TOPOLOGY OF SPACES OF MICRO-IMAGES, 
AND AN APPLICATION TO TEXTURE DISCRIMINATION 

A DISSERTATION 
SUBMITTED TO THE DEPARTMENT OF MATHEMATICS 
AND THE COMMITTEE ON GRADUATE STUDIES 
OF STANFORD UNIVERSITY 
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS 
FOR THE DEGREE OF 
DOCTOR OF PHILOSOPHY 

Jose Andres Perea Benitez 
May 2011
© 2011 by Jose Andres Perea Benitez. All Rights Reserved.
Re-distributed by Stanford University under license with the author.

This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.
http://creativecommons.org/licenses/by-nc/3.0/us/

This dissertation is online at: http://purl.stanford.edu/qs397kn3896
I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gunnar Carlsson, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

George Papanicolaou

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Andras Vasy

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.
Abstract

In the paper *On the Local Behavior of Spaces of Natural Images* [21], the framework of topological data analysis was used by Carlsson et al. to identify the geometric structure of high-density regions in the state space \( \mathcal{M} \subseteq S^7 \), of \( 3 \times 3 \) high-contrast patches from natural images. The astonishing fact discovered in their work: 50% of the points in \( \mathcal{M} \) accumulate around a Klein bottle \( K \), in a tubular neighborhood occupying about 20% of the volume of \( S^7 \).

Driven by this newfound model, we propose in this thesis a novel method for representation and classification of image textures. It is shown that for any gray scale image it is possible to project a sample of its high contrast patches onto \( K \), and then study the underlying probability density function

\[
h : K \rightarrow \mathbb{R}
\]

using Fourier analysis. That is, it is possible to construct a suitable orthonormal basis \( \mathcal{B} \) for \( L^2(K) \), and describe \( h \) in terms of Fourier-like coefficients \( h_1, \ldots, h_N, \ldots \), which we show can be estimated with high confidence.

We then proceed on to studying the type of information contained in the estimates \( [\widehat{h}_1, \ldots, \widehat{h}_N] \), and ways in which they can be used for classification purposes. We propose several dissimilarity measures, and test these ideas on the PMTex database, a collection of 1930 image textures which correspond to 36 different texture samples, with instances taken under varying lightning conditions and camera viewpoint.
## Contents

Abstract iv

1 Introduction 1

2 Homology Inference 4
   2.1 Approximation by Simplicial Complexes 5
   2.2 Persistent Homology 9

3 Topology of Spaces of Micro-Images 15
   3.1 The Mumford Formalism 16
   3.2 Persistent Homology on $\mathcal{M}$ 17

4 Projecting Samples Onto $\mathcal{K}$ 21
   4.1 The Differentiable Case 22
      4.1.1 Direction and Directionality 22
      4.1.2 Projection onto $C_\alpha$ 25
   4.2 The Discrete Case 28
      4.2.1 Direction and Directionality 28
      4.2.2 Estimating $t_k$ 34

5 A Discrete Representation 38
   5.1 Coefficient Estimation in $L^2$ 39
   5.2 An Orthonormal Basis for $L^2(K)$ 43
   5.3 The v-Invariant 45
List of Tables

4.1 Estimated thresholds .............................................. 36
7.1 $|T_p|$ and $|N_p|$ as functions of $p$ ............................ 78
List of Figures

2.1 Bar-code of noisy circle .................................................. 13
3.1 Exemplary natural images ................................................. 15
3.2 DCT basis in patch space ................................................. 17
3.3 Bar-code, three circle model ............................................. 18
3.4 Three circle model in patch space ..................................... 18
3.5 Bar-code Klein bottle ..................................................... 19
3.6 Klein bottle model in patch-space .................................... 19
4.1 Exemplary texture in the USC-SIPI database ....................... 34
4.2 Direction versus directionality ......................................... 35
4.3 Histogram of directionalities, USC Texture Database ............. 36
5.1 Texture of Straw .......................................................... 49
5.2 v-Invariant $v_1(S) = (v_1, \ldots, v_d)$ for Straw ................ 49
5.3 Projected sample and heat map for Straw .......................... 50
6.1 Natural multi-scale distance matrix for ensemble 2 .............. 60
6.2 Classes emerging from $d_4$ distance on Ensemble 2 ............ 61
6.3 Evolution function $\varepsilon(I)$ for images in Ensemble 2 ........ 62
6.4 Computational Scheme for the R-distance $d_R$ .................... 69
7.1 Sample materials, PMTex database ................................. 72
7.2 Instances of a sample in the PMTex database ...................... 73
7.3 Subspace distance matrices ............................................ 74
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>Distance matrix representing $(V, d_R)$</td>
<td>75</td>
</tr>
<tr>
<td>7.5</td>
<td>Material categorization from Hausdorff distance.</td>
<td>76</td>
</tr>
<tr>
<td>7.6</td>
<td>Complete-linkage dendrogram for $d_M$</td>
<td>77</td>
</tr>
<tr>
<td>7.7</td>
<td>Classification success $d_4$, $d_R$ and $\tilde{d}_4$</td>
<td>79</td>
</tr>
<tr>
<td>7.8</td>
<td>Intrinsic distance estimated from $d_R$</td>
<td>82</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The application of topological methods to the analysis of data has produced in the last decade an increasing number of success stories. In neuroscience for instance [24], data collected from electrodes in the visual cortex was analyzed using the theory of persistent homology. In this framework, it was shown that homological signatures (Betti numbers) are powerful enough to distinguish between states of visual stimuli. In medicine, simplicial complexes for data reduction and visualization [23], duplicated almost effortlessly results of Reaven and Miller [43] on the onsets of diabetes.

Topological tools have also been employed in computer vision with surprising results [21]: Non-orientable surfaces appear naturally as the state space of relevant micro-images extracted from photographs of natural scenes (natural images). The geometric structure of these spaces lies at the center of the ideas presented in this work, and gives two elements of novelty to our approach: The continuous character of the texton dictionary, in contrast with the quantized nature of its predecessors [6, 7, 38]; and the meaningfulness of the space $\mathcal{K}$ inside the set of all high contrast patches from natural images.

We believe one of the major contributions of this thesis, lies in the reinterpretation we give to the concept of texton dictionary. As seen in the work of Varma and Zisserman [38], one of the most successful approaches in material discrimination from image textures, consists in learning a finite collection of frequently occurring patches from each material class in a given image database, and then agglomerating them
(over all classes) to form the texton dictionary. Some drawbacks are inherent to this philosophy: The resulting code book is database dependent, and given its quantized nature, important information might get lost when trying to associate an element in the dictionary to a patch in an image.

With the Klein bottle, we show that it is possible to construct code books which are both infinite, and of a continuous nature, but with a simple geometry and low intrinsic dimension. Moreover, this new continuous dictionary captures important properties of textures in natural images which are not tied to an specific image data set. In this setting, the distribution on this space of patches from a particular image, becomes an object associated to the image itself. We believe that with this interpretation, it will be easier to discover extensions as well as applications coming from different angles.

The organization of this thesis goes as follows: In Chapter 2, titled **Homology Inference**, we give some background material in the problem of estimating the homology of a topological space from finite samples. We include several constructions, as well as results with conditions under which the homology estimation problem can be solved. This material is presented for both completeness, and as a motivation for the theory of persistent homology, which we include in the second part of the chapter.

Chapter 3, **Topology of Spaces of Micro-Images**, summarizes the results in the papers by Mumford et. al. [3], and Carlsson et. al. [24]. Our objective with this chapter is to explain the way in which the Klein bottle fits as a relevant part of the space of natural micro-images.

In Chapter 4, **Projecting Samples Onto \( \mathcal{K} \)**, we provide a computational framework for evaluating the closest point projection

\[
p : \mathcal{T} \longrightarrow \mathcal{K}
\]

where \( \mathcal{T} \) is a small tubular neighborhood of the Klein bottle \( \mathcal{K} \).

Chapter 5, **A Discrete Representation**, is devoted to the problem of estimating the Fourier-like coefficients associated to a probability density function \( f : \mathcal{K} \longrightarrow \mathbb{R} \), underlying a given projected sample of high-contrast patches. We first go over the general theory of estimation by orthogonal series, and then construct a particular
basis for $L^2(\mathcal{K})$. The way we do the latter, is by noticing that $\mathcal{K}$ is a quotient of the torus $S^1 \times S^1$, and thus Fourier analysis on $S^1$ yields an orthonormal basis for $L^2(\mathcal{K})$. We end chapter 5 with some particular examples of the computations involved.

In Chapter 6, **Dissimilarity Measures**, we construct several (pseudo) metrics in the frequency domain, so that we can compare the different sets of Fourier-like coefficients defined in chapter 5. In particular we show that image rotation has a simple interpretation in the frequency domain, and thus it is possible to define rotation-invariant notions of dissimilarity.

In Chapter 7, **Classification Results**, we show how the theoretical constructions of the previous chapter behave in a particular texture database.

We end with **Conclusions and Final Remarks** in Chapter 8.

**Computational Details**

Computations, data processing and associated plots were generated with MATLAB R2009b, used under the Single Machine License - Stanford Extended Set.
Chapter 2

Homology Inference

In recent years, our capability to collect data has greatly surpassed our ability to analyze it. With data sets sitting in spaces of very large dimension, in highly-nonlinear ways, global and qualitative information are becoming more relevant. In short, methods which are somewhat independent from choices of metric, that can deal with large dimensions, and are robust with respect to noise or missing points, are in high demand. It is in this scenario that topological methods for the analysis of data [19] are becoming relevant.

One of the driving ideas behind topological data analysis is the following: A finite set $X \subseteq \mathbb{R}^n$ can be thought of as a random sample taken from a topological space $\mathcal{X} \subseteq \mathbb{R}^n$, drawn with respect to some probability density function. The hope is then that topological information about $\mathcal{X}$ can be recovered from $X$, and that phenomena encoded in the data can be revealed from the topology.

One of the driving ideas behind topological data analysis is the following: A finite set $X \subseteq \mathbb{R}^n$ can be thought of as a random sample taken from a topological space $\mathcal{X} \subseteq \mathbb{R}^n$, drawn with respect to some probability density function. The hope is then that topological information about $\mathcal{X}$ can be recovered from $X$, and that phenomena encoded in the data can be revealed from the topology.

Take clustering for example: Given a data set $X$ and a notion of similarity or distance between its points, clustering attempts to partition $X$ into meaningful subsets. This of course, can be thought of as the statistical counterpart of finding the connected components of $\mathcal{X}$, which is a topological notion. The higher-level geometric organization of the space can also reveal important information: Holes or voids in $X$, which could translate into lack of coverage in a sensor network, the presence of cancer in medical image data, or periodic phenomena, can be detected by the rank of the homology groups $H_k(\mathcal{X})$. 
In summary, the topology of $X$, or at least its homology, is an important object of study and the question becomes: How do we go from $X \subseteq \mathbb{R}^n$ to $H_k(X)$? We present now some of the answers found in the literature.

### 2.1 Approximation by Simplicial Complexes

The category of compact subsets of Euclidean space with positive \textit{reach}, introduced by Federer in 1959 [27], comprises most of the examples one expects to find in the study of naturally occurring data, at least up to homotopy equivalence. Intuitively, a compact subset of $\mathbb{R}^n$ has positive reach if it does not have sharp cone-like points and examples include, but are not limited to, all smoothly embedded closed manifolds.

**Definition 2.1.1.** Given a compact subset $K$ of $\mathbb{R}^n$, its medial axis $m(K)$ is the set of points $x \in \mathbb{R}^N$ having at least two closest points in $K$. The reach of $K$, $rch(K)$, is the distance from $K$ to $m(K)$.

Since local smoothing does not change the homotopy type, it follows that the class of compact sets with positive reach is general enough for a theory of homology inference from point cloud data. What we will see now is that in addition to generality, compact subsets with positive reach can be reconstructed (up to deformation) from densely enough samples.

**Proposition 2.1.2.** Let $X \subseteq \mathbb{R}^N$ be a compact set with $R = rch(X) > 0$, and let $X \subseteq X$ be $\epsilon$-dense. That is, for all $x \in X$ there exists $y \in X$ so that $\|x - y\| < \epsilon$. If $\epsilon < (3 - \sqrt{8})R$ and $\epsilon + \sqrt{R\epsilon} < r < R - \sqrt{R\epsilon}$ then

$$X^r = \bigcup_{x \in X} B(x, r)$$

deformation retracts onto $X$. $B(x, r)$ denotes the open ball of radius $r$ centered at $x$.

**Proof.** The inequalities $\epsilon + \sqrt{R\epsilon} < r < R - \sqrt{R\epsilon}$, which make sense given that $\epsilon < (3 - \sqrt{8})R$, imply that

$$r < R - \epsilon \quad \text{and} \quad R - \sqrt{(R - \epsilon)^2 - r^2} < r - \epsilon.$$
The result is now a consequence of Theorem 4 in [11].

It follows that the homology of $X$ coincides with that of $X^r$, for $r$ in certain range depending on both sample density and reach. The computation of $H_k(X^r, \mathbb{Z})$, on the other hand, reduces to the problem of calculating the Smith Normal Form over a Principal Ideal Domain. Indeed, let $\mathcal{C}(X, r)$ be the abstract simplicial complex on $X$ so that $\{x_0, \ldots, x_k\} \in \mathcal{C}(X, r)$ if and only if

$$B(x_0, r) \cap \ldots \cap B(x_k, r) \neq \emptyset.$$ 

We refer to $\mathcal{C}(X, r)$ as the r-Czech complex on $X$. As a consequence of the Nerve Theorem [9], and the equivalence of simplicial and singular homology [2] we have:

**Theorem 2.1.3 (Čech Reconstruction).** The geometric realization of the Čech complex, $|\mathcal{C}(X, r)|$, is homotopy equivalent to $X^r$. Thus, the simplicial homology of $\mathcal{C}(X, r)$ coincides with the singular homology of $X$, provided $X$ is $\epsilon$-dense in $X$, $R = rch(X) > 0$, $\epsilon < (3 - \sqrt{8})R$ and $\epsilon + \sqrt{R\epsilon} < r < R - \sqrt{R\epsilon}$.

There is a large body of work dedicated to the efficient computation of simplicial homology over Principal Ideal Domains, via the Smith Normal Form of the corresponding boundary matrix [31].

The previous reconstruction theorem while encouraging, puts forward many of the difficulties one encounters when studying point clouds coming from real world data. Consider for instance the construction of $\mathcal{C}(X, r)$. Since this complex takes as vertex set all data points, it is in general extremely large, with simplices of dimension further exceeding even that of the ambient space. In practice, it only takes a few thousand points to make the homology computation too expensive in terms of memory resources. As an attempt to remedy this problem, de Silva and Carlsson introduced the Witness Complex in [48]. The main idea behind their construction is that it is often unnecessary to use every point in the data set as a vertex, and thus one can use a much smaller, sparse and well distributed subset, while letting the remaining data points guide the construction of the complex.
Definition 2.1.4. Let $(X,d)$ be a metric space, and let $L \subseteq X$ be finite. A point $w \in X - L$ is called an $r$-witness for $\sigma \subseteq L$ if

$$\max_{l \in \sigma} d(w,l) \leq r + d(w,L - \sigma)$$

The r-Witness Complex $W(X,L,r)$ is the abstract simplicial complex on $L$ so that $\emptyset \neq \sigma \subseteq L$ is a simplex if and only if every $\emptyset \neq \tau \subseteq \sigma$ has an $r$-witness in $X - L$.

By a theorem of de Silva (Corollary 9.7 in [47]), $W(X,L) = W(X,L,0)$ recovers the homotopy type of $X$ under some density conditions for $L$, and of curvature for $X$.

Theorem 2.1.5 (Witness Reconstruction). Let $X$ be a complete Riemannian manifold with constant sectional curvature and injectivity radius $\rho_X$. Suppose that $L \subseteq X$ is $\epsilon$-dense (with respect to the geodesic distance), where $\epsilon \leq \rho_X/3$ or $\epsilon \leq \rho_X/6$ if $X$ is the sphere of dimension at least 2. Then $|W(X,L)|$ is homotopy equivalent to $X$.

It is known that for smoothly embedded curves on the plane and surfaces in $\mathbb{R}^3$, $|W(X,L)|$ is homeomorphic to $X$ under some mild sampling conditions on $L$ [12, 39, 40]. Thus, the requirement of constant sectional curvature can be discarded in low dimensions. By contrast, due to the presence of badly shaped simplices, called slivers, such guarantees are unavailable for arbitrary smooth submanifolds of $\mathbb{R}^n$ with dimension $k \geq 3$. This is true, unfortunately, even under strong sampling conditions [45]. This difficulty in higher dimensions can be overcome, as suggested in [30], using an enriched version of the Witness Complex, where vertices are assigned weights in order to prevent the formation of slivers.

Another important construction related to the Čech complex, is the Vietoris-Rips complex.

Definition 2.1.6. Let $(X,d)$ be a pseudo-metric space and let $r > 0$. The Vietoris-Rips complex on $X$ at scale $r$, denoted by $VR(X,r)$, is the abstract simplicial complex on $X$ so that $\{x_0, \ldots, x_k\} \in VR(X,r)$ if and only if $d(x_i,x_j) < r$ for all $0 \leq i, j \leq k$. 

When $X \subseteq \mathbb{R}^N$ is finite, $VR(X, r)$ is the largest simplicial complex having the same 1-skeleton as $\hat{C}(X, r)$, and one can check that

$$VR(X, r) \subseteq \hat{C}(X, 2r) \subseteq VR(X, 2r).$$

Optimal bounds for these inclusions, which turn out to depend on $N$, have been derived by de Silva and Ghrist (Theorem 2.5 in [49]).

While the considerable size of $VR(X, r)$ is often burdensome, this is ameliorated by removing the ambient space from the definition of the complex. As an example of the usefulness of this generalization, let us consider an undirected connected graph: The minimum number of edges joining two nodes induces a metric on the vertex set, and therefore the Vietoris-Rips complex can be used to study the higher-order geometric organization and connectivity of the graph. Many data sets, specially those coming from biology, can be modeled as a graph and thus the importance of the Vietoris-Rips complex.

As observed by Jean-Claude Hausmann [29], this construction has specially nice properties for Riemannian manifolds.

**Theorem 2.1.7 (Rips Reconstruction).** Let $X$ be a Riemannian manifold with positive injectivity radius $r(X)$. If $0 < r \leq r(X)$ then $|VR(X, r)|$ is homotopy equivalent to $X$. Moreover, the homotopy equivalence can be chosen to be natural with respect to the inclusion map $VR(X, r) \hookrightarrow VR(X, r')$, for $r \leq r' \leq r(X)$.

**Proof.** See theorem 3.5 in [29].

As a consequence of the discussion following 3.11 in the same paper, one obtains:

**Proposition 2.1.8.** Let $X$ be a compact Riemannian manifold with injectivity radius $r(X)$, and let $X \subseteq \mathbb{X}$ be $\epsilon$-dense for $0 < \epsilon < \frac{1}{4}r(X)$. Then for every $0 < \eta \leq r(M) - 4\epsilon$, the image of the homomorphism between $k$-th homology groups induced by the inclusion map

$$VR(X, 2\epsilon + \eta) \hookrightarrow VR(X, 4\epsilon + \eta)$$

is isomorphic to $H_k(X, \mathbb{Z})$. 
We have summarized in this section a number of constructions which allow one to recover the homotopy/homology type, of a class of topological spaces underlying a given point cloud $X$ (Čech, Witness and Rips reconstruction Theorems). These guarantees however, depend on several sampling conditions, regularity assumptions and good parameter choices which in reality cannot be checked. Thus the need for theorems of statistical nature, and invariants which are free from parameter choices.

On the statistical front, interesting results have appeared regarding the estimation of Betti numbers from random samples [42], and the study of the topology of random simplicial complexes [35]. With respect to parameter choices (e.g. $r$ in the definition of Č and VR), proposition 2.1.8 can be interpreted as an advancement in the philosophy of homological estimation. Namely, that homology plus functoriality is a more robust invariant than homology alone, and that this approach can be used to avoid parameter choices. We discuss in the following section how this philosophy can be implemented, as well as some results regarding its correctness.

### 2.2 Persistent Homology

A family of simplicial complexes $\{K_r\}_{r \in \mathbb{R}}$, with vertex set $X$ and so that $K_r \subseteq K_{r'}$ whenever $r \leq r'$, is defined by two kinds of data: The spaces $K_r$, and the containment relations between them. As we saw in the Čech, Rips and Witness reconstruction theorems, under some conditions it is enough to consider each space $K_r$ independently, and in a particular range, in order to solve the homology inference problem. On the other hand, as proposition 2.1.8 indicates, by including the data coming from the relations between members of the filtration, one can eliminate some of the stringent hypotheses. Indeed, even if one cannot assure that a given member in the filtration recovers the right topology, perhaps the containment data can.

The idea of observing how topological features evolve with the filtration was first used by Edelsbrunner, Letscher and Zomorodian in [26]. Their approach was to construct a special sequence of simplicial complexes for point clouds in $\mathbb{R}^3$, and rank topological attributes (measured by homology with $\mathbb{Z}/2$ coefficients) based on how long they lasted in the filtration. A generalization of this to general filtrations of
abstract simplicial complexes, as well as structure theorems and a computational framework, was provided by Zomorodian and Carlsson in [4]. We provide in what follows a very brief summary of their ideas. For a more thorough treatment refer to [19] and [25].

**Definition 2.2.1.** Let $\mathcal{C}$ be any category, and let $\mathcal{P}$ be a partially ordered set. $\mathcal{P}$ yields a category $\mathcal{P}$ where $Ob(\mathcal{P}) = \mathcal{P}$, and there exists a unique morphism from $p$ to $q$ if and only if $p$ is less than or equal to $q$ in $\mathcal{P}$. A $\mathcal{P}$-persistent object in $\mathcal{C}$ is a functor

$$F : \mathcal{P} \rightarrow \mathcal{C}$$

and it follows that $\mathcal{P}$-persistent objects in $\mathcal{C}$ along with natural transformations between them, define a category $\mathcal{P}_{pers}(\mathcal{C})$. We refer to $\mathcal{P}_{pers}(\mathcal{C})$ as the category of $\mathcal{P}$-persistent objects in $\mathcal{C}$.

**Example 1:** Let $X$ be a subset of $\mathbb{R}^N$, and let $\{K_r\}_{r \in \mathbb{R}}$ be a nested family of abstract simplicial complexes with vertex set $X$. That is, $K_r \subseteq K_{r'}$ whenever $r \leq r'$. By applying $k$-th homology with coefficients in a group $G$ one gets a functor

$$H : \mathbb{R} \rightarrow Mod_G,$$

$$r \mapsto H_k(K_r; G)$$

and thus an $\mathbb{R}$-persistent object in the category of modules over $G$.

**Example 2:** Let us assume now that $X$ is finite, let $\{K_r\}_{r \in \mathbb{R}}$ be as before, and let $Vect_{\mathbb{F}}$ be the category of vector spaces over a field $\mathbb{F}$. Then it is possible to choose a strictly increasing sequence $r_n \uparrow \infty$ so that $K_{r_n} = K_{r'}$ whenever $r_n < r, r' < r_{n+1}$, and thus we get an $\mathbb{N}$-persistent object in $Vect_{\mathbb{F}}$ given by

$$T : \mathbb{N} \rightarrow Vect_{\mathbb{F}},$$

$$n \mapsto H_k(K_{r_n}; \mathbb{F})$$

Notice that $T(n)$ is finite dimensional for all $n \in \mathbb{N}$, and there exists $M \in \mathbb{N}$ so that $M \leq n \leq m$ implies that the morphism $T(n) \rightarrow T(m)$ is an isomorphism. An $\mathbb{N}$-persistent object in $Vect_{\mathbb{F}}$ with these two properties is said to be **tame**, and it
follows that tame $\mathbb{N}$-persistent objects in $\mathbf{Vect}_\mathbb{F}$ form a subcategory of $\mathbb{N}_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$, which we denote by $\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$.

**Example 3:** Given $X \subseteq \mathbb{R}^N$ finite and $L \subseteq X$, we have that $VR(X, r)$, $\check{C}(X, r)$ and $W(X, L, r)$ yield families of nested simplicial complexes. Indeed, if $r \leq r'$ then $VR(X, r) \subseteq VR(X, r')$, $\check{C}(X, r) \subseteq \check{C}(X, r')$ and $W(X, L, r) \subseteq W(X, L, r')$. Thus the Vietoris-Rips construction, the Čech complex and the Witness complex, give rise to objects in $\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$.

**Example 4:** Let $K$ be an abstract simplicial complex, and let

$$f : K \rightarrow \mathbb{R}$$

be a non-decreasing function, i.e. $f(\tau) \leq f(\sigma)$ whenever $\emptyset \neq \tau \subseteq \sigma$. Then $f$ induces a nested family of simplicial complexes $\{K_r\}_{r \in \mathbb{R}}$, where $K_r = f^{-1}(-\infty, r]$, and thus an object in $\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$.

As the previous examples indicate, given $X \subseteq \mathbb{R}^N$ it is possible to capture the evolution of several simplicial complexes based on $X$, as objects in $\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$. Since the data of homology plus functoriality is encoded in such objects, it is of the upmost importance to understand their representation and classification. What we will see now is that there is a very convenient representation for elements in $\text{Ob}(\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F}))$, so that relevant aspects of the evolution of the previous complexes, can be easily identified as salient features in the representation.

**Theorem 2.2.2.** Let $\mathbb{F}[x]$ be the polynomial ring on a variable $x$ with coefficients in a field $\mathbb{F}$, and let $g\text{Mod}_{\mathbb{F}[x]}$ be the category of non-negatively graded modules over $\mathbb{F}[x]$. Then the functor

$$\theta : \mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F}) \rightarrow g\text{Mod}_{\mathbb{F}[x]}$$

taking $\left(\{V_n\}_{n \in \mathbb{N}}, \{\psi_{n,m} : V_n \rightarrow V_m\}_{n \leq m}\right)$ to $\bigoplus_n V_n$, with module structure on homogenous elements $v \in V_n$ given by $x^{m-n} \cdot v = \psi_{n,m}(v)$, is an equivalence of categories.

Moreover, $\theta$ identifies $\mathbb{N}^t_{\text{pers}}(\mathbf{Vect}_\mathbb{F})$ with the category of finitely generated non-negatively graded modules over $\mathbb{F}[x]$. 
Since \( \mathbb{F}[x] \) is a Principal Ideal Domain, then a graded version of the structure theorem for finitely generated modules over PID’s (see theorem 7.5 in [44]) implies:

**Theorem 2.2.3.** Let \( T \) be an object in \( \mathbb{N}_{\text{pers}}(\text{Vect}_{\mathbb{F}})^t \). Then there exists an isomorphism of graded \( \mathbb{F}[x] \)-modules

\[
\theta(T) \cong \bigoplus_{i=1}^{R} \sum_{r=1}^{i_r} \mathbb{F}[x] \oplus \bigoplus_{s=1}^{S} \sum_{j=1}^{j_s} \left( \mathbb{F}[x]/(x^{n_s}) \right)
\]

where \( n_s \leq n_{s+1} \) and \( \sum^j \) is the operator which shifts grading by \( j \).

That is, \( T \in \text{Ob}\left( \mathbb{N}_{\text{pers}}(\text{Vect}_{\mathbb{F}})^t \right) \) can be identified with \( \theta(T) \) and the latter is uniquely determined by the set of intervals

\[
\left\{ [i_r, \infty), [j_s, j_s + n_s) \mid r = 1, \ldots, R, s = 1, \ldots, S \right\}
\]

where \([i_r, \infty)\) is associated to the free summand \( \sum_{i_r} \mathbb{F}[x] \), and \([j_s, j_s + n_s)\) corresponds to the torsion component \( \sum_{j_s} \left( \mathbb{F}[x]/(x^{n_s}) \right) \).

Let us see how this representation applies to the problem of homology inference from point cloud data. Following examples 2 and 3, given \( X \subseteq \mathbb{R}^N \) we have a filtration by simplicial complexes with vertex set \( X \)

\[
\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_M = K_{M+1} = \cdots
\]

and for any \( k \in \mathbb{N} \), we get the tame \( \mathbb{N} \)-persistent object in \( \text{Vect}_{\mathbb{F}} \)

\[
H_k(K_0; \mathbb{F}) \to H_k(K_1; \mathbb{F}) \to \cdots \to H_k(K_M; \mathbb{F}) \to H_k(K_{M+1}; \mathbb{F}) \to \cdots
\]

Thus we obtain the graded \( \mathbb{F}[x] \)-module \( \bigoplus_i H_k(K_i; \mathbb{F}) \), and the \( \mathbb{F}[x] \)-isomorphism

\[
\bigoplus_i H_k(X_i; \mathbb{F}) \cong \bigoplus_{i=1}^{R} \sum_{r=1}^{i_r} \mathbb{F}[x] \oplus \bigoplus_{s=1}^{S} \sum_{j=1}^{j_s} \left( \mathbb{F}[x]/(x^{n_s}) \right)
\]

yields bases \( B_i \) for each \( H_k(K_i; \mathbb{F}) \), so that if \( \psi_{i,j} : H_k(K_i; \mathbb{F}) \to H_k(K_j; \mathbb{F}) \) is the
CHAPTER 2. HOMOLOGY INFERENCE

homomorphism induced by the inclusion $K_i \subseteq K_j$, $i \leq j$, then $\psi_{i,j}(v) \in B_j \cup \{0\}$ for all $v \in B_i$. The corresponding set of intervals

$$\left\{(i_r, \infty), [j_s, j_s + n_s) \mid r = 1, \ldots, R, s = 1, \ldots, S\right\}$$

which we refer to as the bar-code, can now be interpreted as follows:

- The interval $[b, \infty)$ indicates that there exists $v \in B_b$ so that $v \notin \text{im}(\psi_{b-1,b})$, and $\psi_{b,d}(v) \neq 0$ for all $d \geq b$.

- The interval $[b, d)$ for $d$ finite, indicates that $v \notin \text{im}(\psi_{b-1,b})$, $\psi_{b,d-1}(v) \neq 0$ and $\psi_{b,d}(v) = 0$.

In cases like $K_i = VR(X, r_i)$ (or $\check{C}(X, r_i)$), it is more convenient to consider $[r_b, r_d)$ instead of $[b, d)$. With this modification, the previous interpretation for the interval $[b, d)$ translates into a homological feature which is born at $r_b$, and then dies at $r_d$, where $r$ can be thought of as a time parameter.

The bar-code associated to a filtration of simplicial complexes, can be summarized in a diagram as the one in figure 2.1.

![Figure 2.1: Example of the bar-code for a noisy circle](image)

The intuition behind persistent homology is then that long intervals should indicate real features, while short ones should be regarded as topological noise.
Frédéric Chazal and Steve Y. Oudot have proved several theorems with guarantees for bar-codes from the Čech, Witness and Vietoris-Rips filtrations, for a very general class of spaces, and for samples including noise (Theorems 3.5, 3.6 and 3.7 in [18]). What the authors show is that if $X$ is a compact subset of $\mathbb{R}^n$, then topological noise in the respective bar-codes can be quantified in terms of the Hausdorff distance between $X$ and $X$ (and that between $L \subseteq X$ and $X$ for the Witness complex), provided this distance is small enough with respect to the weak feature size of $X$, a notion generalizing $rch(X)$. As a result, one recovers the homology of $X$ by removing the noise from the bar-codes.

Given its apparent complexity, it is perhaps somewhat surprising that there exists a fast algorithm for computing the persistent homology of a filtration of simplicial complexes. It is shown in [4] by Carlsson and Zomorodian, that computing bar-codes is in the worst case as complicated as Gaussian elimination over a Field. An implementation of their algorithm comes with the PLEX package, available at http://comptop.stanford.edu/u/programs/jplex/index.html.
Chapter 3

Topology of Spaces of Micro-Images

In [3], Mumford et. al. initiated the study of spaces of patches extracted from natural images, with respect to their geometric properties (disregarding both contrast and brightness), and distribution in the state space $S^7$. The van Hateren Natural Image Database [32], is a collection of approximately 4,000 monochrome, calibrated images, photographed by Hans van Hateren around Groningen (Holland), in town and the surrounding countryside.

Figure 3.1: Exemplary images from the van Hateren Natural Images Database.
CHAPTER 3. TOPOLOGY OF SPACES OF MICRO-IMAGES

3.1 The Mumford Formalism

For each image in the database, 5,000 $3 \times 3$ patches were selected at random, and then the logarithm of each entry in the patch is calculated. According to Weber's law, the ratio $\Delta L / L$ between the just noticeable difference $\Delta L$ and the ambient luminance $L$, is constant for a wide range of values of $L$. Since it is believed that the human visual system uses this adaptation, Mumford et. al. attempt to model the set of patches as close as possible to the way we perceive them.

Next, from the 5,000 $3 \times 3$ patches of a single image, the top 20 percent with respect to contrast is selected. Contrast is measured by means of the $D$-norm $\| \cdot \|_D$, where an $n \times n$ patch $P = [p_{ij}]$ yields the vector $v = [v_1, \ldots, v_{n^2}]^t$ given by $v_{n(j-1)+i} = p_{ij}$, and one defines

$$\|P\|_D^2 = \sum_{r \sim t} (v_r - v_t)^2.$$ 

Here $v_r = p_{ij}$ is related ($\sim$) to $v_t = p_{kl}$, if and only if $p_{ij}$ is the 4-connected neighborhood of $p_{kl}$, i.e. if and only if $|i-k| + |j-l| \leq 1$. The combinatorial data in the relation $\sim$ can be encoded in a symmetric $n \times n$ matrix $D$, and it follows that $\|P\|_D^2 = \langle v, Dv \rangle$. In the case $n = 3$ the matrix $D$ takes the form:

$$D = \begin{bmatrix}
2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\
0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\
0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2
\end{bmatrix}$$

The vectors from the resulting log-intensity, high-contrast patches are centered by subtracting their mean, and normalized by dividing by their $D$-norm. This yields approximately $4 \times 10^6$ points in $\mathbb{R}^9$, on the surface of a 7-dimensional ellipsoid.
There is a specially convenient basis for $\mathbb{R}^9$ of eigenvector of $D$ called the DCT-basis. This basis consists of the constant vector $[1, \ldots, 1]^t$, which is perpendicular to the hyperplane containing the centered and normalized data, and eight vectors which in patch space take the form:

![Figure 3.2: DCT basis in patch space](image)

By taking the coordinates with respect to the DCT-basis of the vectors in the aforementioned 7-ellipsoid, we get a subset $\mathcal{M}$ of the 7-sphere $S^7$.

### 3.2 Persistent Homology on $\mathcal{M}$

Using the formalism of persistent homology, Carlsson et. al. [21] studied the structure of (topological spaces represented by) subsets of $\mathcal{M}$ with high-density. To this end, they let $k \in \mathbb{N}$, $p \geq 0$ and define

$$\rho_k : \mathcal{M} \rightarrow \mathbb{R}$$

$$x \mapsto \|x - x_k\|$$

where $x_k \in \mathcal{M}$ is the $k$-th nearest neighbor of $x$ in $\mathcal{M}$. It follows that $\rho_k$ is inversely proportional to the density at $x$, and the value of $k$ determines the local or global character of the estimate. Let $X(k, p)$ be the set of points in $\mathcal{M}$ whose $\rho_k$-value, falls in the bottom $p$ percent.

The topology of the space $X(k, p)$ for different values of $k, p$ revealed very interesting structure: For $(k, p) = (1200, 30)$, the set $X(1200, 30)$ approximates a topological space modeled by a wedge of five circles, as shown by the respective bar-codes reported in figure 3.3.
Figure 3.3: Bar-code for $X(1200, 30)$ shows the topology of a wedge of five circles.

With respect to patches, this space can be realized as the three circle model:

Figure 3.4: (a) A space with the homotopy type of a wedge of 5 circles. The space is made up of three circles, of which two are disjoint, and the third intercepts each of the others at two points. (b) Patches included in the three circle model.

The natural question arising from these findings is whether or not one could discover in the data, a higher dimensional entity containing the 3-circle model. The answer was affirmative, and the topological space that emerged was a familiar surface:
The set \( X(100, 10) \) has topological signatures consistent with either the Klein bottle or the torus:

![Bar-code for \( X(100, 10) \) with \( \mathbb{Z}/2 \)-coefficients reveals a surface.](image)

(a) \( H_1(\cdot; \mathbb{Z}/2) \)-persistence

(b) \( H_2(\cdot; \mathbb{Z}/2) \)-persistence

Figure 3.5: Bar-code for \( X(100, 10) \) with \( \mathbb{Z}/2 \)-coefficients reveals a surface.

In order to decide between the Klein bottle and the torus, one only needs to change the field of coefficients to \( \mathbb{F}_3 \), and repeat the persistent-homology computation. With this change, the Klein bottle emerged, and a closer study revealed the model:

![Space of patches corresponding to \( X(100, 10) \) parametrizes a Klein bottle.](image)

Figure 3.6: Space of patches corresponding to \( X(100, 10) \) parametrizes a Klein bottle.
This space can be understood as the set of 3-by-3 patches, whose intensity function is given by polynomials

\[ p(x, y) = c(ax + by) + d(ax + by)^2 \]

where \( a^2 + b^2 = c^2 + d^2 = 1 \). Let this set of polynomials be denoted by \( \mathcal{K} \).

The interpretation of the polynomial model is as follows: The angle \( \alpha \in [\pi/4, 5\pi/4) \) satisfying

\[ e^{i\alpha} = a + ib \]

captures the preferred direction of the patch, and it is parametrized by the horizontal direction on the Klein bottle model. The vertical direction, given by an angle \( \beta \in [-\pi/2, 3\pi/2) \) so that

\[ e^{i\beta} = c + id \]

determines how linear or quadratic the patch will be. Using Monte Carlo simulations, it was shown that 50% of the points in \( \mathcal{M} \) accumulate around \( \mathcal{K} \), in a tubular neighborhood occupying about 20% of the volume of \( S^7 \). This should be contrasted with the fact that \( \mathcal{M} \) accounts for 84% of the total volume in the 7-sphere.

With this interpretation of the polynomial model \( \mathcal{K} \), we will give in the next chapter a solution to the follow problem: Given a natural image, most of its high-contrast patches should be in a small tubular neighborhood \( \mathcal{T} \) of \( \mathcal{K} \). How do we compute the projection map \( \Phi : \mathcal{T} \rightarrow \mathcal{K} \)?
Chapter 4

Projecting Samples Onto $\mathcal{K}$

As discussed in the previous chapter, the Klein bottle model captures two patch features. Namely, the first coordinate measures the preferred direction of the patch, and once the direction is identified, the second coordinate captures how quadratic or linear the patch is. Following this insight, we present in this chapter a method for associating to most high-contrast patches in an image, a point on the idealized Klein bottle $\mathcal{K}$. Namely, we show that for most high-contrast patches it is possible to measure its preferred direction as an angle $\alpha \in [\pi/4, 5\pi/4)$, and given the vector

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix}$$

one can compute the orthogonal projection of the patch (regarded as a function of two variables $x, y$) onto the space of polynomial functions

$$\mathcal{C}_\alpha = \{ c(ax + by) + d(ax + by)^2 \mid c^2 + d^2 = 1 \}$$

provided the patch is directional enough.

As a methodological approach, we will first develop the concepts of direction, directionality and projection onto $\mathcal{C}_\alpha$ for patches represented by differentiable functions. Having this, we will extend the notions to discrete ones. Notice that this method can be regarded as a continuous version of the Maximum Response 8 (MR8) filter.
4.1 The Differentiable Case

4.1.1 Direction and Directionality

Let \( I = [-1, 1] \). A square patch from an image in gray scale can be thought of as a function

\[
f : I^2 \rightarrow I
\]

where \( f(x, y) \) is the intensity at the pixel \((x, y)\), \( f(x, y) = -1 \) if \((x, y)\) is a purely black pixel and \( f(x, y) = 1 \) if \((x, y)\) is a purely white pixel.

**Definition 4.1.1.** A patch \( f : I^2 \rightarrow I \) is said to be purely directional, if there exists a function \( g : [-\sqrt{2}, \sqrt{2}] \rightarrow I \) and a unitary vector \([a, b]\) so that

\[
f(x, y) = g(ax + by)
\]

for all \((x, y) \in I^2\).

Notice that such an \( f \) will be constant along lines parallel to \([-b, a]\).

**Theorem 4.1.2.** Let \( f : I^2 \rightarrow I \) be a differentiable patch. Then \( Q_f : \mathbb{R}^2 \rightarrow \mathbb{R} \) given by

\[
Q_f(v) = \int \int_{I^2} \langle \nabla f, v \rangle^2 dx dy
\]

is a positive semi-definite quadratic form so that:

1. If \( f(x, y) = g(ax + by) \), with \( a^2 + b^2 = 1 \) then

\[
Q_f(a, b) \geq Q_f(v)
\]

for all \( \|v\| = 1 \).

2. Let

\[
j : S^1 \rightarrow \mathbb{RP}^1
\]

\[
v \mapsto \{v, -v\}
\]
be the projection map and assume the eigenvalues of \( A_f \), the matrix representing \( Q_f \), are distinct. Then the direction map

\[
\text{Dir}(f) = j\left( \arg \max_{\|v\|=1} Q_f(v) \right)
\]

is a well defined continuous function of \( f \) in the \( C^1 \) topology. Moreover, the maximum is attained at the eigenvectors of \( A_f \) with largest eigenvalue.

Proof.

1. If \( f(x, y) = g(ax + by) \), then we have the identity

\[
b \frac{\partial f}{\partial x} = a \frac{\partial f}{\partial y}
\]

and therefore \( Q_f(-b, a) = 0 \). Notice that by positive semi definiteness of \( Q_f \), we conclude that \( Q_f \) attains its global minimum on \( S^1 \) at \( \begin{bmatrix} -b \\ a \end{bmatrix} \). On the other hand, by the spectral theorem, there exists an orthogonal matrix \( B \) so that if \( \lambda_1, \lambda_2 \) are the eigenvalues of \( A_f \), we can write

\[
A_f = B^t \cdot \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \cdot B.
\]

Since \( B \) is orthogonal, \( \|v\| = 1 \) implies \( \|Bv\| = 1 \) and if \( Bv = \begin{bmatrix} r \\ s \end{bmatrix} \) we have

\[
Q_f(v) = (1 - s^2)\lambda_1 + s^2\lambda_2
\]

It follows that the minimum of \( Q_f \) on the unit circle is the smallest eigenvalue of \( A_f \), and it is attained at the unitary vectors in its eigenspace.

Likewise, the largest eigenvalue of \( A_f \) is the maximum of \( Q_f \), which is attained at any unitary vector in its eigenspace. Therefore, if \( Q_f(-b, a) \) is the minimum, \( Q_f(a, b) \) is the maximum and \( Q_f(a, b) \geq Q_f(v) \) for all \( \|v\| = 1 \).
2. As it was shown (1), \( \max_{\|v\|=1} Q_f(v) \) is attained at the unitary eigenvectors in the eigenspace of the largest eigenvalue of \( A_f \). Since the eigenvalues of \( A_f \) are distinct, their eigenspaces are one dimensional and it follows that the unitary vectors in each eigenspace are antipodal. Thus \( \text{Dir} \) is well defined.

\( \square \)

Remarks:

1. From here on, we will assume that \( \iint \|\nabla f\|^2 dxdy = 1 \) for all differentiable patches \( f : I^2 \rightarrow I \) in consideration. This condition is related to normalization by the \( D \)-norm. Indeed, the \( D \)-norm is a discrete version of

\[
F \mapsto \sqrt{\iint \|\nabla f\|^2 dxdy}
\]

which is the unique scale-invariant norm on images \( f(x, y) \).

2. If we let \( S^1 = [0,2\pi]/\sim \), where 0 is identified with 2\( \pi \), then \( \mathbb{R}P^1 \) can be identified with \( [\pi/4,5\pi/4]/\sim \) where \( \pi/4 \sim 5\pi/4 \). Via these identifications, we will regard \( \text{Dir} \) as a function onto \( [\pi/4,5\pi/4) \).

3. The condition of \( A_f \) not being a scalar matrix cannot be dropped in order to define the direction of \( f \). Consider the patch \( f(x, y) = \sqrt{\frac{32}{3}} (x^2 + y^2) \). It follows that

\[
\iint \|\nabla f\|^2 dxdy = 1 \text{ and } \iint (\nabla f, \begin{bmatrix} a \\ b \end{bmatrix})^2 dxdy = \frac{1}{2}
\]

for all \( a^2 + b^2 = 1 \) and therefore \( Q_f = (1/2)Id \). Notice that \( f \) is a radial patch, and therefore its direction is not well defined.

4. For each differentiable patch \( f \), we can define its directionality

\[
dir(f) = |\lambda_1 - \lambda_2|
\]

where \( \lambda_1, \lambda_2 \) are the eigenvalues of \( A_f \).
CHAPTER 4. PROJECTING SAMPLES ONTO $K$

It follows that $\text{dir}(f)$ is maximized when $f$ is a purely directional patch, for radial patches $\text{dir}(f) = 0$, and therefore $\text{dir}(\cdot)$ measures how directional a patch is.

A discrete version of $\text{dir}(f)$ was used in [20] as a thresholding function, (along with two other quantities measuring characteristics of a patch) in order to recover the high density regions in $\mathcal{M}$.

5. The same results described above remain true for differentiable functions

$$f : I^2 \longrightarrow [u, v]$$

where $-\infty < u < v < \infty$. In particular, when $f$ is a log-intensity patch.

4.1.2 Projection onto $C_\alpha$

Let $\alpha \in [\pi/4, 5\pi/4)$, $e^{i\alpha} = a + ib$, $u = ax + by$ and $v = -bx + ay$ for $(x, y) \in I^2$. Let

$$U = U_\alpha = \{(ax + by, -bx + ay) \mid (x, y) \in I^2\}$$

and let $L^2(U)$ be the space of real valued square-integrable functions on $U$, equipped with the inner product

$$\langle f, g \rangle = \int_U f(u, v)g(u, v)d\mu.$$ 

From the Stone-Weierstrass theorem and the fact that continuous functions are dense in $L^2$, we conclude that the set of polynomials in the variables $u, v$ is dense in $L^2(U)$.

Let $f \in L^2(U)$ be so that $\langle f, u^n v^m \rangle = 0$ for all $n, m \in \mathbb{N} \cup \{0\}$ and let us assume that $\|f\| > 0$. Thus, there exists a polynomial $p(u, v)$ so that $\|f - p\| < \frac{\|f\|}{2}$ and from the orthogonality relations $\langle f, u^n v^m \rangle = 0$ we get

$$\|f\|^2 = \|f - p\|^2 - \|p\|^2 < \frac{\|f\|^2}{4} - \|p\|^2$$
which is a contradiction. This shows that by applying Gram-Schmidt to the set of monomials \( \{u^n v^m \mid n, m \geq 0 \} \) with indexing \((n, m) \mapsto (n + m)(n + m + 1)/2 + n\), we obtain a complete orthonormal subset \( S \) of \( L^2(U) \).

It follows that any \( f \in L^2(U) \) can be polynomially represented as

\[
\sum_{s \in S} \langle f, s \rangle s.
\]

Now, for each function \( f : I^2 \to I \) one can define \( f_\alpha(u, v) = f(au - bv, bu + av) \) and when \( f \) is a differentiable patch, which is also purely directional with direction \( \alpha \), we have

\[
\frac{\partial f_\alpha}{\partial v} = 0.
\]

From this perspective, \( \alpha = Dir(f) \) can be interpreted as the angle that in average, makes \( f_\alpha \) as independent from \( v \) as possible. It follows that one can represent \( f_\alpha \) in terms of polynomials \( s \in S \) depending only on \( u \), and if one further requires the patch to be centered, that is \( \int_U f_\alpha d\mu = 0 \), then the degree two polynomial approximation becomes

\[
f_\alpha \approx \langle f_\alpha, u \rangle \frac{u}{\|u\|^2} + \langle f_\alpha, u^2 \rangle \frac{u^2}{\|u^2\|^2}
\]

\[
= r_\alpha (c(ax + by) + d(ax + by)^2)
\]

for some \( r_\alpha \geq 0 \) and \( c^2 + d^2 = 1 \). This follows by tracing back the Gram-Schmidt orthonormalization on the set \( \{1, v, u, v^2, uv, u^2\} \) and checking that \( u \) and \( u^2 \) are perpendicular.

**Definition 4.1.3.** Let \( f : I^2 \to I \) be a differentiable patch so that \( dir(f) > 0 \). If \( Dir(f) = \alpha \), then \( f_\alpha \approx r_\alpha (c(ax + by) + d(ax + by)^2) \) is its degree two polynomial approximation and \( r_\alpha > 0 \), then

\[
\Phi(f) = c(ax + by) + d(ax + by)^2
\]

is the projection of \( f \) onto \( C_\alpha \).
Remark: Even when \( C_\alpha \) is not a linear subspace of \( L^2(U) \), we use the term “projection” in the sense that \( c(ax + by) + d(ax + by)^2 \) attempts to minimize the \( L^2 \)-distance from \( f_\alpha \) to \( C_\alpha \). Since in the case of linear subspaces of \( L^2(U) \) the concepts of orthogonal projection and distance minimizer coincide, then the abuse of language is mildly justified.

Now, since \( K = \bigcup C_\alpha \) then by composing with the inclusion \( C_\alpha \subseteq K \), we get the projection of \( f \) onto the idealized Klein bottle \( K \).

We will now describe a model for \( K \) which is more amenable for computations. Let \( S^1 \subseteq \mathbb{C} \) be the unit circle inside the complex plane, and let \( T = S^1 \times S^1 \) be the torus. Notice that \( T \) is equipped with a free \( \mathbb{Z}/2\mathbb{Z} \) action, defied by

\[
1 \cdot (z, w) = (-z, \overline{w})
\]

with orbit space homeomorphic to \( K = [\pi/4, 5\pi/4] \times [-\pi/2, 3\pi/2]/\sim \), where \( \sim \) stands for the identification of \( (x, -\pi/2) \) with \( (x, 3\pi/2) \) and that of \( (\pi/4, \pi + y) \) with \( (5\pi/4, \pi - y), |y| \leq \pi \). It follows that \( K \) is the usual model for the Klein bottle and we have a quotient map

\[
q : S^1 \times S^1 \longrightarrow K.
\]

It is this quotient map and the induced homomorphism \( q^* : L^2(K) \longrightarrow L^2(S^1 \times S^1) \) which will allow us to do Fourier analysis on the Klein bottle.

**Definition 4.1.4.** Let \( \Phi(f) = c(ax + by) + d(ax + by)^2 \) and let \( (\alpha, \beta) \in K \) be so that \( e^{i\alpha} = a + ib \) and \( e^{i\beta} = c + id \). We let

\[
\Psi(f) = (\alpha, \beta) \in K
\]

be the projection of \( f \) onto \( K \).
4.2 The Discrete Case

4.2.1 Direction and Directionality

Let $P$ be a $3 \times 3$ patch represented by the $3 \times 3$ matrix $P = [p_{ij}]$, and let us identify it with the function that takes the value $p_{ij}$ on the region

$$R_{ij} = \{(x,y) \in I^2 : |x - r_j| + |y - r_i| < 1/3\} \ r_j = 2(j - 2)/3, \ r_i = 2(2 - i)/3.$$

A delicate step toward extending the definition of the direction of a differentiable patch to a discrete one, is the choice of discretization for the gradient. For the integral we use

$$\iint F(x,y)dxdy \approx \sum_{ij} F(r_j, r_i).$$

Let us show with an example why this is a subtle issue. If we discretize the partial derivatives of $P$ using the formulae:

$$P_x(x,y) = \begin{cases} p_{i2} - p_{i1} & \text{if } (x,y) \in R_{i1} \\ \frac{p_{i3} - p_{i1}}{2} & \text{if } (x,y) \in R_{i2} \\ p_{i3} - p_{i2} & \text{if } (x,y) \in R_{i3} \end{cases}$$

(the ones for $P_y$ being similar) then for the patch

which can be represented (up to a positive multiple) by the matrix

$$F = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$
we get that
\[ A_F(x, y) = \begin{bmatrix} 9/2 & 0 \\ 0 & 9/2 \end{bmatrix} \]
and therefore, every vector in the unit circle corresponds to a unitary vector in the eigenspace of the largest eigenvalue of \( A_F \). One would like, however, this patch to have direction \( e^{3\pi i/4} \). If instead, using the first order approximation
\[
\nabla f(dx) \approx \nabla f(0) + \text{Hess } f(0) dx
\]
we let
\[
\nabla P(x, y) = \nabla P(0, 0) + \text{Hess } P(0, 0) \cdot \begin{bmatrix} j - 2 \\ 2 - i \end{bmatrix}, \quad (x, y) \in R_{ij}
\]
where
\[
\nabla P(0, 0) = \frac{1}{2} \begin{bmatrix} p_{23} - p_{21} \\ p_{12} - p_{32} \end{bmatrix}
\]
\[
\text{Hess } P(0, 0) = \frac{1}{2} \begin{bmatrix} p_{23} - 2p_{22} + p_{21} & \frac{p_{13} - p_{11} - p_{33} + p_{31}}{2} \\ \frac{p_{13} - p_{11} - p_{33} + p_{31}}{2} & p_{12} - 2p_{22} + p_{32} \end{bmatrix}
\]
then applying this definition to our example we get
\[
Q_F(x, y) = \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 5/2 & -2 \\ -2 & 5/2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.
\]
Notice that the unitary vector corresponding to the largest eigenvalue is now \( e^{3\pi i/4} \).

The conclusion from this example is that when trying to approximate the gradient at a point, one should use as many neighboring pixels as possible. This, to ensure that the global direction is measured correctly. We will study in the next paragraphs how to carry out this computations for patches of size \( k \times k \), for some choices of \( k > 3 \).
An important feature of natural images is their multi-scale hierarchical organization. In textures for example, one encounters that observations at various distances may reveal different patterns. It is for this reason that robust invariants should include multi-scale information, and ways of tracking the scale evolution of features.

The distribution on the Klein bottle of high-contrast patches coming from an image, can be turned into a multi-scale invariant by analyzing the evolution of the distribution as the number of pixels increases. We have chosen here to study patches of size \((3(2n-1)) \times (3(2n-1))\) for \(n = 1, 2, 3, 4, 5\). The reasons behind this choice are the following: When trying to describe a discretization for the gradient, there are two things to keep in mind. On the one hand, when approximating the partial derivatives at a point one should use as much information as possible from neighboring pixels, so that the global direction is measured accurately. On the other hand, the derivative is only a local concept and therefore its computation should not involve interactions between distant pixels. Thus, we choose sizes that allows us to make the computations locally and the localization coherent as we move throughout the patch.

Given a patch \(P\) of size \((3(2n-1)) \times (3(2n-1))\), this can be divided into disjoint \(3 \times 3\) sub-patches, whose centers we label. For any pixel \(p\) in \(P\), there exists a unique labeled center \(c\) closest to it, and using this center we compute the first order approximation

\[
\nabla P(p) = \nabla P(c) + HessP(c) \cdot u
\]

where \(u \in \{-1, 0, 1\} \times \{-1, 0, 1\}\) is the position of \(p\) with respect to \(c\).

If by way of example we represent the \(3 \times 3\) sub-patch with center \(c\) containing \(p\) (in position \((1, 1)\)) as

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(c_{-1,1})</td>
<td>(c_{0,1})</td>
<td>(p)</td>
</tr>
<tr>
<td>(c_{-1,0})</td>
<td>(c)</td>
<td>(c_{1,0})</td>
</tr>
<tr>
<td>(c_{-1,-1})</td>
<td>(c_{0,-1})</td>
<td>(c_{1,-1})</td>
</tr>
</tbody>
</table>
then \( u = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \) and \( \nabla P(c) = \frac{1}{2} \begin{bmatrix} c_{1,0} - c_{-1,0} \\ c_{0,1} - c_{0,-1} \end{bmatrix} \).

An expression for \( HessP(c) \) is a little more involved. In general, a patch \( P \) of size \((3(2n - 1)) \times (3(2n - 1))\) has three types of \(3 \times 3\) sub-patches, according to their location with respect to the boundary of \( P \):

\[
\begin{array}{ccccccc}
I & II & II & II & I \\
II & III & III & III & II \\
II & III & III & III & II \\
II & III & III & III & II \\
I & II & II & II & I \\
\end{array}
\]

so we compute the Hessian for the center of each type as follows:

- For sub-patches of type I, \( 4HessP(c) \) is given according to its location by

\[
\begin{pmatrix}
c_{2,0} - 3c + 2c_{-1,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,-2} - 3c + 2c_{0,1}
\end{pmatrix}
\]

\begin{align*}
\text{top left} & : \\
\begin{pmatrix}
c_{2,0} - 3c + 2c_{-1,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,2} - 3c + 2c_{0,-1}
\end{pmatrix} & : \text{bottom left}
\end{align*}
CHAPTER 4. PROJECTING SAMPLES ONTO $K$

For sub-patches of type II, $4HessP(c)$ is given according to its location by

$$HessP(c) = \frac{1}{4} \begin{bmatrix}
2c_{1,0} - 3c + c_{-2,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,-2} - 3c + 2c_{0,1}
\end{bmatrix}$$

- top right

$$\begin{bmatrix}
2c_{1,0} - 3c + c_{-2,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,2} - 3c + 2c_{0,-1}
\end{bmatrix}$$

- bottom right

- For sub-patches of type III

$$HessP(c) = \frac{1}{4} \begin{bmatrix}
c_{2,0} - 2c + c_{-2,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,-2} - 3c + 2c_{0,1}
\end{bmatrix}$$

- top

$$\begin{bmatrix}
c_{2,0} - 2c + c_{-2,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,2} - 3c + 2c_{0,-1}
\end{bmatrix}$$

- bottom

$$\begin{bmatrix}
c_{2,0} - 3c + 2c_{-1,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,2} - 2c + c_{0,-2}
\end{bmatrix}$$

- left

$$\begin{bmatrix}
c_{-2,0} - 3c + 2c_{1,0} & c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} \\
c_{1,1} - c_{1,-1} - c_{-1,1} + c_{-1,-1} & c_{0,2} - 2c + c_{0,-2}
\end{bmatrix}$$

- right

Equipped with the discretization for the gradient, we can know compute $A_P$, $dir(P)$, $Dir(P)$ whenever $dir(P) \geq t_k$ and $\Psi(P)$ whenever $rad(P) > 0$. Here $t_k$ ($k = 1, \ldots, 5$) is a threshold to be determined.
Let $I$ be a digital image in gray scale. We describe now the process for constructing families of samples on the Klein bottle, from the sets of high-contrast log-intensity $k \times k$ patches of $I$.

**Constructing the Family of Samples**

1. For each $k = 3(2n - 1)$ ($n = 1, \ldots, 5$), let $\mathcal{P}_k$ be the set of $k \times k$ patches from $I$.

2. According to its bit depth, a patch $P \in \mathcal{P}_k$ is an integer matrix of size $k \times k$, with entries between 0 and $2^d - 1$. Here $d$ is the bit depth, which is usually equal to 8 or 16. Let $\tilde{P}$ be the patch obtained from $P$ by adding 1 to each entry and computing its logarithm. Let $\tilde{\tilde{P}}$ be the patch obtained from $\tilde{P}$ by subtracting its mean from each entry and denote the resulting set by $\tilde{\tilde{\mathcal{P}}}_k$.

3. For each $k$ and each $\tilde{\tilde{P}} \in \tilde{\mathcal{P}}_k$, compute its $D$-norm $\| \tilde{\tilde{P}} \|_D = \sqrt{\sum_{i \sim j} (p_i - p_j)^2}$

   where $\tilde{\tilde{P}} = [p_{rs}]$, $p_{k(s-1)+r} = p_{rs}$ and $i \sim j$ if and only if $p_i$ is in the 4-connected neighborhood of $p_j$. Keep the patch $\tilde{\tilde{P}}$ if its $D$-norm falls within the top 30% with respect to $\tilde{\mathcal{P}}_k$.

4. For each $k = 3(2n - 1)$ ($n = 1, \ldots, 5$) let

   $$S_n(I) = \{ \Psi(\tilde{\tilde{P}}) \mid \tilde{\tilde{P}} \in \tilde{\mathcal{P}}_k, \ dir(\tilde{\tilde{P}}) \geq t_k \text{ and } rad(\tilde{\tilde{P}}) > 0 \}.$$ 

   It follows that $S_n(I) \subseteq K$ is a sample on the Klein bottle.

We have now, modulo estimating the $t_k$’s, a complete method for associating to each digital image $I$ in gray scale, a sequence $S_i(I)$ of samples on the Klein bottle $K$. We will study the estimation problem in what remains of this section.
4.2.2 Estimating $t_k$

As discussed earlier, $dis(\cdot)$ is a measure of how directional a patch is and $Dir(P)$ computes this preferred direction whenever $dir(P) > 0$. Nevertheless, it is expected that the determined and perceived directions disagree for small values of $dir$, and therefore the role of $t_k$ will be to assure their agreement, whenever $dir \geq t_k$. In other words, evaluating the condition $dir(P) \geq t_k$ amounts to deciding whether $P$ is close enough to $\mathcal{K}$ so that applying the projection $\Phi : \mathcal{T} \rightarrow \mathcal{K}$ makes sense.

The estimation of the thresholds $t_k$ was carried out using the USC-SIPI Texture database. We describe now this set of image textures. The USC-SIPI Image Database is a collection of digitized images, maintained by the Signal and Image Processing Institute (SIPI) at the University of Southern California [46]. The Textures volume contains 154 images, all monochrome of sizes $512 \times 512$ and $1024 \times 1024$. Most of the images in this set are digitized versions of pictures in the Brodatz book [41].

![Figure 4.1: Exemplary images from the Brodatz collection, digitized in the USC-SIPI Texture database.](image)

Using the USC database, we now study how the perceived direction is related to $Dir(P)$, as $dir(P)$ increases. To this end, we construct the sets $\tilde{P}_k$ (refer to step 2 in Constructing the Family of Samples) for all images in the USC database and record for each patch its direction in directionally. The results for $k = 3, 9$ are summarized in figure 4.2.
Figure 4.2: Perceived directionality and \( \text{Dir} \) as a functions of \( \text{dir} \). \( \text{dir} \) increases to the right and down. \( \text{Dir} \) is measured in degrees.
The threshold $t_k$ is then determined (using the full set of $k \times k$ patches from the USC Database) as the value of $dir$, in which the perceived and computed directions begin to agree in a consistent fashion. In table 4.1 we summarize the estimated thresholds.

<table>
<thead>
<tr>
<th>$k$</th>
<th>3</th>
<th>9</th>
<th>15</th>
<th>21</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_k$</td>
<td>0.030</td>
<td>0.013</td>
<td>0.045</td>
<td>0.043</td>
<td>0.070</td>
</tr>
</tbody>
</table>

To have an idea of the percentage of patches we discard when thresholding by $t_k$, we present in figure 4.3 the histograms of directionalities for $k = 3, 9$.

(a) Directionality $3 \times 3$ patches  
(b) Directionality $9 \times 9$ patches  

Figure 4.3: Histogram of directionalities, USC Texture Database
In summary, we have developed a method for associating to each image $I$, a sequence of samples $S_i(I)$ on the Klein bottle $K$ with underlying probability density functions

$$h_i(I): K \rightarrow \mathbb{R}.$$ 

One of our final goals is to study what attributes of $I$ are captured by the sequence $\{h_i(I)\}$, and compare these sequences for different images $I$. In order to do so, it is necessary to device compact, discrete and faithful representations of estimators for $h_i$, as well as ways of comparing these representations. This problem will be the main focuss of the next chapter, where ideas from Fourier analysis and density estimation will be combined.
Chapter 5

A Discrete Representation

Density estimation from finite samples lies at the center of statistics as one of the most fundamental questions. In fact, many applications of statistical methods reduce to the estimation of some parameters, or the full probability density function (PDF) underlying the phenomenon. It is for this reason that several methods, following various heuristics, have been devised to solve different instances of the density estimation problem. See for example [8] for a nice survey.

One idea that stands out from this constellation of strategies, is that estimating a PDF can be translated into estimating an at most countable set of scalars, which is extremely appealing for finite representation purposes. If the PDF is, for instance, assumed to be in a particular parametric family, then one estimates the set of parameters. In this direction, the most general approach is perhaps the one given by orthogonal series estimators. Let $H$ be a separable Hilbert space and let $\{\phi_k\}_{k \in \mathbb{Z}}$ be an orthonormal basis for $H$. Under the assumption that $f \in H$, one can write

$$f = \sum_k f_k \phi_k$$

where convergence of the series is in the topology of $H$, and the estimation of $f$ can be translated into the estimation of the $f_k$’s. Notice that a finite number of the $f_k$’s might be enough for representation purposes, depending on the rate of convergence of the series.
A popular choice for $H$ is the space of square-integrable functions on the unit circle, where one can take $\phi_k(x) = e^{ikx}$. In what follows we will take $H = L^2(M)$, where $(M, \mathcal{A}, \mu)$ is a measure space, $\mathcal{A}$ countably generated and $\mu \sigma$-finite, to assure the separability of $L^2(M)$.

Natural questions arise in this setting:

- Are there “good” estimators $\hat{f}_k$ for $f_k$?
- How good is the statistic
  $$\hat{f} = \sum_k \hat{f}_k \phi_k.$$  
- How does the accuracy depend on the sample size?

Fortunately, the literature in statistics gives encouraging answers to these questions, with inspiration coming from the case $M = \mathbb{R}^n$.

### 5.1 Coefficient Estimation in $L^2$

Let $\mathcal{S}(\mathbb{R}^n)$ be the space of rapidly decreasing complex valued functions on $\mathbb{R}^n$ (see [36]). It follows that for each $c \in \mathbb{R}^n$, the Dirac delta function centered at $c$

$$\delta(x - c) : \mathcal{S}(\mathbb{R}^n) \longrightarrow \mathbb{C}$$

$$\varphi \mapsto \varphi(c)$$

is a continuous linear functional and therefore a tempered distribution. That is, an element in the dual space $\mathcal{S}'(\mathbb{R}^n) = \mathcal{S}(\mathbb{R}^n)^*$. An application of integration by parts shows that:

**Proposition 5.1.1.** For every $\varphi \in \mathcal{S}(\mathbb{R}^n)$ one has that

$$\lim_{k \to \infty} \int_{\mathbb{R}^n} \varphi(x) \frac{k}{\sqrt{\pi^n}} e^{-k^2\|x-c\|^2} d\mu(x) = \varphi(c).$$

Thus $\delta(x - c)$ is a weak limit of Gaussians with mean $c$, as the variance approaches zero.
CHAPTER 5. A DISCRETE REPRESENTATION

40

What this result tells us, is that $\delta(x - c)$ can be interpreted as the generalized function bounding unit volume, that is zero at $x \neq c$ and an infinite spike at $c$. Moreover, if $X_1, \ldots, X_N$ are i.i.d. random variables with probability density function

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

where $f \in L^2(\mathbb{R}^n)$, then as a “generalized statistic”

$$\hat{f}_\delta(X) = \frac{1}{N} \sum_{i=1}^{N} \delta(X - X_i)$$

can be thought of as the Gaussian Kernel estimator (see [8]) with infinitely small width.

Let $\{\phi_k\}$ be an orthonormal basis for $L^2(\mathbb{R}^n)$. What we will see shortly, is that an appropriate definition of the coefficients of $\hat{f}_\delta$ with respect to $\{\phi_k\}$, yields good estimators for those of $f$. To motivate the definition, let us consider the following example: Let $\phi_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by

$$\phi_\alpha(x_1, \ldots, x_n) = h_{\alpha_1}(x_1) \cdots h_{\alpha_n}(x_n)$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ and $h_k$ is the $k$-th Hermite function (or harmonic oscillator wave function) on $\mathbb{R}$. It follows that $\{\phi_\alpha\}$ is an orthonormal basis for $L^2(\mathbb{R}^n)$ (Lemma 3, page 142, [36]) and from the $N$-representation theorem (V.14 , [36]) we get that

$$\sum_\alpha T(\overline{\phi_\alpha})\phi_\alpha$$

converges to $T$ in the weak-* topology, for all $T \in \mathcal{S}'(\mathbb{R}^n)$. Here we are using $\mathcal{S}(\mathbb{R}^n) \subseteq L^2(\mathbb{R}^n) \subseteq \mathcal{S}'(\mathbb{R}^n)$, and $T(\overline{\phi_\alpha}) = T(\phi_\alpha)$.

From this example, it is natural to make the following definition.

**Definition 5.1.2.** Let $\{\phi_k\}$ be an orthonormal basis for $L^2(\mathbb{R}^n)$, and let $T \in \mathcal{S}'(\mathbb{R}^n)$. We define $T_k = T(\overline{\phi_k})$ as the coefficients of $T$ with respect to $\{\phi_k\}$. 
Proposition 5.1.3. Let \((M, \mathcal{A}, \mu)\) be as before and let \(\{\phi_k\}\) be an orthonormal basis for \(L^2(M)\). If \(X_1, \ldots, X_N\) are i.i.d. random variables with PDF \(f : M \to \mathbb{R}\), then the coefficients
\[
\hat{f}_k = \frac{1}{N} \sum_{i=1}^{N} \overline{\phi_k(X_i)}
\]
of the generalized statistic \(\hat{f}_\delta\) with respect to \(\{\phi_k\}\), are unbiased estimators for the coefficients \(f_k\) of \(f\). Moreover, \(\hat{f}_k\) converges almost surely to \(f_k\) as \(N \to \infty\).

Proof. Indeed, for each \(k \in \mathbb{Z}\)
\[
\mathbb{E}[\hat{f}_k] = \frac{1}{N} \int_{M} f \cdot \sum_{i=1}^{N} \overline{\phi_k} d\mu = \int_{M} f \cdot \overline{\phi_k} d\mu = f_k.
\]
The convergence result follows from the theorem of Large Numbers.

The good news, there are reasonably good estimators for the coefficients of \(f\) in very general settings. The bad news, since they are basis dependent, is that it is hard to identify just how much information from \(f\) is actually encoded in this representation. Let us try to make this idea more transparent. The first thing to note is that the sequence of estimators
\[
S_K(\hat{f}) = \sum_{|k| \leq K} \hat{f}_k \phi_k
\]
does not converge to \(f\) but to \(\hat{f}_\delta\) (in the weak-* topology) regardless of the sample size, and provided
\[
\sum \psi_k \phi_k(x) = \psi(x)
\]
for all \(\psi \in \mathcal{S}(\mathbb{R}^n), \ x \in \mathbb{R}^n\) (This is true for bases consisting of Hermite functions and trigonometric polynomials). More specifically, we are approximating a linear combination of delta functions, as opposed to the function \(f\). Thus, even when the \(\hat{f}_k\) are almost surely correct for large \(N\), the convergence in probability is not enough to make the \(\hat{f}_k\) decrease fast enough as \(k\) gets larger.
The conclusion: taking more coefficients is not necessarily a good thing, so instead we could try to make an educated guess about \( \{ \phi_k \} \) that improves our chances in terms of representation accuracy.

Some observations about the type of probability density functions modeling the distribution of small patches coming from natural scenes, are pertinent:

1. Natural images often come as compositions of various scenes. That is, small patches coming from different locations in the image are likely to have different directions and different gradients. Also, since the camera averages intensities locally, most transitions from linear to quadratic gradients in the same direction, vary continuously. This implies that any random sample of high density patches tends to cover large connected portions of the Klein bottle, and therefore the associated PDF is not likely to be localized in space. This observation suggests that the basis \( \{ \phi_k \} \) should be of global character, eliminating candidates such as the ones coming from wavelets.

2. An aspect inherent to the geometric model, is that elements in \( L^2(K) \) will have as domain a quotient of \( S^1 \times S^1 \). Thus by pre-composing with the quotient map, they can be regarded as \( 2\pi \)-periodic functions. Appropriate linear combinations of the set of functions

\[
\phi_{nm}(x, y) = e^{inx+imy}
\]

emerges then as a natural candidate.

Since the higher degree Fourier coefficients tune the fine scale details of the approximation, it follows that most of the information about the PDF is encoded in the low frequencies. That is, truncating the series is sensible in a heuristic sense.

It is known ([22]) that truncation is also nearly optimal with respect to mean integrated square error (MISE), within a natural class of statistics for \( f \). Indeed, if within the family of estimators

\[
f^*_N = \sum_{k=1}^{\infty} \lambda_k(N) \hat{f}_k \phi_k
\]
one looks for the sequence \( \{\lambda_k(N)\}_{k \in \mathbb{N}} \) that minimizes
\[
\text{MISE} = \mathbb{E}[\|f_N^* - f\|^2_2]
\]
then provided the \( f_k \) \((k \leq K(N))\) are large compared with \( n^{-1}\text{var}(\phi_k) \) and the \( f_k \) \((k > K(N))\) are negligible, then \( \lambda_k(N) = 1 \) for \( k \leq K(N) \) and zero otherwise, has a nearly optimal mean square error. Notice that this is the expected behavior for trigonometric bases. Refer to [37] for another inclusion criterion using only the derived estimates \( \hat{f}_k \), and the sample size \( N \).

In summary, not only is it possible to encode the information from \( f \) in a finite set of easily computable estimators, but this truncation is nearly optimal in a precise sense. We will proceed now to find a version of the Fourier basis for the case in which \( M \) is the Klein bottle.

### 5.2 An Orthonormal Basis for \( L^2(K) \)

Let \( S^1 \) be the unit circle, and let \( M(S^1, \mathbb{C}) \) be the set of complex valued (Lebesgue) measurable functions on \( S^1 \). We identify \( M(S^1, \mathbb{C}) \) with the set of functions \( f : \mathbb{R} \rightarrow \mathbb{C} \) which are measurable and so that \( f(x + 2\pi) = f(x) \) for all \( x \in \mathbb{R} \). If \( f, g \in M(S^1, \mathbb{C}) \), we define
\[
\langle f, g \rangle_{S^1} = \frac{1}{2\pi} \int_{S^1} f \overline{g} d\tau
\]
where \( \tau \) is the Lebesgue measure on \([0, 2\pi] \). Let \( L^2(S^1) \) be the set of equivalence classes \([f]\) (with respect to equality almost everywhere) of functions \( f \in M(S^1, \mathbb{C}) \) so that \( \langle f, f \rangle_{S^1} < \infty \). By abuse of notation we will write \( f \) instead of \([f]\).
Theorem 5.2.1. \( \langle \cdot, \cdot \rangle_{S^1} \) defines an inner product on \( L^2(S^1) \) so that \( (L^2(S^1), \langle \cdot, \cdot \rangle_{S^1}) \) is a Hilbert space. Moreover, the set of functions \( \mathcal{B} = \{ e^{i\tau n} \mid n \in \mathbb{Z} \} \) is an orthonormal basis for \( L^2(S^1) \). That is, if \( \hat{f}(n) = \langle f, e^{i\tau n} \rangle_{S^1} \) then

\[
S_N(f)(\tau) = \sum_{n=-N}^{N} \hat{f}(n)e^{i\tau n}
\]

converges to \( f \) in the \( L^2 \)-norm

\[
\|f\|_{S^1}^2 = \langle f, f \rangle_{S^1}.
\]

We will show now that a similar result holds for the Klein bottle. Let \( K \) be as in definition 4.1.4, and let

\[
M(K, \mathbb{C}) = \{ f : S^1 \times S^1 \to \mathbb{C} \mid f \text{ is measurable and } f(x+\pi,-y) = f(x,y) \}
\]

be the set of complex valued measurable functions on \( K \). Let

\[
\langle f, g \rangle_K = \frac{1}{(2\pi)^2} \int_K \overline{f} \overline{g} \, d\kappa
\]

for \( f, g \in M(K, \mathbb{C}) \), \( \kappa \) being the Lebesgue measure on \([0,\pi] \times [0,2\pi]\), and define

\[
L^2(K) = \{ \overline{f} \mid f \in M(K, \mathbb{C}) \text{ and } \langle f, f \rangle_K < \infty \}.
\]

Theorem 5.2.2. Let \( n \in \mathbb{Z} \), \( m \in \mathbb{N} \cup \{0\} \) and let

\[
\phi_{n,m}(x,y) = e^{inx+imy} + (-1)^n e^{inx-imy}.
\]

Then \( (L^2(K), \langle \cdot, \cdot \rangle_K) \) is a Hilbert space and \( \mathcal{B} = \{ \phi_{n,m} \mid m = 0 \text{ implies } n = 2k \} \) is an orthogonal basis for \( L^2(K) \) with respect to \( \langle \cdot, \cdot \rangle_K \). That is, if

\[
\|f\|_K^2 = \langle f, f \rangle_K
\]
is the induced $L^2$-norm and $\tilde{f}(n, m) \cdot \|\phi_{n,m}\|_K^2 = \langle f, \phi_{n,m} \rangle_K$, then for every $f \in L^2(K)$

$$\lim_{N \to \infty} \left\| \sum_{m=0}^{N} \sum_{n=-N}^{N} \tilde{f}(n, m) \phi_{n,m} - f \right\|_K = 0.$$ 

Proof. We defer the proof of this result to appendix A. □

In the case of real valued functions we have the following

**Corollary 5.2.3.** Let $f : K \rightarrow \mathbb{R}$ be a square integrable function. Then

$$\sum_{m=0}^{N} \sum_{n=-N}^{N} \tilde{f}(n, m) \phi_{n,m} = \frac{1}{2\pi^2} + \sum_{m=1}^{N} a_m(2 \cos my) + \sum_{n=1}^{N_0} b_n(2 \cos 2nx) + c_n(2 \sin 2nx)$$

$$+ \sum_{n,m=1}^{N} d_{nm}\left(2\sqrt{2} \cos(nx) \cdot \sin\left(my + \frac{\pi}{4}(1 + (-1)^n)\right)\right)$$

$$+ \sum_{n,m=1}^{N} e_{nm}\left(2\sqrt{2} \sin(nx) \cdot \sin\left(my + \frac{\pi}{4}(1 + (-1)^n)\right)\right)$$

converges to $f$ with respect to the norm $\| \cdot \|_K$ as $N \to \infty$. Here $N_0 = \lfloor \frac{N}{2} \rfloor$. The functions between parentheses on the right hand side of the equality are orthonormal with respect to $\langle \cdot, \cdot \rangle_K$, and thus the coefficients $a_m, b_n, c_n, d_{dm}, e_{nm}$ can be computed through inner products with $f$. That is,

$$a_m = \langle f, 2 \cos(my) \rangle_K, \quad b_n = \langle f, 2 \cos(2nx) \rangle_K, \quad \text{etc.}$$

We will refer to this set of functions as the trigonometric basis on the Klein bottle.

### 5.3 The $v$-Invariant

Let $I$ be an image in gray scale and for each $k = 3(2i - 1)$, let $S_i(I) \subseteq K$ be the projection onto the Klein bottle of (a sample of) the high-contrast log-intensity $k \times k$ patches from $I$. If $h_i(I) : K \rightarrow \mathbb{R}$ is the underlying PDF, then under the assumption of square-integrability, it can be written as an infinite series in terms of
the trigonometric basis on the Klein bottle as

\[ h_i(I) = \frac{1}{2\pi^2} + \sum_{m=1}^{\infty} a_m (2 \cos my) + \sum_{n=1}^{\infty} b_n (2 \cos 2nx) + c_n (2 \sin 2nx) \]

\[ + \sum_{n,m=1}^{\infty} d_{nm} \left( 2\sqrt{2} \cos(nx) \cdot \sin(my + \frac{\pi}{4} (1 + (-1)^n)) \right) \]

\[ + \sum_{n,m=1}^{\infty} e_{nm} \left( 2\sqrt{2} \sin(nx) \cdot \sin(my + \frac{\pi}{4} (1 + (-1)^n)) \right). \]

If we write \( S_i(I) = \{P^1_k, \ldots, P^N_k\} \) where \( P^r_k = (X^r_k, Y^r_k) \), then for each \( n, m \in \mathbb{N} \) we have the unbiased estimates (by proposition 5.1.3 and corollary 5.2.3)

\[ \hat{a}_m = \frac{1}{2N\pi^2} \sum_{r=1}^{N} \cos(mY^r_k) \]

\[ \hat{b}_n = \frac{1}{2N\pi^2} \sum_{r=1}^{N_0} \cos(2nX^r_k) \]

\[ \hat{c}_n = \frac{1}{2N\pi^2} \sum_{r=1}^{N_0} \sin(2nX^r_k) \]

\[ \hat{d}_{nm} = \frac{1}{\sqrt{2}N\pi^2} \sum_{r=1}^{N} \cos(nX^r_k) \cdot \sin(mY^r_k + \frac{\pi}{4} (1 + (-1)^n)) \]

\[ \hat{e}_{nm} = \frac{1}{\sqrt{2}N\pi^2} \sum_{r=1}^{N} \sin(nX^r_k) \cdot \sin(mY^r_k + \frac{\pi}{4} (1 + (-1)^n)) \]

for the coefficients of \( h_i(I) \) with respect to the trigonometric basis on the Klein bottle.

From the frequencies of the various sines and cosines involved, we can assign to each estimated coefficient a degree

\[ \text{deg}(\hat{a}_m) = m \]

\[ \text{deg}(\hat{b}_n) = \text{deg}(\hat{c}_n) = 2n \]

\[ \text{deg}(\hat{d}_{nm}) = \text{deg}(\hat{e}_{nm}) = n + m \]
and endow the multi-set

\[ C_i(I) = \{a_m, b_n, c_n, \hat{d}_{nm}, \hat{e}_{nm} \mid n, m \in \mathbb{N} \} \]

with a partial order \( \preceq \) as follows: Let \( u, v \in C_i(I) \) be distinct (not necessarily as values, but as elements of the multi-set),

1. If \( \text{deg}(u) < \text{deg}(v) \) then \( u \prec v \).

2. If \( \text{deg}(u) = \text{deg}(v) \) and \( u \) is less than \( v \) with respect to the grammatical order then \( u \prec v \). The grammatical order on \( C_i(I) \) is given by the letters without their sub-indices (e.g. \( \hat{a}_m \prec \hat{b}_n \), for all \( m, n \)).

3. If \( \text{deg}(u) = \text{deg}(v) \) and \( u \approx v \) with respect to the grammatical order, (e.g. \( \hat{d}_{12} \) and \( \hat{d}_{21} \)) then \( u \prec v \) if \( u \) is less than \( v \) with respect to the natural order \( (n, m) \mapsto (n + m - 2)(n + m - 1)/2 + n \).

**Definition 5.3.1.** Let \( I \) be an image in gray scale and let \( i \in \mathbb{N} \). Given a maximum degree \( D \in \mathbb{N} \), the \( v \)-Invariant at scale \( k = 3(2i - 1) \) bounded by \( D \) is the vector

\[ v_i(I) = (v_{i1}, \ldots, v_{id}) \]

where \( \{v_{i1}, \ldots, v_{id}\} = \{v \in C_i(I) \mid \text{deg}(v) \leq D\} \) and the indexing is induced by \( \preceq \).

We define the (multi-scale) \( v \)-Invariant \( v(I) \) as the matrix whose \( i \)-th row is \( v_i(I) \).

The \( v \)-Invariant is the construction we propose in this thesis, as a way of capturing the distribution of high-contrast small patches from an image. In what follows we will study visualization methods, dissimilarity measures for discrimination purposes, and its behavior under image rotation.
In order to visualize the information contained in \( v_i(I) \), we will use the heat-map associated to the estimated probability density function

\[
\hat{h}_i(I)(x, y) = \frac{1}{2\pi^2} + \sum_{n=1}^{D} \hat{a}_m(2 \cos my) + \sum_{n=1}^{D_0} \hat{b}_n(2 \cos 2nx) + \hat{c}_n(2 \sin 2nx)
\]

\[
+ \sum_{n+m=1}^{D} \hat{d}_{nm}(2\sqrt{2} \cos(nx) \cdot \sin(my + \frac{\pi}{4}(1 + (-1)^n)))
\]

\[
+ \sum_{n+m=1}^{D} \hat{e}_{nm}(2\sqrt{2} \sin(nx) \cdot \sin(my + \frac{\pi}{4}(1 + (-1)^n))).
\]

where \( D_0 = \lfloor \frac{D}{2} \rfloor \). That is, the color distribution on the fundamental domain \([\pi/4, 5\pi/4] \times [-\pi/2, 3\pi/2]\), where the color value at \((x, y)\) is given by \( \hat{h}_i(I)(x, y) \).
5.4 Preliminary Examples

Let us consider the image texture of Straw (D15 in [41]) obtained from scattered dry stalks

and let \( S \) be the corresponding digital image in gray scale. Following the construction described in definition 5.3.1 and taking \( D = 6 \) (and therefore \( d = 42 \)), we obtain the \( v \)-Invariant at scale 3

![v-Invariant](image)

Figure 5.1: Texture of Straw.

Given the \( v \)-Invariant, we proceed to compute the heat map for the estimated probability density function \( \hat{h}_1(S) \):
CHAPTER 5. A DISCRETE REPRESENTATION

Figure 5.3: (a) Projected Sample of $3 \times 3$ patches from Straw. (b) Estimated heat map for $D = 6, d = 42$.

The representation power of the v-Invariant is evident. With a total degree as small as 6, one already has a heat-map that accurately describes the distribution of $S_1(S)$. It is worth comparing the heat-map with the Klein bottle model.

In order to understand what patches appear in $S_1(S)$, and their distribution on $K$. 
From this comparison we can draw the following conclusions:

1. The hotter spots on the heat-map (modes of $\hat{h}_1(S)$) correspond roughly to linear patches with direction $\frac{3\pi}{4}$. Thus, the horizontal position of hot spots in the heat map encode observable direction of the image texture. On the other hand, the vertical position determines the size of the features in the image. That is, features coming from quadratic patches are regarded as smaller than the ones coming from linear ones.

2. The extent to which the heat map is confined to a vertical strip and the width of the band, reflect the directionality of the texture at the given scale. Indeed, if the image is purely directional, then so are its high-contrast patches, they will all have roughly the same direction and therefore the heat-map of the estimated PDF will be confined to a narrow vertical strip. The converse is not true in general; that is, there are textures with heat maps (at some scales) confined to a narrow vertical strip even when the images are not directional. The idea is that small purely directional patches, all with the same direction, can be arranged in patterns distorting the perceived global directionality. One expects nevertheless, that if the *vertical strip condition* holds across all scales, then the image will be purely directional.

3. There is a dominance in linear patches with an specific gradient orientation. The orientation of the gradient is dominant in the southeast direction and comes from the intensity of the heat at $y = \frac{\pi}{2}$. By zooming into the image, we see that this observation is consistent with the shadowing of the stalks and their location against the black background. It follows that the v-Invariant is sensitive to changes in light source.

4. The diffusion in the heat-map contains information regarding the roughness of the material. Indeed, diffusion in the heat can be interpreted as local small variations of direction and feature size of the high-contrast patches. This is exactly the effect created by irregular shadowing throughout the texture, which in turn is caused by lack of smoothness in the material.
5. The *vertical alignment* of the modes of $\hat{h}_1(S)$ is also a measure of the dominant direction in the texture. In rough materials, where the diffusion on the heat-map tends to be large, the vertical strip condition may not hold. Nevertheless, the vertical alignment of hot spots is an evidence of a dominant direction in the image.

6. The *transition in gradient orientation* between linear patches, is only through quadratic patches depicting bright configurations. Since this is certainly not an artifact of the representation, it would be interesting to understand what this observation translates into, when one looks at the actual image.

7. Another observation that requires a deeper understanding, is the strong *mirror symmetry* with respect to $x = \frac{3\pi}{4}$ and $y = \frac{\pi}{2}$.

A crucial aspect about these observation is that they indeed refer to properties of image textures captured by the heat-map, as opposed to mere artifacts of the construction. In what follows we provide examples of how all these properties vary as a function of the input texture. Let us take **ensemble 1** to be the collection:
The corresponding Heat-Maps are:

In which each row shows the heat-maps for a given image texture at all scales (size of patches). Notice the asymmetry in Grass 3, as well as the different types of heat diffusion, locations of modes, vertical alignments and gradient orientations across all examples.

We present in ensemble 1, two families of image textures from the USC and PMTex databases (see chapter 7). The ensemble contains: three versions of the texture produced by grass in the top row, and three images in the bottom row corresponding to three instances of the same material. The chosen material is a textured wall, and the different instances presented are rotations of light-source and the material itself. The goal with this collection is to show that even within the same texture class (Grass, USC Texture database) and several instances of the same material (Wall, PMTex database), the v-Invariant is powerful enough to make a clear distinction between
them. One of the main objectives, certainly, is to quantify how this distinctions are made. We discuss some ideas in this regard on chapter 6.

We now turn our attention to ensemble 2:

in which we present a wide range of textures; going from naturally occurring (Grass) to man-made materials (Fabric), regular (Fabric 3) to irregular patterns (Fabric 4), with a multi-scale hierarchy (Fabric 3) and all levels of competing directions (Plastic bubbles, Pebbles, Metal Plates). The main goal with this ensemble, is to use the diversity in the collection to show how the observations about the heat map for $\hat{h}_1(S)$, vary in terms of the input image.

Next we compute the Heat-Maps at various scales, for all images in ensemble 2:
We end this section with the following observations:

**Vertical Strip Condition** Refers to the extent to which the heat is confined to a vertical strip, and how narrow this band is. As we discussed earlier, this condition is related to the directionality of the image. As one can see, Straw is the texture with a most directional trend at all scales, supporting the fact that, out of all images in ensemble 2, its heat maps are the most accurately described by the vertical strip condition. Fabric 4 is an example where the vertical strip condition holds at some scales, but the image is not directional. In this case, the weave is diagonal but the colors in the fibers create an irregular pattern. The heat-maps in this case capture the directionality of the weave.

**Vertical Alignment of Modes** It is interesting to compare the heat maps of Fabric 1 and Metal Plates. The observation here is that as the scale increases, we get a better alignment of the distribution modes. Indeed, the heat maps for Fabric 1 show that the texture can be regarded as directional only at large scales, where the weave can be disregarded. The interesting part about the Metal Plates, is that the alignment in modes is due to shadowing and is only present at large scales, in which the conflicting vertical and horizontal direction can be ignored.

**Diffusion of the Heat** Is associated to roughness of the texture, and can be seen in Fabric 1, Bark, Pebbles or Corn.

**Gradient Orientation** It is interesting how this feature (location of modes in the $y$ axis) depends on shadowing in the image texture, due to roughness in the material and light source. Compare for example the heat maps (at small scale) of the metal plates, where the gradient orientation is consistent with the shadowing. In this case, light source location can be deduced from surface information (front versus back). If, by contrast, one considers pebbles and corn then even when there is a clear gradient orientation, this does not give away the location of the light source. Thus, gradient orientation of modes depends on both light source and surface properties, in a non-straightforward manner. Nevertheless, it shows that the v-Invariant is sensitive to these changes.
Feature Size is related to the role of quadratic patches in the distribution. In this case, the more small scale features we encounter in the image, the more quadratic patches will be present in the heat maps. Consider for example Pebbles and Fabric 3 versus Straw, Grass and Fabric 4.
Chapter 6

Dissimilarity Measures

With the v-Invariant at hand, not only do we have a way of representing the distribution on the Klein bottle of high-contrast patches, but we also have natural ways of assigning distances (i.e. using the $L^2$ norm) between such representations.

6.1 Naive Examples

When humans compare or classify a set of textures according to their similarities, we tend to answer questions regarding differences in directionality, regular versus irregular or stochastic, concordance at some scale, etc. This observation implies that given a way of representing images, one should device different notions of similarity between the representations, depending on which features one wishes to compare. With this heuristic in mind, we propose the following notions of distance:

**Best free uni-scale matching** Since images often exhibit similarities when compared at (potentially) different scales than the ones at which they are first presented, the best free uni-scale matching distance attempts to capture this similarity disregarding as much as possible the initial states. Given images $I, J$, this distance will be denoted as $d_1(I, J)$ and computed via the expression:

$$d_1(I, J) = \min_{1 \leq i, j \leq 5} \| \hat{h}_i(I) - \hat{h}_j(J) \|_{L^2}.$$
Best natural uni-scale matching Measures similarity at the same scale, for the scale in which the two distribution are the most similar. This distance will be denoted as $d_2(I, J)$ and computed via the expression:

$$d_2(I, J) = \min_{1 \leq i \leq 5} \left\| \hat{h}_i(I) - \hat{h}_i(J) \right\|_{L^2}.$$

Best free multi-scale matching Is denoted as $d_3(I, J)$ and computed via:

$$d_3(I, J) = \min_{\sigma \in S_5} \sum_{i=1}^{5} \left\| \hat{h}_i(I) - \hat{h}_{\sigma(i)}(J) \right\|_{L^2}.$$

Here $S_5$ denotes the symmetric group in 5 elements.

Natural multi-scale matching Tracks all scales at once, in the order given by the initial states. An expression for this distance is given by

$$d_4(I, J) = \sum_{i=1}^{5} \left\| \hat{h}_i(I) - \hat{h}_i(J) \right\|_{L^2}.$$

Since the $L^2$-norm for a real valued function $f : K \rightarrow \mathbb{R}$ is equal to the $\ell^2$-norm of its coefficients with respect to the trigonometric basis, we get the equality

$$\left\| \hat{h}_i(I) - \hat{h}_j(J) \right\|_{L^2} = \| v_i(I) - v_j(J) \|_{\mathbb{R}^d}.$$

Here $\| \cdot \|_{\mathbb{R}^d}$ is the Euclidean norm in $\mathbb{R}^d$.

One of the most important features of the multi-scale approach presented here, is that it allows one to measure how the distribution of high-contrast patches changes as a function of the patch size. An alternative for quantifying such change, is by looking at the evolution of the heat-map and measure how it changes from one stage to the next. It is in this spirit that we propose the evolution function $\varepsilon$. 


Definition 6.1.1. Let $I$ be an image with estimated distributions $\hat{h}_1(I), \ldots, \hat{h}_5(I)$ on $K$. Let
\[
\varepsilon(I) = \sqrt{\sum_{i=1}^{4} \| \hat{h}_i(I) - \hat{h}_{i+1}(I) \|^2_{L^2}}
\]
be its evolution function. Using $\varepsilon$, we define the evolution pseudo-distance
\[
d_\varepsilon(I, J) = |\varepsilon(I) - \varepsilon(J)|.
\]

The intuition is that the value of $\varepsilon$ will be large for images with a rich multi-scale structure (i.e. Fabric 1). The applications we have in mind for this function is to serve as a filter for (hierarchical, partial) clustering purposes.

We will devote the second part of this section to analyzing the aforementioned dissimilarity measures for ensemble 2. Computing the pairwise Natural-multi scale matching for this collection of images, we get the dissimilarity matrix:

![Natural Multi-Scale Matching](image)

Figure 6.1: Natural multi-scale distance matrix for ensemble 2.
Here, distances decrease (resp. increase) when color goes toward blue (resp. red). This ensemble serves as a great example of how easy it is for humans to rank similarity between textures, and then how hard it is to actually identify the criteria involved in the decision making process. If one takes for example Fabric 3, Metal Front and Metal Back, then it is clear that the first two are closer together. This is accurately captured by the v-Invariant and the $d_4$ distance. From the matrix we also get that Bubbles and Grass are the closest, the second closest pair would be Corn and Pebbles, while Fabric 1 and Metal Back are the two farthest apart. One would be tempted to say that Fabric 3 and Metal Front are the closest, but this is only true at large scales, when one ignores the weave pattern in the fabric.

Using complete linkage, in which two clusters are as similar as their most dissimilar points, we compute the dendrogram associated to the previous distance matrix:

![Dendrogram](image)

Figure 6.2: Classes emerging from $d_4$ distance on Ensemble 2.

A natural question that arises is how to use ensemble-specific information and the v-Invariant, to achieve classifications say by material or natural versus synthetic textures. We hint now to how variations across scales can be used to this end.
We begin by computing the evolution function $\varepsilon$ for all images in ensemble 2:

![Figure 6.3: Evolution function $\varepsilon(I)$ for images in Ensemble 2.](image)

$\varepsilon$ provides a natural partition into classes for ensemble 2. Notice that, with the exception of Plastic Bubbles, all man-made patterns rank above naturally occurring textures when measured by $\varepsilon$.

The concluding remark from this section, is that the v-Invariant is a simple and compact representation for images, that allows one to easily compare texture attributes. The dissimilarity measures presented in this section, can be thought of as the naive distances coming for free with the definition of the v-Invariant.

How far can one take this representation when looking for purpose-specific dissimilarity measures? In the next section we treat this question when the goal is to analyze surface and image rotation.

## 6.2 Image Rotation

A crucial aspect toward exploiting the full power of an invariant associated to images, is understanding its behavior under geometric transformations such as translation, rotation and scaling of the input.

We will show in this section that image rotation has a particularly simple effect on the distribution of patches on the Klein bottle, as well as a straightforward interpretation in the spectral domain of the v-Invariant. With this interpretation at hand, we present a mathematical and computational framework for estimating dissimilarities between images in a rotation invariant a fashion.
More specifically, we introduce the R-distance $d_R$ as a rotation invariant version of $d_4$, and through the computational model we show how the rotation factor can be recovered. We end this section with an $SO(2, \mathbb{R})$-equivariant version of the v-Invariant.

### 6.2.1 Behavior of the v-Invariant under the $SO(2, \mathbb{R})$ action.

**Proposition 6.2.1.** Let $T \in SO(2, \mathbb{R})$ be counter-clockwise rotation by an angle $\tau \in [-\pi, \pi)$ and let $h \in L^2(K, \mathbb{R})$. If the coefficients of $h$ and $h^\tau(x, y) = h(x - \tau, y)$ with respect to the trigonometric basis on $K$ are

$$a_1, \ldots, a_m, b_n, c_n, d_{nm}, e_{nm}, \ldots \quad \text{and} \quad a_1^\tau, \ldots, a_m^\tau, b_n^\tau, c_n^\tau, d_{nm}^\tau, e_{nm}^\tau, \ldots$$

then

$$a_m^\tau = a_m, \quad \begin{bmatrix} b_n \\ c_n \end{bmatrix}^\tau = T^{2n} \begin{bmatrix} b_n \\ c_n \end{bmatrix}, \quad \begin{bmatrix} d_{nm} \\ e_{nm} \end{bmatrix}^\tau = T^n \begin{bmatrix} d_{nm} \\ e_{nm} \end{bmatrix}$$

for all $n, m \in \mathbb{N}$.

**Proof.** We present the proof for $\tau \geq 0$, the one for $\tau < 0$ being similar. It follows from 5.2.3 that

$$a_m^\tau = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\pi h(x - \tau, y) \cos(my) dxdy$$

$$= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_{-\tau}^{\pi-\tau} h(x, y) \cos(my) dxdy$$

$$= \frac{1}{(2\pi)^2} \left( \int_0^{2\pi} \int_0^\pi h(x, y) \cos(my) dxdy - \int_0^{2\pi} \int_{\pi-\tau}^\pi h(x, y) \cos(my) dxdy \right.$$

$$\left. + \int_0^{2\pi} \int_{-\tau}^0 h(x, y) \cos(my) dxdy \right).$$

If for the second term in the last equation we perform the change of coordinates $(x, y) \mapsto (x - \pi, 2\pi - y)$, then using $h(x - \pi, 2\pi - y) = h(x, y)$ we conclude that $a_m^\tau = a_m$. 
Following a similar argument we see that

\[
b_n^\tau = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\pi h(x-\tau, y) \cos(2n) dxy
\]

\[
= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_{-\tau}^{\pi-\tau} h(x, y) \left( \cos(2nx) \cos(2n\tau) + \sin(2nx) \sin(2n\tau) \right) dxdy
\]

\[
= b_n \cos(2n\tau) + c_n \sin(2n\tau)
\]

\[
c_n^\tau = -b_n \sin(2n\tau) + c_n \cos(2n\tau)
\]

The result follows now by writing these equations using the matrix of \(T\) with respect to the standard basis.

The result follows now by writing these equations using the matrix of \(T\) with respect to the standard basis.

Using the linearity of the expected value, one obtains.

**Corollary 6.2.2.** Under the hypotheses of the previous proposition, one has that

\[
\hat{a}_m^\tau = \hat{a}_m, \quad \begin{bmatrix} \hat{b}_n^\tau \\ \hat{c}_n^\tau \end{bmatrix} = T^{2n} \begin{bmatrix} \hat{b}_n \\ \hat{c}_n \end{bmatrix}, \quad \begin{bmatrix} \hat{d}_{nm}^\tau \\ \hat{e}_{nm}^\tau \end{bmatrix} = T^n \begin{bmatrix} \hat{d}_{nm} \\ \hat{e}_{nm} \end{bmatrix}
\]  

(6.1)

yield unbiased estimators for the coefficients of \(h^\tau\) with respect to the trigonometric basis on \(K\).

Let us now see how these results apply to image rotation. From the Klein bottle model (Fig 3.6), it follows that the rotation (with respect to its center) of a continuous approximation \(I\) for an image \(I\), has the effect of translating horizontally the probability density function associated to its small patches. That is \(h(I^\tau) = h^\tau(I)\), where \(I^\tau\) is the counter-clockwise rotation of \(I\) by an angle \(\tau\).

In the discrete case, the local averages taken to form the pixels of the image have a non-trivial effect on the resulting rotation. Nonetheless, if \(S = \{P_1, \ldots, P_N\} \subseteq K\) is a sample associated to \(I\), where \(P_i = (X_i, Y_i)\), then \(S^\tau = \{P_1^\tau, \ldots, P_N^\tau\}\) where \(P_i^\tau = (X_i + \tau, Y_i)\), can be regarded a sample drawn from \(h(I^\tau)\) and thus can be used for the coefficient estimation stage in our pipeline. That is, \(h(I^\tau) = h^\tau(I)\).
The coefficients computed using this sample coincide with those in equation 6.1, and thus it makes sense to take them as the \( v \)-Invariant at the corresponding scale, for \( I^\tau \). If follows from proposition 6.2.1, that given images \( I, J \) and \( \tau \in [-\pi, \pi) \) one can compute \( \phi(I, J, \tau) = d_4(I^\tau, J)^2 \) in terms of \( v(I) \), \( v(J) \) and \( \tau \) alone. Let

\[
v_k(I) = (a_{k,1}, \ldots, a_{k,m}, b_{k,n}, c_{k,n}, d_{k,nm}, e_{k,nm}, \ldots, e_{k,RS})
\]

\[
v_k(J) = (a'_{k,1}, \ldots, a'_{k,m}, b'_{k,n}, c'_{k,n}, d'_{k,nm}, e'_{k,nm}, \ldots, e'_{k,RS})
\]

then

\[
\phi(I, J, \tau) := d_4(I^\tau, J)^2 
\]

\[
= \sum_{k=1}^{5} \|v_k(I^\tau) - v_k(J)\|^2 
\]

\[
= \|v(I)\|^2 + \|v(J)\|^2 - 2A(I, J) - 2\varphi(I, J, \tau) 
\]

where

\[
\|v(I)\|^2 = \sum_{k=1}^{5} \|v_k(I)\|^2 
\]

\[
A(I, J) = \sum_{k=1}^{5} \sum_{m=1}^{D} a_{k,m}a'_{k,m} 
\]

\[
\varphi(I, J, \tau) = \sum_{k=1}^{5} \left( \sum_{n=1}^{D_0} \left< T^{2n} \begin{bmatrix} b_{k,n} \\ c_{k,n} \end{bmatrix}, \begin{bmatrix} b'_{k,n} \\ c'_{k,n} \end{bmatrix} \right> + \sum_{n+m=1}^{D} \left< T^{n} \begin{bmatrix} d_{k,nm} \\ e_{k,nm} \end{bmatrix}, \begin{bmatrix} d'_{k,nm} \\ e'_{k,nm} \end{bmatrix} \right> \right) 
\]

\[
= \sum_{n=1}^{D} \alpha_n \cos(n\tau) + \beta_n \sin(n\tau). 
\]

Combining these formulae and proposition 6.2.1 we get
Proposition 6.2.3. For all images $I, J$ and all $\tau, \kappa, \sigma \in [-\pi, \pi]$ we have:

1. $\|v(I^\tau)\| = \|v(I)\|$. 

2. $A(I^\tau, J^\kappa) = A(I, J)$. 

3. $\varphi(I^\tau, J^\kappa, \sigma) = \varphi(I, J, \sigma + \tau - \kappa)$. 

Definition 6.2.4. Let $I, J$ be images in gray scale. We define the R-distance $d_R$ as 

$$d_R(I, J) = \min_{\tau \in [-\pi, \pi]} d_4(I^\tau, J). \quad (6.4)$$

Theorem 6.2.5. The R-distance $d_R$, defines a metric on the set of conjugacy classes of images under the $SO(2, \mathbb{R})$ action.

Proof. Symmetry of $d_R$ follows directly from the formulae in proposition 6.2.3, so let us show that the triangular inequality holds. To this end, let $I, J, K$ be images and let $\tau, \kappa \in [-\pi, \pi]$ be so that 

$$d_R(I, J) = d_4(I^\tau, J), \quad d_R(J, K) = d_4(J^\kappa, K).$$

Using the formulae from proposition 6.2.3 and the fact that $d_4$ satisfies the triangular inequality, we conclude that 

$$d_R(I, J) + d_R(J, K) = d_4(I^\tau, J) + d_4(J^\kappa, K)$$

$$= d_4(I^{\tau + \kappa}, J^\kappa) + d_4(J^\kappa, K)$$

$$\geq d_4(I^{\tau + \kappa}, K)$$

$$\geq d_R(I, K).$$
6.2.2 Computing the R-distance

In order for \( d_R \) to be useful in real world image comparisons, it is imperative that we device a suitable computational scheme. The first thing to notice is that

\[
\tau^* = \arg\max_{\tau \in [-\pi, \pi)} \varphi(I, J, \tau)
\]  

(6.5)

if and only if \( d_4(I^{\tau^*}, J) = d_R(I, J) \). Writing the objective function from 6.5 in terms of \( \tau \) as

\[
\varphi(I, J, \tau) = \sum_{n=1}^{D} \alpha_n \cos(n\tau) + \beta_n \sin(n\tau)
\]

where \( \alpha_n, \beta_n \) depend only on \( v(I) \) and \( v(J) \) (see equation 6.3), it follows that \( \tau^* \) satisfies the equation

\[
\sum_{n=1}^{D} n\beta_n \cos(n\tau^*) - n\alpha_n \sin(n\tau^*) = 0.
\]

This way, if

\[
q(z) = \sum_{n=1}^{D} (n\beta_n - i\alpha_n)z^n \quad \text{and} \quad \tilde{q}(z) = \sum_{n=1}^{D} (n\beta_n + i\alpha_n)z^n
\]

then \( z^* = e^{i\tau^*} \) is a root of the degree \( 2D \) complex polynomial

\[
p(z) = z^D \cdot (q(z) + \tilde{q}(1/z)).
\]

The previous analysis can be summarized as follows:

**Lemma 6.2.6.** Let

\[
p(z) = \sum_{n=1}^{D} n(\beta_n - i\alpha_n)z^{D+n} + n(\beta_n + i\alpha_n)z^{D-n}
\]

where the \( \alpha_n, \beta_n \) are given by 6.3. If \( \tau^* \) is a minimizer for 6.4 and \( z^* = e^{i\tau^*} \) then \( p(z^*) = 0 \). Therefore, computing \( d_R(I, J) \) reduces to minimizing \( \phi(\tau) = \phi(I, J, \tau) \).
(refer to equation 6.2) over the set of arguments from (unitary) roots of \( p(z) \).

The roots of \( p(z) \) can in principle be computed as the eigenvalues of its companion matrix. This is how the MATLAB routine \texttt{roots} is implemented. It is known ([10]) that \texttt{roots} produces the exact eigenvalues of a matrix within roundoff error of the companion matrix of \( p \). This, however, does not mean that they are the exact roots of a polynomial with coefficients within roundoff error of those of \( p \).

Let \( \hat{p} \) be the monic polynomial having \texttt{roots}(\( p \)) as its exact roots. It follows that \( \hat{p} \) is close to \( p \) in a precise sense [1]. Indeed, let \( \epsilon \) be the machine precision and let \( E = [E_{ij}] \) be a dense perturbation matrix, where the \( E_{ij} \) are small multiples of \( \epsilon \). If \( A \) is the companion matrix of

\[
p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1z + a_0
\]

and \( \hat{p}(z) \) is the characteristic polynomial of \( A + E \), then to first order, the coefficient of \( z^{k-1} \) in \( \hat{p}(z) - p(z) \) is

\[
\sum_{m=0}^{k-1} a_m \sum_{i=k+1}^{n} E_{i,i+m-k} - \sum_{m=k}^{n} a_m \sum_{i=1}^{k} E_{i,i+m-k}
\]

where \( a_n \) is defined to be 1.

This result suggests in our case, that even when \texttt{roots}(\( p \)) does not return the exact roots of \( p \), but those of a slight perturbation \( \hat{p} \), their arguments

\[
\text{Arg}(\texttt{roots}(p)) \subseteq [-\pi, \pi)
\]

can be used as the initial guess in an iterative approximation, e.g. Newton’s method, looking to minimize \( \phi(\tau) \). Given that simple expressions for \( \phi'(\tau) \), \( \phi''(\tau) \) and good initial points are readily available, one expects a fast convergence in methods based on gradient descent.

Let

\[
a^* = \text{GradientDescent}(\phi, a)
\]

be found in a Newton-type method, looking to minimize \( \phi \) with \( a \) as the starting
point. For this step, we use the `fminunc` routine included in the Statistics MATLAB Toolbox, which has a fast convergence and high accuracy when provided with $\phi''$.

The computational scheme for calculating $d_R(I, J)$ can be summarized as follows:

Input  v-Invariants $v(I)$ and $v(J)$ for $I$ and $J$.

Output R-distance $d_R(I, J)$

minimizer $\tau^* = \arg\min_{\tau \in [-\pi, \pi)} d_4(I^\tau, J)$

1. Determine the polynomial $p(z)$ and the arguments
   
   \[ A = \text{Arg(roots}(p)) \]

2. for each $a \in A$ do
   
   \[ a^* = \text{GradientDescent}(\phi, a) \]

end

3. Find

   \[ \tau^* = \arg\min_{a^* \in A^*} \phi(a^*) \]
   
   where $A^* = \text{GradientDescent}(\phi, A)$.

Figure 6.4: Computational Scheme for the R-distance $d_R$

6.2.3 A rotation invariant version of $v(I)$

The version of $v(I)$ that we have in mind is an association

\[ I \mapsto \tilde{v}(I) \]

so that $\tilde{v}(I) = v(I^{\tau_0})$ for some $\tau_0 \in [-\pi, \pi)$ depending only on $v(I)$, and so that $\tilde{v}(I^{\tau}) = \tilde{v}(I)$ for all $\tau \in [-\pi, \pi)$. This implies that whenever $v(I)$ is rotation invariant, that is $v(I) = v(I^{\tau})$ for all $\tau \in [-\pi, \pi)$, $\tilde{v}$ is given by

\[ \tilde{v}(I) = v(I). \]
For the general case, let us assume that 

\[
v(I) = [v_{ij}] = (v_1(I); \cdots; v_5(I)) \]

where

\[
v_k(I) = (a_{k,1}, \ldots, a_{k,m}, b_{k,n}, c_{k,n}, d_{k,nm}, e_{k,nm}, \ldots, e_{k,RS})
\]

is not rotation invariant, so it follows that there exists a nonzero vector of the form

\[\begin{bmatrix} b_{k,n} \\ c_{k,n} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} d_{k,nm} \\ e_{k,nm} \end{bmatrix}.\]

The first entries of such vectors can be ordered with respect to their position in 

\[v(I) = [v_{ij}]\]

by means of 

\[(i, j) \mapsto (i + j - 1)(i + j - 2)/2 + i,\]

so let \(v_{i_0,j_0}\) be the smallest of such entries. Let \(u_0\) be the corresponding vector, write 

\[e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\]

and let 

\[T \in SO(2, \mathbb{R})\]

be the clockwise rotation given by

\[\frac{u_0}{\|u_0\|} = \begin{cases} 
T^{2n_0}(e_1) & \text{if } v_{i_0,j_0} = b_{k_0,n_0} \\
T^{n_0}(e_1) & \text{if } v_{i_0,j_0} = d_{k_0,n_0m_0}
\end{cases}\]

If \(\tau_0 \in [-\pi, \pi)\) is the angle associated to \(T\), then by defining \(\tilde{v}(I) = v(I^{\tau_0})\) we get that \(I \mapsto \tilde{v}(I)\) has the desired properties.

**Remarks:**

1. Let \(I\) be an image, \(\kappa \in [-\pi, \pi)\) and let \(J = I^\kappa\). An important aspect of the ideas presented in this section, is how simple it becomes to estimate \(\kappa\) (up to self-similarities of \(I\)) from \(v(I)\) and \(v(J)\). Indeed, if \(\tau_0(I) \in [-\pi, \pi)\) is so that 
\[v(I^{\tau_0(I)}) = \tilde{v}(I)\]

then
\[\kappa \equiv \tau_0(I) - \tau_0(J) \mod 2\pi.\]

In the case where \(J\) is not exactly the result of rotating \(I\), e.g. when \(I\) and \(J\) are images of the same object under different lightning conditions and viewpoint, the minimizer computed in figure 6.4 serves as a generalization of \(\kappa\).

In order to deal with the non-uniqueness of the minimizer, a natural choice is
to require $\tau^*$ to be the one with the smallest absolute value.

2. Recall the definition of $d_4(I, J)$ as the quantity

$$d_4(I, J) = \sqrt{\sum_{i=1}^{5} \|v_i(I) - v_i(J)\|^2}$$

where $v(I) = (v_1(I); \ldots; v_5(I))$. If we write $\tilde{v}(I) = (\tilde{v}_1(I); \ldots; \tilde{v}_5(I))$ then we can define a rotation invariant version of $d_4$ by the formula

$$\tilde{d}_4(I, J) = \sqrt{\sum_{i=1}^{5} \|\tilde{v}_i(I) - \tilde{v}_i(J)\|^2}$$

from which it follows that

$$\tilde{d}_4(I, J) = d_4(I^\tau(I), J^\tau(J))$$

$$= d_4(I^\tau(I) - \tau(J), J)$$

$$\geq \min_{\tau \in [-\pi, \pi]} d_4(I^\tau, J)$$

$$= d_R(I, J)$$

Even when $\tilde{d}_4$ is fairly simple to compute in comparison to $d_R$, the best classification results were obtained with the latter.

3. Recall the evolution function $\epsilon$ defined as

$$\epsilon(I) = \sqrt{\sum_{i=1}^{4} \|v_{i}(I) - v_{i+1}(I)\|^2}$$

It follows from this formula and proposition 6.2.1 that $\epsilon$ is rotation invariant.
Chapter 7

Classification Results

The Photometric 3D Texture Database (PMTex database) is a collection of gray scale images from real world textures, photographed by Jerry Wu [34] under controlled illumination conditions. The database contains a total of 36 sample materials (which we will refer as samples):

![Sample materials, PMTex database.](image)

The samples exhibit a wide range of physical surfaces: Directional, bidirectional and multidirectional, with small and large height variations, heavy shadowing, etc.
Each sample is then captured under different poses and illumination conditions:

![Figure 7.2: Instances of a sample in the PMTex database.](image)

The different poses of the sample correspond to counterclockwise rotations of the material, with respect to the texture plane. The resulting positions are of the form $\frac{i\pi}{6}$ where $i = 0, \ldots, 6$. The rotation of the light source, which is located with a slant of $\frac{\pi}{4}$ at a fixed distance from the center of the material, occurs in increments of $\pi/4$ or $\pi/2$. This yields a database of 1930 texture images, which we will analyze in this chapter.

### 7.1 Material Categories Via Hausdorff Distance

From an appearance-based viewpoint, categories of materials emerge naturally when textural properties of pairs of samples under varying lighting conditions and camera pose, are used as a measure of similarity. In particular, samples can be categorized by the way in which their textural information changes when seen from different angles.

As an example to fix ideas, woolen sweaters knitted in different patterns may present fairly different textural information, while the way in which the two textures change with varying lighting conditions is often very similar. We propose the Hausdorff distance between collections of sample instances, as a way of comparing this changes in textural information.

Let $S$ be a sample (e.g. a piece of wood, fabric, terrain, etc) for which images $S_1, \ldots, S_n$ taken under several lighting conditions and camera poses, are available.
CHAPTER 7. CLASSIFICATION RESULTS

We can associate to $S$ the metric space $(v(S), d_R)$, $v(S) = \{v(S_1), \ldots, v(S_n)\}$, which is a subspace of $(V, d_R)$, where $V = \{v(T) \mid T$ is a PMTEx image texture$\}$.

Below we show some of the distance matrices associated to samples in the PMTEx database.

![Image of distance matrices]

Figure 7.3: First and third row, instances of samples in the PMTEx database; rows two and four, corresponding distance matrix for the associated metric space.

All pairwise distances between instances of one sample in the PMTEx database are encoded in the corresponding matrix representation. Each one of such matrices is either $28 \times 28$ or $56 \times 56$, depending on the number of instances available. For each sample $S$ included in the PMTEx database, one can represent the different instances by $S_{ij}$ where $i = 0, \ldots, 6$ and $j = 0, \ldots, k$, $k = 3, 7$. That is, $S_{ij}$ is the texture image we get when the sample $S$ and the light source are rotated $\frac{i\pi}{6}$ and $\frac{2j\pi}{k+1}$ respectively. With this notation, $d_R(S_{ij}, S_{i'j'})$ corresponds to the $(i + j(k + 1), i' + j'(k + 1))$ entry in the matrix associated to $v(S)$. 
Remark: The regularity in the patterns observed in figure 7.3, serves as strong evidence of the robustness of our method to sampling effects when computing $v(S_{ij})$, and numerical errors when calculating $d_R$. The idea we explore in this section, is the extent to which these patterns serve as fingerprints for material samples.

Understanding how the scale of colors reflects distance between textures in the PMTex database, is perhaps easier when looking at the distance metric representing $(V,d_R)$. We include it below:

Figure 7.4: 1930-by-1930 distance matrix representing $(V,d_R)$. All matrices in figure 7.3 are present as diagonal blocks of this one.
Using the Hausdorff distance we define the $M$-distance $d_M$ ($M$ for material classification) as:

**Definition 7.1.1.** Given samples $S, S'$ with texture images $S_1, \ldots, S_n$ and $S'_1, \ldots, S'_m$ taken under distinct lighting conditions and camera poses, we let

$$d_M(S, S') = d_H(v(S), v(S')) = \max \left\{ \max_i \min_j d_R(S_i, S'_j), \max_j \min_i d_R(S_i, S'_j) \right\}.$$  

The material categories should emerge now as clusters, with distance measured by $d_M$. We use complete clustering, in which two clusters are as similar as their most dissimilar elements, and report the obtained classes.

The corresponding dendrogram follows.
CHAPTER 7. CLASSIFICATION RESULTS

7.2 Texture Classification From Known Instances

It is often the case that instances from families of textures are readily available for classification purposes. The categorization power of an invariant and its associated dissimilarity measures, can then be gauged through classification of novel instances. In this machine-learning framework, the first step is to randomly select a subset of instances for each sample in the PMTex database. Using the resulting training set, we categorize the remaining texture images via nearest neighbor for each one of the distances introduced in 6.2. We now describe the details:

1. First, we choose a percentage $0 < p \leq 100$, which will determine the size of the training set.

2. For each sample $S$ in the PMTex database we do the following: let $S_1, \ldots, S_n$ be the instances available for $S$ ($n$ will be either 28 or 56), and let $k = \left\lfloor \frac{pn}{100} \right\rfloor$. 

Figure 7.6: Complete-linkage dendrogram for $d_M$
Randomly choose an order preserving, injective function

\[ \sigma : \{1, \ldots, k\} \rightarrow \{1, \ldots, n\} \]

and let \( T_p(S) = \{S_{\sigma(1)}, \ldots, S_{\sigma(k)}\} \).

3. Let

\[ T_p = \bigcup_S T_p(S) \]

where the union is taken over all samples in the PMTex database. \( T_p \) will be regarded as a random training set for the percentage \( p \). The set of novel instances to be classified is now \( N_p = \bigcup_S \left( \{S_1, \ldots, S_n\} - T_p(S) \right) \).

4. For each \( S \in N_p \), let \( S' \in T_p \) be the closest element in \( T_p \) as measured by \( d \in \{d_4, d_R, \tilde{d}_4\} \). If \( S \) and \( S' \) are instances of the same sample, we will deem the classification as successful. The classification success of \( d \) in \((T_p, N_p)\) can now be computed as the percentage

\[ \frac{\text{successful classifications}}{|N_p|} \times 100. \]

We repeat this process on 1,000 random training sets, for different percentages \( p = 7, 14, 25, 50 \). The mean and standard deviation of classification success, are then computed for each distance. We compare in figure 7.7, the classification success of \( d_4 \), \( d_R \) and \( \tilde{d}_4 \) at \( p = 7, 14, 25, 50 \). Nodes in the plot correspond to the computed mean at the given percentage, and the error bars to the standard deviation.

To have an idea of the numbers, we show in table 7.1 the sizes of \( T_p \) and \( N_p \) as functions of \( p \).

<table>
<thead>
<tr>
<th>( p )</th>
<th>( 7 )</th>
<th>( 14 )</th>
<th>( 25 )</th>
<th>( 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>T_p</td>
<td>)</td>
<td>135</td>
<td>270</td>
</tr>
<tr>
<td>(</td>
<td>N_p</td>
<td>)</td>
<td>1,750</td>
<td>1,660</td>
</tr>
</tbody>
</table>

Table 7.1: \( |T_p| \) and \( |N_p| \) as functions of \( p \)
With 36 samples on the PMTex database and at least 28 instances per sample, the number of possible training sets for each value of $p$ is greater than $(\binom{28}{14})^{36} \geq 4 \times 10^{252}$. While the number of randomly generated training sets (per value of $p$) is only $10^3$, the small variance in computed success percentage, strengthens the validity of our estimation.

![Classification Success Plot](image)

Figure 7.7: Classification success of $d_4$, $d_R$ and $\tilde{d}_4$ on the PMTex database. We select $p\%$ instances per sample as training set, and classify novel textures using nearest neighbors. The error-bar plot shows the mean and standard deviation of the classification success, using 1,000 random training sets.

The classification success of $d_R$ is, as expected, greater than that of $d_4$ and $\tilde{d}_4$ on the PMTex database. This comparison makes the case for the importance of understanding the behavior of image invariants with respect to image transformations.
We have only scratched the surface with respect to actions of the general linear group $GL(3, \mathbb{R})$. Scaling can be taken into consideration with distances like $d_1$ (refer to section 6.1), while the study of shear mappings could shed light on the effects of 3D motion and perspective. Is it worth noticing how the mathematical framework for distributions of high contrast patches (thought of as functions on a meaningful space like the Klein bottle, and efficiently represented by Fourier-like coefficients), allows one not only to interpret, but to quantify the changes in the representation.

A success rate of almost 97% via nearest neighbor implies a little bit more than good classification power. It shows that the notion of similarity being used, is an adequate local approximation to the intrinsic distance. In this sense, it shows that the local behavior of $(V, d_R)$, can be used to infer global information of how the image of the v-Invariant (as a function from images to coefficients) sits inside $\mathbb{R}^{42}$. In the next section, we will present a construction pursuing this goal.

7.3 Local to Global Information: Intrinsic distance reconstruction and Class Partitioning

As discussed in the previous section, the local information carried by the v-Invariant and the $R$-distance, is well suited for texture classification and class partitioning. The main issue with distances which are estimated or defined following certain heuristics, is that they often fail at capturing more than the local behavior. It is exactly the same difficulty one runs into when trying to measure distances in an embedded manifold using the ambient metric, as opposed to the geodesic distance. If we look for example at figures 7.3 and 7.4, we will see that $d_R$ suffers of this drawback. On the one hand, figure 7.3 and the previous section suggest that it handles local distances accurately; while on the other hand, figure 7.4 clearly shows that $d_R$ by itself does not do a very good job at separating instances of the same sample when lighting conditions are varied. We propose in this section a local to global approach based on nearest neighbor graphs, in order to recover the intrinsic distance. See [28] for other applications of this idea.
Recall that a path from $a$ to $b$ in a directed graph $G$, is a sequence of vertices $g_0, \ldots, g_N$ so that $g_0 = a$, $g_N = b$ and $(g_{i-1}, g_i)$ is an edge in $G$ for all $i = 1, \ldots, N$. The length of such path is defined to be $N$. We say that there is a loop in $G$ containing the vertices $a, b$ if there are paths from $a$ to $b$ and from $b$ to $a$.

Notice that

“either $a = b$, or $a$ and $b$ are in a loop”

defines an equivalence relation on the set of vertices of $G$, and the equivalence classes are referred to as the strongly connected components of $G$. There is a natural distance in $G$ arising from this data. Indeed, if $a \neq b$ are in the same strongly connected component, we define $d_G(a, b)$ as one half of the length of the shortest loop containing them. If $a = b$ we let $d_G(a, b) = 0$ and if $a, b$ are not in a loop in $G$, we let $d_G(a, b) = \infty$.

Given a finite metric space $(X, d_X)$ and $k \in \mathbb{N}$, one can associate a directed graph $G_k(X) = G_k(X, d_X)$ with vertex set $X$, and where an edge is drawn from $x$ to $y$ if $y \neq x$ falls within the $k$-th nearest neighbors of $x$. The idea we exploit in this section, is that the strongly connected components of $G_k(V) = G_k(V, d_R)$ should recover the sets of instances of samples in the PMTex database. The extent to which this is true, can be measured using the separation of classes under the graph distance.

Let us denote by $d_k$ the distance associated to the directed graph $G_k(V)$, and replace $d_k$ by $\min(d_k, 20)$. These two metrics are equivalent in the sense that they induce the same topology, but the second is more suitable for computations and comparisons. We will denote the finite metric, by abuse of notation, also as $d_k$.

Given that each class within the PMTex database gets connected at a potentially different rate as we go from $G_k(V)$ to $G_{k+1}(V)$, then the choice of a particular value of $k$ is somewhat arbitrary and unmotivated. To see this, compare the distance matrices in figure 7.3. Another observation in this regard, is that it should be more meaningful to look at the evolution of $d_k$ as $k$ increases, to rule out “false” edges and give a better estimate of the global distance.

In order to overcome these difficulties, three ideas will guide our approach: First, we deem distances between elements as more relevant, if they do not change too much as the graph gets more connected. The second observation is that for each
class, there is a range of values of $k$ in which inner-class distances do not change too much. This, due to the fact that edges are being drawn namely between elements from the same class. The third and final observation is that, as suggested by figure 7.7, the stabilization occurs as the class approach full connectedness.

Based on these observations, we propose a weighted average of the distances $d_1, \ldots, d_N$ as an approximation to the intrinsic distance. We let $N = 20$ as this value is roughly half the size of all classes in the PMTex database. We define

$$d = \frac{1}{210} (d_1 + 2d_2 + 3d_3 + \cdots + 20d_{20})$$

and the distance matrix corresponding to $(V, d)$ is now

\[ \text{Figure 7.8: 1930-by-1930 distance matrix for } (V, d). \text{ Each diagonal block in dark blue represents a class in the PMTex database.} \]
By comparing this matrix to the one on figure 7.4, one can see how the intrinsic metric does a much better job at separating between classes in the PMTex database. Each diagonal block in dark blue, represents the distances between all instances of a given texture sample. The conclusion we can draw from this analysis, is that the strongly connected components from nearest neighbor graphs, were in this case successful at recovering the initial texture classes.
Chapter 8

Conclusions and Final Remarks

The application of Gabor filters for feature extraction, the maximum response 8 (MR8) filters for texture analysis, and the use of the Discrete Cosine Transformation (DCT) in image compression, are all closely related approaches. They all attempt to project small patches onto a finite number of more basic ones, and probe the result in order to extract meaningful information. This strategy is an old, and very successful one.

Projecting onto the Klein bottle follows this idea with two new features built in. First, the actual domain of filters (the Klein bottle) is continuous, which increases the fidelity of the filtered data. Notice that better accuracy in the feature extraction stage does not become an obstacle for compact representation, for two numbers are enough. Second, by taking into account the actual statistical and geometric structure of the space of optical patches, one gains insight into phenomena and applications that were hidden before having the model. Just as the description of distributions of high contrast patches from an image, and the effect of rotation became quite apparent having the model at our disposal, many more uses of this structure are yet to be discovered.

The idea of describing a patch as a point on a Klein bottle works for all images. In this paper we used an statistical approach to study images whose statistic properties capture most of the information available. It would be interesting to explore other ways of using this representation.
The results presented here are encouraging as seen from the texture analysis and classification viewpoint: On the one hand, the Fourier representation proved to be compact and powerful for discrimination purposes. On the other hand, it is important to notice that we only used (purely) linear and quadratic information from the patches we selected. It would be an interesting direction for future research to enlarge the Klein bottle model, in order to include patches which are frequently encountered in texture images, and contain higher order information.
Appendix A

Proof of theorem 5.2.2

Since $⟨·,·⟩_K$ is readily a positive definite Hermitian form, we will proceed to check the orthogonality relations, the convergence assertion and the completeness statement. Let $δ_{ij}$ be the Kroneker delta function. That is, $δ_{ij} = 1$ if $i = j$ and zero otherwise. Notice that for $n,n' \in \mathbb{Z}$ and $m,m' \in \mathbb{N} \cup \{0\}$ we have

$$(2π)^2 δ_{nn'} δ_{mm'} = (2π)^2 ⟨e^{inx}, e^{in'x}⟩_{S^1} · ⟨e^{imy}, e^{im'y}⟩_{S^1}$$

$$= \int_0^{2π} e^{i(n-n')x} dx \cdot \int_0^{2π} e^{i(m-m')y} dy$$

$$= \int_0^{2π} \left( \int_0^{2π} e^{inx+im'y} · e^{-in'x-im'y} dx \right) dy$$

$$= \int_0^{2π} \int_0^{2π} e^{inx+im'y} · e^{-in'x-im'y} dxdy + \int_{-2π}^{2π} \int_0^{2π} (-1)^n e^{inx-im'y} · (-1)^{n'} e^{-in'x+im'y} dxdy$$

$$= \int_0^{2π} \int_0^{2π} e^{i(n-n')x+i(m-m')y} + (-1)^{n-n'} e^{i(n-n')x-i(m-m')y} dxdy.$$
On the other hand,

\[
\int_0^{2\pi} \int_0^\pi \left( (-1)^{n'} e^{i(m+m')y} + (-1)^n e^{-i(m+m')y} \right) e^{i(n-n')x} \, dx \, dy
\]

\[
= \left( \pi \delta_{nn'} + (\delta_{nn'} - 1) \frac{1 - (-1)^{n-n'}}{(n-n')i} \right) \int_0^{2\pi} e^{i(m+m')y} (-1)^{n'} + e^{-i(m+m')y} (-1)^n \, dy
\]

\[
= \left( \pi \delta_{nn'} + (\delta_{nn'} - 1) \frac{1 - (-1)^{n-n'}}{(n-n')i} \right) 2\pi \delta_{mm'} \left( (-1)^{n'} + (-1)^n \right)
\]

\[
= (-1)^n (2\pi)^2 \delta_{nn'} \delta_{m0} \delta_{0m'}
\]

and adding the derived equalities we get

\[
\langle \phi_{nm}, \phi_{n'm'} \rangle_K = \delta_{nn'}(\delta_{mm'} + (-1)^n \delta_{m0} \delta_{0m'})
\]

which implies the orthogonality relations.

Let us now check the convergence assertion. Let \( q : S^1 \times S^1 \to K \) be the quotient map induced by the equivalence relation \((x, y) \sim (x', y')\) if and only if \(x - x' = \pi\) and \(y + y' = 0\). Then \( q \) induces a map

\[
q^* : M(K, \mathbb{C}) \to M(S^1 \times S^1, \mathbb{C})
\]

give by \( q^*(f) = f \circ q \), and it follows that for each \((y, f) \in S^1 \times L^2(K)\), we have a function \( f_y \in M(S^1, \mathbb{C}) \) given by \( f_y(x) = q^*(f)(x, y) \).
The first thing to notice is that if \( f, g \in L^2(K) \) then
\[
\langle f, g \rangle_K = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\pi f(x, y) \overline{g(x, y)} \, dx \, dy
\]
\[
= \frac{1}{2(2\pi)^2} \int_0^{2\pi} \int_0^\pi f(x, y) \overline{g(x, y)} \, dx \, dy + \frac{1}{2(2\pi)^2} \int_{-2\pi}^0 \int_0^{2\pi} f(x + \pi, -y) \overline{g(x + \pi, -y)} \, dx \, dy
\]
\[
= \frac{1}{2(2\pi)^2} \int_{S^1 \times S^1} q^*(f) \cdot \overline{q^*(g)} \, d(\tau \times \tau)
\]
so in particular
\[
\|f\|_K^2 = \frac{1}{4\pi} \int_0^{2\pi} \|f_y\|_{S^1}^2 \, dy < \infty
\]
which shows that \( f_y \in L^2(S^1) \) for almost every \( y \in S^1 \). This way, by theorem 5.2.1, we conclude that
\[
\lim_{N \to \infty} \left\| f_y - S_N(f_y) \right\|_{S^1} = 0 \quad \text{and} \quad \|f_y\|_{S^1}^2 = \sum |\widehat{f_y}(n)|^2
\]
for almost every \( y \) in \( S^1 \). Notice that
\[
S_N(f_y)(x + \pi) = \sum_{n=-N}^N \frac{1}{2\pi} \int_0^{2\pi} q^*(f)(\tau, -y) e^{-in\tau} \, d\tau \cdot (-1)^n e^{inx}
\]
\[
= \sum_{n=-N}^N \frac{1}{2\pi} \int_{-\pi}^{\pi} q^*(f)(\tau, y) e^{-in\tau} \, d\tau \cdot e^{inx}
\]
\[
= \sum_{n=-N}^N \widehat{f_y}(n) e^{inx}
\]
so as a function of two variables, \( S_N(f_y)(x) \in L^2(K) \).
An application of Young’s inequality with exponent 2 shows that

\[
\frac{1}{2} \left\| f - \sum_{m=0}^{N} \sum_{n=-N}^{N} \tilde{f}(n, m) \phi_{n,m} \right\|_K^2 \leq \left\| f - S_N(f_y) \right\|_K^2 + \left\| S_N(f_y) - \sum_{m=0}^{N} \sum_{n=-N}^{N} \tilde{f}(n, m) \phi_{n,m} \right\|_K^2 \tag{A.1}
\]

where

\[
\left\| f - S_N(f_y) \right\|_K^2 = \langle f - S_N(f_y), f - S_N(f_y) \rangle_K
\]

\[
= \frac{1}{2(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} |f_y(x) - S_N(f_y)(x)|^2 dxdy
\]

\[
= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \left\| f_y - S_N(f_y) \right\|_{S^1}^2 dy
\]

\[
\frac{1}{2} \left\| f_y - S_N(f_y) \right\|_{S^1}^2 \leq \left\| f_y \right\|_{S^1}^2 + \left\| S_N(f_y) \right\|_{S^1}^2
\]

\[
= \left\| f_y \right\|_{S^1}^2 + \sum_{n=-N}^{N} |\hat{f}_y(n)|^2 \leq 2\left\| f_y \right\|_{S^1}^2.
\]

Since the function \( y \mapsto \left\| f_y \right\|_{S^1}^2 \) is integrable, Lebesgue’s dominated convergence theorem implies that \( \lim_{N \to \infty} \left\| f - S_N(f_y) \right\|_K = 0 \).

To see that the second term in the right hand side of inequality A.1 also goes to zero as \( N \) goes to infinity, notice that

\[
\left\| \phi_{nm} \right\|_K^2 = 1 + (-1)^n \delta_{m0} \quad \text{and} \quad \tilde{f}(n, m) \left\| \phi_{nm} \right\|_K^2 = \langle f, \phi_{nm} \rangle_K
\]

imply that \( \tilde{f}(n, m) \) can be computed as

\[
\tilde{f}(n, m) = \left( 1 - \frac{1}{4} \delta_{m0}(3 - (-1)^n) \right) \langle f, \phi_{nm} \rangle_K.
\]
Thus, for each $n \in \mathbb{Z}$

$$\sum_{m=0}^{N} \tilde{f}(n, m) \phi_{nm} \cdot e^{-inx} = \frac{1 + (-1)^n}{2} \langle f, \phi_{n0} \rangle_K + \sum_{m=1}^{N} \langle f, \phi_{nm} \rangle_K \left( e^{imy} + (-1)^n e^{-imy} \right)$$

where

$$(2\pi)^2 \langle f, \phi_{nm} \rangle_K = \int_{0}^{2\pi} \int_{0}^{\pi} f(x, y)(e^{-inx-Imy} + e^{-in(x-\pi)+Imy})dxdy$$

$$= \int_{0}^{2\pi} \int_{0}^{\pi} f(x, y)e^{-inx-Imy}dxdy + \int_{-2\pi}^{0} \int_{\pi}^{2\pi} j^* f(x, y)e^{-inx-Imy}dxdy$$

$$= 2\pi \int_{0}^{2\pi} \hat{f}_y(n)e^{-imy}dy = (2\pi)^2 \hat{f}_y(n, m).$$

Since this can also be computed as

$$\langle f, \phi_{nm} \rangle_K = \frac{1}{2(2\pi)^2} \int_{0}^{2\pi} \int_{0}^{2\pi} j^* (f)(x, y) \left( e^{-inx-Imy} + (-1)^n e^{-inx+Imy} \right) dxdy$$

$$= \hat{f}_y(n, m) + (-1)^n \hat{f}_y(n, -m)$$

we conclude that

$$\sum_{m=0}^{N} \tilde{f}(n, m) \phi_{nm} \cdot e^{-inx} = \sum_{m=-N}^{N} \hat{f}_y(n, m)e^{imy} = S_N(\hat{f}_y(n)).$$
Now, from the previous calculation and the orthogonality relations in $L^2(S^1)$

$$
\|S_N(f_y) - \sum_{n=-N}^N \sum_{m=0}^N \tilde{f}(n, m) \phi_{nm}\|_K^2
= \frac{1}{2(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \left| \sum_{n=-N}^N \left( \hat{f}_y(n) - \sum_{m=0}^N \tilde{f}(n, m)(e^{iny} + (1)^n e^{-iny}) \right) e^{inx} \right|^2 \, dx \, dy
= \sum_{n=-N}^N \frac{1}{4\pi} \int_0^{2\pi} \left| \hat{f}_y(n) - S_N(\hat{f}_y(n)) \right|^2 \, dy
= \frac{1}{2} \sum_{n=-N}^N \left\| \hat{f}_y(n) - S_N(\hat{f}_y(n)) \right\|_{S^1}^2.
$$

Since for all $N, N' \in \mathbb{N}$

$$
\sum_{|n| \geq N} \frac{1}{4\pi} \int_0^{2\pi} \left| \hat{f}_y(n) - S_{N'}(\hat{f}_y(n)) \right|^2 \, dy
\leq \sum_{|n| \geq N} \frac{1}{2\pi} \int_0^{2\pi} |\hat{f}_y(n)|^2 \, dy + |S_{N'}(\hat{f}_y(n))|^2 \, dy
= \sum_{|n| \geq N} \frac{1}{2\pi} \int_0^{2\pi} |\hat{f}_y(n)|^2 \, dy + \sum_{|n| \geq N} \|S_{N'}(\hat{f}_y(n))\|_{S^1}^2
\leq \sum_{|n| \geq N} \frac{1}{2\pi} \int_0^{2\pi} |\hat{f}_y(n)|^2 \, dy + \sum_{|n| \geq N} \|\hat{f}_y(n)\|_{S^1}^2
= \frac{1}{\pi} \sum_{|n| \geq N} \int_0^{2\pi} |\hat{f}_y(n)|^2 \, dy
$$

and for almost every $y \in S^1$

$$
\lim_{N \to \infty} \sum_{|n| \leq N} |\hat{f}_y(n)|^2 = \|f_y\|_{S^1}^2
$$

then by Lebesgue’s dominated convergence theorem, given $\epsilon > 0$, we can pick $N_0 \in \mathbb{N}$
so that for all $N \in \mathbb{N}$

$$\sum_{|n| \geq N_0} \frac{1}{4\pi} \int_0^{2\pi} \left| \hat{f}_y(n) - S_N(\hat{f}_y(n)) \right|^2 dy < \epsilon.$$ 

This way

$$\lim_{N \to \infty} \sum_{n=-N}^{N} \frac{1}{4\pi} \int_0^{2\pi} \left| \hat{f}_y(n) - S_N(\hat{f}_y(n)) \right|^2 dy 
\leq \frac{1}{2} \sum_{|n| \leq N_0} \left\| \hat{f}_y(n) - S_N(\hat{f}_y(n)) \right\|_{S^1}^2 + \epsilon 
= \epsilon$$

and since this is true for all $\epsilon > 0$, letting it go to zero we get the result.
Bibliography


[34] Jerry Wu’s PMTex database, available at: http://www.taurusstudio.net/research/pmtexdb/index.htm


[46] USC-SIPI Texture database, available at:
   http://sipi.usc.edu/database/database.cgi?volume=textures


