.74)	74.59	87.50	40.54	22.35	64.11	17.98	51.18(±28.57)	51.86
$\overline{CGFV}$	70.55	79.05	72.90	44.71	52.40	55.56	**62.53**(±13.52)	80.27	90.22	76.35	22.35	57.14	47.75	**62.35**(±25.04)	**62.44**
$\overline{R_{(s,\theta)}}$	70.00	78.10	69.09	51.81	51.30	53.90	62.37(±11.46)	81.35	89.13	77.03	25.29	55.05	42.70	61.76(±24.92)	62.06

Leave-one-patient-out cross-validation was performed in the BWH dataset, Table 6.8 summarized rates and comparisons of all the methods. The best rate was achieved with our proposal, $\overline{ECGFV}_{16,2}$. Note that when it comes BWH, the variance es greater than in BS, this increase of the variance was due to BWH includes three subtypes of CL. However, CL is not a single morphological feature.

Table 6.8: *Comparison rates in the BWH dataset using leave-one-patient-out cross validation. All the data are expressed in (%)*

	Descriptor	**Pr**							**Se**							**F_1-Score**
		NT	PL	PS	CL1	CL2	CL3	mean(±std)	NT	PL	PS	CL1	CL2	CL3	mean(±std)	
single	$LBP^{uni}_{8,1}$	55.89	68.49	58.04	26.09	52.16	26.53	47.87(±17.56)	74.32	81.52	43.92	10.59	63.07	14.61	48.01(±30.25)	47.94
	$LBP^{uni}_{16,2}$	57.65	79.90	51.76	32.97	55.46	31.52	51.54(±17.90)	78.38	90.76	29.73	17.65	68.99	16.29	50.30(±32.93)	50.91
	$LBP^{uni}_{24,3}$	58.95	82.81	57.01	29.32	55.39	30.77	52.38(±19.99)	72.97	86.41	41.22	22.94	66.20	17.98	51.29(±28.07)	51.83
	$\overline{CGFV}$	69.72	79.43	69.62	41.05	51.15	50.00	**60.17**(±14.85)	80.27	90.22	74.32	22.94	54.36	40.45	**60.43**(±25.76)	**60.30**
	$\overline{R_{(s,\theta)}}$	68.69	77.36	72.26	34.57	49.17	50.00	58.68(±16.62)	79.46	89.13	75.68	16.47	51.57	44.94	59.54(±27.07)	59.11
extended	$\overline{ECGFV}_{8,1}$	77.94	84.26	73.47	59.17	62.15	54.29	68.55(±11.76)	85.95	90.22	72.97	41.76	70.38	42.70	67.33(±20.84)	67.94
	$\overline{ECGFV}_{16,2}$	80.84	85.35	77.33	60.34	61.76	59.86	**70.91**(±11.54)	88.92	91.85	78.38	41.18	68.64	49.44	**69.74**(±20.79)	**70.32**
	$\overline{ECGFV}_{24,3}$	78.61	87.30	77.08	52.31	59.94	58.16	68.90(±13.93)	88.38	89.67	75.00	40.00	66.20	46.07	67.55(±20.98)	68.22

6.7 Conclusions

We proposed a new approach to quantify up to six emphysema patterns based on complex Gabor filters and local binary patterns. This joint model allows to encode global texture characteristics with local information simultaneously. We presented the complex Gabor model for texture analysis and summarized its properties related to the HVS. Since Gabor-based methods transform images into a high-dimensional feature vectors, we applied kernel Fisher discriminant analysis via the kernel trick to avoid computing a mapping function and to find the most discriminant non-linear boundary among classes. We performed 10-fold cross-validation and leave-one-patient-out cross validation to assess our proposal. In general, extended descriptors increased classification rates around 15% in the BW dataset. Concerning the BWH dataset our method achieved a F_1-Score of 70.32; that means 10% above any single descriptor. This rate was mainly influenced by the misclassified three severity levels of CL. However, our proposal outperformed all the methods. These results have shown that the proposed method is a promising technique that yields a good performance in emphysema classification. Furthermore, this approach may be useful in other textural classification scenarios beyond medical imaging.

> *– If people let the government decide what foods they eat and what medicines they take, their bodies will soon be in as sorry a state as the souls who live under tyranny.*
>
> Thomas Jefferson

7

Conclusions and future work

In recent years bio-inspired image models have captured the attention of researchers; such a phenomenon has been partially supported because of the advance in understanding of the HVS, which is a complex system based on a retinotopic organization that performs parallel processing by lateral, feedforward, and feedback connections. Among all the components that form the HVS, the retina and the striate cortex are the ones where we have achieved a better knowledge. The former is regarded as a preprocessing stage that adapts *visual stimuli* for subsequent processing, while the striate cortex is considered as a low-level descriptor.

Understanding the encoding process of visual information may lead to overcome limitations of existing signal-processing-based methods. Although the models used in image processing and computer vision are simplifications derived from psychophysical experiments, they have incorporated more complex features. In the previous chapters we established that the HVS decomposes *visual stimuli* in multiple frequency bands and different orientations where lower frequencies represent the context of the scene and higher frequency bands contain edges, texture information, and noise.

Literature distinguishes two types of image models: those ones that incorporate perceptual aspects such as contrast adaptation and visual masking; and models that mimic the structure and organization of the HVS. The latter is the most popular model. Researchers have focused efforts to propose innovative designs based on it for image processing and computer vision tasks. The perception of objects depends primarily on the ability of the HVS to extract continuous edges from a visual scene. However, the retina received fragmented edges due to the size of receptive fields.

Therefore it must implement mechanisms for integrating information and building the perception of continuous edges.

In this dissertation we analyzed and proposed an image model based on the studies of Daugman and Field. Our proposal is selective in both orientation and frequency bands. We applied the logarithmic function to Gabor filters in order to remove the DC component that disturbs frequency bands. In addition, we rotated the even bands in order to better cover the Fourier plane.

The HVS exploits redundancy as well. It converts *visual stimuli* into signals and then encodes them; therefore it must distinguish useful signals and rule out redundant information; for example, signals received by adjacent receptive fields. As we saw in Chapter 2, redundant systems are more robust than orthogonal ones because they are less sensitive to noise and can expand images in different ways. Gabor filters meet the previous characteristics, they have been used successfully in texture analysis and other applications which do no require image reconstruction such as denoising and edge extraction because it is not possible to reconstruct images with no artifacts from the expansion. However, Gabor filters are considered as optimal filters and reach the optimal trade-off between space and frequency domains.

We compared the classical Gabor model and our proposal in a texture segmentation problem. We revisited the Jain and Farrokhnia's algorithm, which is an unsupervised texture segmentation method applied on the Brodatz dataset. The results showed that our proposal outperformed the classical model because their frequency bands are better distributed, such a configuration leads to a better management of redundancy. This experiment allowed to discover that there is an optimal point where is possible to take advantage of the redundant information to discriminate textures; before or after this point the discrimination rate decays. Further experiments were performed with our proposal in texture classification and image retrieval using several distance metrics among histograms.

The second part of this dissertation stated that joint representations improve texture characterization because they simultaneously encode global characteristics with local information. This combined approach leads to a better texture classification rates. Furthermore, we included a study on local operators for the purpose of characterizing textures locally. Furthermore texture classification experiments were conducted in order to assess their performance and robustness to rotational changes.

The last part was dedicated to solve a classification problem of lung CT images. We included phase information to build vectors from complex Gabor functions. Furthermore, we appended LBP histograms to each vector to improve texture characterization. We also analyzed a set of global and local texture descriptors to characterize emphysema morphology. The results have shown the effectiveness of our proposal and that the combination of global and local descriptors provides robust feature vectors that lead to an improvement in the classification rate. The results suggest that combined descriptor methods such as Gabor schemes and LBPs are robust descriptors that outperform the state-of-the-art models.

However, combined methods have many parameters and their optimization depends mainly on the application. No single approach did perform best or very close to the best for all images and experiments. Therefore, it is difficult to propose a general or universal solution. This thesis also made clear that composite models are required to better characterize textures.

For four years we have studied textures from different approaches and in a variety of applications. We have left evidence that our proposals compete with state-of-art methods. We have taken advantage of the overcomplete information but there is much work to be done. To our knowledge the next step involves sparsity and sparse models. There is evidence that through sparse algorithms, overcomplete information can be optimized. Also edge extraction must be included in the sparse coding. Such improvements should yield a better characterization of textures and offer efficient descriptors for object recognition.

Bibliography

[1] H. Kolb, Simple anatomy of the retina, in: H. Kolb, R. Nelson, E. Fernández, B. Jones (Eds.), Webvision: The Organization of the Retina and Visual System, University of Utah, John Moran Eye Center, New York, 2012, pp. 1–2.

[2] J. Isaacs, R. Roberts, Constructions of equiangular tight frames with genetic algorithms, in: IEEE International Conference on Systems, Man and Cybernetics, SMC, pp. 595–598.

[3] J. Daugman, Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression, IEEE Transactions on Acoustics, Speech and Signal Processing 36 (1988) 1169–1179.

[4] R. Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010.

[5] A. Rosenfeld, J. L. Pfaltz, Sequential operations in digital picture processing, J. ACM 13 (1966) 471–494.

[6] E. Meijering, A chronology of interpolation: from ancient astronomy to modern signal and image processing, Proceedings of the IEEE 90 (2002) 319–342.

[7] J. Morlet, G. Arens, E. Fourgeau, D. Giard, Wave propagation and sampling theory; Part I, complex signal and scattering in multilayered media, Geophysics 47 (1982) 203–221.

[8] J. Morlet, G. Arens, E. Fourgeau, D. Giard, Wave propagation and sampling theory; Part II, sampling theory and complex waves, Geophysics 47 (1982) 222–236.

[9] A. Grossmann, J. Morlet, Decomposition of hardy functions into square integrable wavelets of constant shape, SIAM Journal on Mathematical Analysis 15 (1984) 723–736.

[10] Y. Meyer, Wavelets and Operators, Cambridge University Press, 1993.

[11] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 674–693.

[12] P. Burt, E. Adelson, The laplacian pyramid as a compact image code, IEEE Transactions on Communications 31 (1983) 532–540.

[13] D. Donoho, M. Vetterli, R. DeVore, I. Daubechies, Data compression and harmonic analysis, IEEE Transactions on Information Theory 44 (1998) 2435–2476.

[14] L. G. Roberts, Machine Perception of three-dimensional solids, Ph.D. thesis, Massachusetts Institute of Technology, 1963.

[15] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (1986) 679–698.

[16] D. Marr, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, Henry Holt and Co., Inc., New York, NY, USA, 1982.

[17] H. B. Barlow, Possible principles underlying the transformation of sensory messages, in: Sensory Communication, MIT Press, Boston, USA, 1961, pp. 217–234.

[18] B. A. Olshausen, D. J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1?, Vision Research 37 (1997) 3311–3325.

[19] M. Petrou, F. Faille, An imaging architecture based on derivative estimation sensors, in: E. Bayro-Corrochano, J.-O. Eklundh (Eds.), Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, volume 5856 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 3–18.

[20] D. H. Hubel, T. N. Wiesel, Receptive fields, binocular interaction and functional architectures in the cat's visual cortex, The Journal of Physiology 160 (1962) 106–154.

[21] X. G. Troncoso, S. L. Macknik, S. Martinez-Conde, Vision's first steps: Anatomy, physiology, and perception in the retina, lateral geniculate nucleus, and early visual cortical areas, in: G. Dagnelie (Ed.), Visual Prosthetics, Springer US, 2011, pp. 23–57.

[22] M. A. MacNeil, R. H. Masland, Extreme diversity among amacrine cells: Implications for function, Neuron 20 (2005) 971–982.

[23] V. Mante, V. Bonin, M. Carandini, Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli, Neuron 58 (2008) 625–638.

[24] D. H. Hubel, Eye, brain, and vision, W. H. Freeman., 1995.

[25] D. H. Hubel, T. N. Wiesel, Brain and Visual Perception: The Story of a 25-year Collaboration, Oxford University Press., New York, 2005.

[26] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, N. C. Rust, Do we know what the early visual system does?, The Journal of Neuroscience 25 (2005) 10577–10597.

[27] B. A. Olshausen, D. J. Field, How close are we to understanding V1?, Neural Computation 17 (2005) 1665–1699.

[28] A. K. Moorthy, Z. Wang, A. C. Bovik, Visual perception and quality assessment, in: G. Cristóbal, P. Schelkens, H. Thienpont (Eds.), Optical and Digital Image Processing, Wiley-VCH Verlag GmbH & Co. KGaA, 2011, pp. 419–439.

[29] J. L. Puchalla, E. Schneidman, R. A. Harris, M. J. Berry, Redundancy in the population code of the retina, Neuron 46 (2005) 493–504.

[30] D. Mumford, Neuronal architectures for pattern-theoretic problems, in: Large-Scale Theories of the Cortex, MIT Press, 1994, pp. 125–152.

[31] I. Daubechies, A. Grossmann, Y. Meyer, Painless nonorthogonal expansions, Journal of Mathematical Physics 27 (1986) 1271–1283.

[32] R. J. Duffin, A. C. Schaeffer, A class of nonharmonic Fourier series, Transactions Amer. Math. Soc. 72 (1952) 341–366.

[33] S. Mallat, A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, Academic Press, 3rd edition, 2008.

[34] J. J. Benedetto, G. E. Pfander, Wavelet transform and time-frequency signal analysis, Applied and Numerical Harmonic Analysis, Birkhäuser, pp. 3–36.

[35] Y. Xu, J. B. Weaver, D. M. Healy, J. Lu, Wavelet transform domain filters: a spatially selective noise filtration technique, IEEE Transactions on Image Processing 3 (1994) 747–758.

[36] N. Kingsbury, The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement, in: Proc. Europe Signal Processing Conf., volume 98, pp. 319–322.

[37] H. Bölcskei, H. Hlawatsch, Gabor Analysis and Algorithms: Theory and Applications, Birkhäuser, Boston, MA, USA, pp. 295–322.

[38] J. Kovačević, A. Chebira, Life beyond bases: The advent of frames (Part I), IEEE Signal Processing Magazine 24 (2007) 86–104.

[39] J. Kovačević, A. Chebira, Life beyond bases: The advent of frames (Part II), IEEE Signal Processing Magazine 24 (2007) 115–125.

[40] M. S. Lewicki, T. J. Sejnowski, Learning overcomplete representations, Neural Computation 12 (2000) 337–365.

[41] E. Simoncelli, W. Freeman, E. Adelson, D. Heeger, Shiftable multiscale transforms, IEEE Transactions on Information Theory 38 (1992) 587–607.

[42] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (1993) 3397–3414.

[43] D. J. Field, What is the goal of sensory coding?, Neural Comput. 6 (1994) 559–601.

[44] J. F. Murray, K. Kreutz-Delgado, Visual recognition and inference using dynamic overcomplete sparse learning, Neural Computation 19 (2007) 2301–2352.

[45] S. Marčelja, Mathematical description of the responses of simple cortical cells$*$, J. Opt. Soc. Am. 70 (1980) 1297–1300.

[46] C. Zetzsche, U. Nuding, K. S. Schill, Nonlinear overcomplete coding in visual cortex, Journal of Vision 5 (2005) 674.

[47] S. Fischer, F. Šroubek, L. Perrinet, R. Redondo, G. Cristóbal, Self-invertible 2d log-Gabor wavelets, Int. J. Comput. Vision 75 (2007) 231–246.

[48] G. H. Granlund, In search of a general picture processing operator, Computer Graphics and Image Processing 8 (1978) 155–173.

[49] J. P. Jones, L. A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, Journal of Neurophysiology 58 (1987) 1233–1258.

[50] R. Navarro, A. Tabernero, G. Cristóbal, Image representation with Gabor wavelets and its applications, Advances in Imaging and Electron Physics 97 (1996) 1–84.

[51] B. Escalante-Ramírez, The hermite transform as an efficient model for local image analysis: An application to medical image fusion, Comput. Electr. Eng. 34 (2008) 99–110.

[52] J. G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Am. A 2 (1985) 1160–1169.

[53] D. Gabor, Theory of communication, J. Inst. Elec. Eng. (London) 93III (1946) 429–457.

[54] H. Nyquist, Certain factors affecting telegraph speed, American Institute of Electrical Engineers, Transactions of the XLIII (1924) 412–422.

[55] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 397–423.

[56] R. Redondo, F. Šroubek, S. Fischer, G. Cristóbal, Multifocus image fusion using the log-Gabor transform and a multisize windows technique, Information Fusion 10 (2009) 163–171.

[57] L.-L. Shen, Z. Ji, Gabor wavelet selection and SVN classification for object recognition, Acta Automatica Sinica 35 (2009) 350–355.

[58] A. C. Bovik, M. Clark, W. S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 55–73.

[59] R. Nava, B. Escalante-Ramírez, G. Cristóbal, Texture image retrieval based on log-Gabor features, in: L. Alvarez, M. Mejail, L. Gomez, J. Jacobo (Eds.), Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, volume 7441 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2012, pp. 414–421.

[60] P. Kovesi, Phase congruency: A low-level image invariant, Psychological Research 64 (2000) 136–148.

[61] M. Gross, R. Koch, Visualization of multidimensional shape and texture features in laser range data using complex-valued Gabor wavelets, IEEE Transactions on Visualization and Computer Graphics 1 (1995) 44–59.

[62] O. Nestares, A. Tabernero, R. Navarro, J. Portilla, Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions, Journal of Electronic Imaging 7 (1998) 166–173.

[63] T. S. Lee, Image representation using 2D Gabor wavelets, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 959–971.

[64] D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. A 4 (1987) 2379–2394.

[65] J. Buhmann, J. Lange, C. von der Malsburg, Distortion invariant object recognition by matching hierarchically labeled graphs, in: Neural Networks, 1989. IJCNN., International Joint Conference on, pp. 155–159 vol.1.

[66] A. Dixit, N. P. Hegde, Image texture analysis - survey, in: Advanced Computing and Communication Technologies (ACCT), 2013 Third International Conference on, pp. 69–76.

[67] B. Julesz, Visual pattern discrimination, IRE Transactions on Information Theory 8 (1962) 84–92.

[68] B. Julesz, E. Gilbert, L. Sheep, H. Frisch, Inability of humans to discriminate between visual textures that agree in second-order statistics – revisited, Perception 2 (1973) 391–405.

[69] R. Paget, Texture modeling and synthesis, in: M. Mirmehdi, X. Xie, J. Suri (Eds.), Handbook of Texture Analysis, Imperial College Press, 2008, pp. 33–60.

[70] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using LBP variance (LBPV) with global matching, Pattern Recognition 43 (2010) 706–719.

[71] S. P. Awate, T. Tasdizen, R. T. Whitaker, Unsupervised texture segmentation with nonparametric neighborhood statistics, in: Proceedings of the 9th European conference on Computer Vision - Volume Part II, pp. 494–507.

[72] T. Randen, J. H. Husøy, Filtering for texture classification: A comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 21 (1999) 291–310.

[73] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local Binary Patterns and its application to facial image analysis: a survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 41 (2011) 765–781.

[74] T. Ojala, M. Pietikainen, D. Harwood, Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, in: Proceedings of the 12th International Conference on Pattern Recognition - Conference A: Computer Vision Image Processing (IAPR), volume 1, pp. 582–585.

[75] K. Seetharaman, Texture analysis based on a family of stochastic models, in: IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 518–523.

[76] D. Dunn, W. E. Higgins, J. Wakeley, Texture segmentation using 2-d Gabor elementary functions, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (1994) 130–149.

[77] T. P. Weldon, W. E. Higgins, D. F. Dunn, Efficient Gabor filter design for texture segmentation, Pattern Recognition 29 (1996) 2005–2015.

[78] D. Dunn, W. Higgins, Optimal Gabor filters for texture segmentation, Image Processing, IEEE Transactions on 4 (1995) 947–964.

[79] A. K. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, Pattern Recognition. 24 (1991) 1167–1186.

[80] H. Feng, Z. Jiang, J. Shi, Unsupervised texture segmentation based on latent topic assignment, Journal of Electronic Imaging 22 (2013) 013026–013026.

[81] A. Khanna, M. Shrivastava, Unsupervised techniques of segmentation on texture images: A comparison, in: IEEE International Conference on Signal Processing, Computing and Control (ISPCC), pp. 1–6.

[82] D. A. Clausi, M. E. Jernigan, Designing Gabor filters for optimal texture separability, Pattern Recognition 33 (2000) 1835–1849.

[83] S. E. Grigorescu, N. Petkov, P. Kruizinga, A comparative study of filter based texture operators using mahalanobis distance, in: 15th International Conference on Pattern Recognition, volume 3, pp. 885–888.

[84] P. S. Bradley, U. M. Fayyad, Refining initial points for k-means clustering, in: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998, pp. 91–99.

[85] P. Brodatz, Textures; a photographic album for artists and designers, Dover Publications New York, 1966.

[86] M. Mirmehdi, X. Xie, J. Suri, Handbook of Texture Analysis, Imperial College Press, London, UK, 2009.

[87] B. Caputo, E. Hayman, M. Fritz, J.-O. Eklundh, Classifying materials in the real world, Image and Vision Computing 28 (2010) 150–163.

[88] M. Varma, A. Zisserman, A statistical approach to texture classification from single images, Int. J. Comput. Vision 62 (2005) 61–81.

[89] L. Wang, D. C. He, Texture classification using texture spectrum, Pattern Recognition 23 (1990) 905–910.

[90] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: Proceedings of the third European conference on Computer Vision (Vol. II), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1994, pp. 151–158.

[91] S. Brahnam, L. C. Jain, L. Nanni, A. Lumini (Eds.), Local Binary Patterns: New Variants and Applications, volume 506 of *Studies in Computational Intelligence*, Springer, 2014.

[92] P. Brodatz, USC-SIPI, `http://sipi.usc.edu/database/database.php?volume=rotate`, 2012. [Online; accessed 1-March-2012].

[93] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Transactions on Image Processing 19 (2010) 1635–1650.

[94] L. Nanni, S. Brahnam, A. Lumini, Survey on LBP based texture descriptors for image classification, Expert Systems with Applications 39 (2012) 3634–3641.

[95] M. Pietikäinen, A. Hadid, G. Zhao, T. Ahonen, Computer vision using Local Binary Patterns, volume 40, Springer, 1st edition, xv, 207 p. 87 illus., 56 in color edition, 2011.

[96] T. Ojala, M. Pietikäinen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.

[97] M. Pietikäinen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, Pattern Recognition 33 (2000) 43–52.

[98] Y. Ma, Number local binary pattern: An extended local binary pattern, in: 2011 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), pp. 272 –275.

[99] L. Liu, P. Fieguth, G. Kuang, Generalized Local Binary Patterns for texture classification, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2011, pp. 123.1–123.11.

[100] X. Fu, W. Wei, Centralized Binary Patterns embedded with image euclidean distance for facial expression recognition, in: 4th International Conference on Natural Computation, volume 4, pp. 115–119.

[101] T. Tan, M. Zhang, F. Liu, Face recognition using Extended Local Binary Patterns and fuzzy information fusion, in: 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), volume 2, pp. 625–629.

[102] Z. Guo, L. Zhang, D. Zhang, A completed modeling of Local Binary Pattern operator for texture classification, IEEE Transactions on Image Processing 19 (2010) 1657–1663.

[103] S. Liao, M. W. K. Law, A. C. S. Chung, Dominant Local Binary Patterns for texture classification, IEEE Transactions on Image Processing 18 (2009) 1107–1118.

[104] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1996) 51–59.

[105] Z. Guo, L. Zhang, D. Zhang, S. Zhang, Rotation invariant texture classification using adaptive LBP with directional statistical features, in: 17th IEEE International Conference on Image Processing (ICIP), pp. 285–288.

[106] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[107] W. Xing-yuan, C. Zhi-feng, Y. Jiao-jiao, An effective method for color image retrieval based on texture, Computer Standards & Interfaces 34 (2012) 31–35.

[108] P. W. Huang, S. K. Dai, Image retrieval by texture similarity, Pattern Recognition 36 (2003) 665–679.

[109] Y. Jie, Z. Qiang, Z. Liang, C. Y. Wuhan, Research on texture images retrieval based on the Gabor wavelet transform, in: International Conference on Information Engineering, ICIE09., volume 1, pp. 79–82.

[110] M. E. ElAlami, A novel image retrieval model based on the most relevant features, Knowledge-Based Systems 24 (2011) 23–32.

[111] R. Nava, B. Escalante-Ramírez, G. Cristóbal, A comparison study of Gabor and log-Gabor wavelets for texture segmentation, in: 7th International Symposium on Image and Signal Processing and Analysis (ISPA), pp. 189–194.

[112] A. Kong, An analysis of Gabor detection, in: Image Analysis and Recognition, volume 5627 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, 2009, pp. 64–72.

[113] B. S. Manjunath, W. Y. Ma, Texture features for browsing and retrieval of image data, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18 (1996) 837–842.

[114] J. Lin, Divergence measures based on the Shannon entropy, IEEE Transactions on Information Theory 37 (1991) 145–151.

[115] M. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, IEEE Transactions on Image Processing 11 (2002) 146–158.

[116] G. Castellano, L. Bonilha, L. Li, F. Cendes, Texture analysis of medical images, Clinical Radiology 59 (2004) 1061–1069.

[117] C. Galban, M. Han, J. Boes, K. Chughtai, C. Meyer, T. Johnson, S. Galban, A. Rehemtulla, E. Kazerooni, F. Martínez, B. Ross, Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression, Nat. Med. 18 (2012) 1711–1715.

[118] H. Brody, Chronic obstructive pulmonary disease, Nature 489 (2012) S1.

[119] D. Sin, Mortality in COPD: The role of comorbidities, in: L. Nici, R. ZuWallack (Eds.), Chronic Obstructive Pulmonary Disease, Respiratory Medicine, Humana Press, 2012, pp. 1–13.

[120] M. Takahashi, J. Fukuoka, N. Nitta, R. Takazakura, Y. Nagatani, Y. Murakami, H. Otani, , K. Murata, Imaging of pulmonary emphysema: A pictorial review, Int J Chron Obstruct Pulmon Dis. 3 (2008) 193–204.

[121] O. Mets, P. Jong, B. Ginneken, H. Gietema, J. Lammers, Quantitative computed tomography in COPD: Possibilities and limitations, Lung 190 (2012) 133–145.

[122] M. Hayhurst, D. Flenley, A. Mclean, A. Wightman, W. Macnee, D. Wright, D. Lamb, J. Best, Diagnosis of pulmonary emphysema by computerised tomography, The Lancet 324 (1984) 320–322.

[123] A. Madani, J. Zanen, V. de Maertelaer, P. A. Gevenois, Pulmonary emphysema: Objective quantification at multi–detector row CT comparison with macroscopic and microscopic morphometry, Radiology 238 (2006) 1036–1043.

[124] N. Sverzellati, G. Randi, P. Spagnolo, A. Marchianò, M. Silva, J.-M. Kuhnigk, C. L. Vecchia, M. Zompatori, U. Pastorino, Increased mean lung density: Another independent predictor of lung cancer?, European Journal of Radiology 82 (2013) 1325–1331.

[125] L. Sørensen, S. Shaker, M. de Bruijne, Quantitative analysis of pulmonary emphysema using Local Binary Patterns, IEEE Transactions on Medical Imaging 29 (2010) 559 –569.

[126] A. Costa, G. Humpire-Mamani, A. Traina, An efficient algorithm for fractal analysis of textures, in: Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on, pp. 39–46.

[127] C. Mendoza, G. Washko, J. Ross, A. Diaz, D. Lynch, J. Crapo, E. Silverman, B. Acha, C. Serrano, R. Estepar, Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions, in: 9th IEEE International Symposium on Biomedical Imaging (ISBI), 2012, pp. 474–477.

[128] L. Sorensen, M. Nielsen, P. Lo, H. Ashraf, J. Pedersen, M. de Bruijne, Texture-based analysis of COPD: A data-driven approach, IEEE Transactions on Medical Imaging 31 (2012) 70–78.

[129] T. Zulueta-Coarasa, S. Kurugol, J. C. Ross, G. G. Washko, R. San Jose Estepar, Emphysema classification based on embedded probabilistic pca, in: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 3969–3972.

[130] A. Depeursinge, A. Foncubierta-Rodríguez, D. Ville, H. Müller, Multiscale lung texture signature learning using the Riesz transform, in: N. Ayache, H. Delingette, P. Golland, K. Mori (Eds.), Medical Image Computing and Computer-Assisted Intervention MICCAI 2012, volume 7512 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2012, pp. 517–524.

[131] W. R. Schwartz, F. Roberti de Siqueira, H. Pedrini, Evaluation of feature descriptors for texture classification, Journal of Electronic Imaging 21 (2012) 023016–1–023016–17.

[132] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Transactions on Systems, Man and Cybernetics, 8 (1978) 460–473.

[133] L. U. Perrinet, M. Samuelides, S. J. Thorpe, Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit., Neurocomputing 57C (2002) 125–134.

[134] L. U. Perrinet, Role of homeostasis in learning sparse representations, Neural Computation 22 (2010) 1812–1836.

[135] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst., Man, Cybern., Syst. SMC-3 (1973) 610–621.

[136] V. Marcos, G. Cristóbal, Texture classification using Tchebichef moments, J. Opt. Soc. Am. A 30 (2013) 1580–1591.

[137] M. Li, R. C. Staunton, Optimum gabor filter design and local binary patterns for texture segmentation, Pattern Recogn. Lett. 29 (2008) 664–672.

[138] J. Lee, M. Verleysen, High-dimensional data, in: J. Lee, M. Verleysen (Eds.), Nonlinear Dimensionality Reduction, Information Science and Statistics, Springer New York, 2007, pp. 1–16.

[139] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition, IEEE Transactions on Image Processing 11 (2002) 467 –476.

[140] R. A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (1936) 179–188.

[141] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K. Mullers, Fisher discriminant analysis with kernels, in: Proceedings of the 1999 IEEE Signal Processing Society Workshop. Neural Networks for Signal Processing IX, pp. 41–48.

[142] S. Varma, R. Simon, Bias in error estimation when using cross-validation for model selection, BMC Bioinformatics 7 (2006) 91–98.

Index

Colophon

THIS thesis was typeset using LaTeX.

This dissertation represents not only my work but also the knowledge of hundreds of researchers and engineers who have devoted hours to reveal the secrets of the human visual system. I want to thank all those who involuntarily contributed to this effort: Joseph Fourier (1768-1830), David Hilbert(1862-1943), Ronald Aylmer Fisher(1890-1962), Dennis Gabor(1900-1979), David Hubel(1926-2013), and Ingrid Daubechies(1954-).