Statistics of Natural Image Geometry

Kim Steenstrup Pedersen

CR Subject Classification: I.4, I.5, G.3

This thesis was set in LaTeX by the author.

Statistics of Natural Image Geometry

Kim Steenstrup Pedersen

This is a Ph.D. thesis given in defense of the Danish Ph.D. degree in Science at the Department of Computer Science, Faculty of Science, University of Copenhagen, Denmark. Supervisors have been Professor Dr. Phil. Peter Johansen, Department of Computer Science, University of Copenhagen and Professor Ph.D. Mads Nielsen, IT University of Copenhagen. The advisory board consisted of Professor Ph.D. Tony Lindeberg, Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Sweden and Associate Professor Ph.D. Ying Nian Wu, Department of Statistics, UCLA, USA.

Copenhagen, 2nd April 2003

Preface

I am grateful for the stipend I received in January 2000 from the faculty of science, University of Copenhagen. The three years of Ph.D. studies have been great fun and of course very educational.

I would like to thank my two supervisors Peter Johansen and Mads Nielsen for their advice, support, and encouragement. Especially, thanks to Peter for providing a positive working environment in the Image group at DIKU, and thanks to Mads for throwing some great parties.

I appreciate DIKU's policy on Ph.D. advisory boards and would like to thank the members of my advisory board, Tony Lindeberg and Ying Nian Wu, for their time and advice.

In the spring semester of 2001 I was so fortunate to be able to visit David Mumford at Brown University. Thank you David for opening up your lab for me, for putting Ann Lee and me together on the 3×3 patch project, and for the nice weekend in Maine. Ann Lee and I have been working together on a couple of projects since then and I would like to thank her for the great collaboration. I hope we can continue to work together in the future.

I also want to thank the rest of the image group at DIKU and at IT-C, especially Ole Fogh Olsen, Grumse (Martin) Lillholm, Erik Dam, and Jon Sporring, for being a group of good colleagues. Niels Holm Olsen — I am sorry that we did not finish the project we started in the summer of 2002. Furthermore, I must not forget the people I shared an office with: Henrik Dohlmann, Arthur Pece, Rasmus Kjeldsen, Andreas Thomsen, Erik, and Ole. I also appreciate the group of people I have had lunch with over the last three years (you know who you are).

At last I would like to express my gratitude to my wife Elisabeth for supporting me and for being who she is.

Kim Steenstrup Pedersen

Contents

List of Publications

This list of publications only includes peer reviewed papers and in the case of several versions of a work, only the latest version is included.

Papers Published in Scientific Journals:

- Ann B. Lee, Kim S. Pedersen and David Mumford: The Nonlinear Statistics of High-Contrast Patches in Natural Images. *In International Journal of Computer Vision*. To appear, 2002.

- Kim S. Pedersen and Mads Nielsen: The Hausdorff Dimension and Scale-Space Normalisation of Natural Images. *In Journal of Visual Communication and Image Representation*, 11(2):266-277, 2000.

Papers Published in Peer Reviewed Conference Proceedings:

- Kim S. Pedersen and Ann B. Lee: Toward a Full Probability Model of Edges in Natural Images. *In Proceedings of the 7th European Conference on Computer Vision, ECCV'02*, Copenhagen, Denmark. Part I, p. 328 – 342, LNCS 2350, Springer-Verlag, May, 2002.

- Ann B. Lee, Kim S. Pedersen and David Mumford: The Complex Statistics of High-Contrast Patches in Natural Images. *In WWW Proceedings of Second International IEEE Workshop on Statistical and Computational Theories of Vision*. Vancouver, Canada, July 2001. Available at http://pascal.stat.ucla.edu/~szhu/Workshops/SCTV2001.html.

- Kim S. Pedersen and Mads Nielsen: Computing Optic Flow by Scale-Space Integration of Normal Flow. *In Scale-Space and Morphology in Computer Vision: Proceedings of Scale-Space 2001*, p. 14 – 25, Lecture Notes in Computer Science LNCS 2106, Springer Verlag, July 2001.

Papers submitted to Conferences:

- Kim S. Pedersen: Properties of Brownian Image Models in Scale-Space. *Scale-Space 2003*. December, 2002.

Published Technical Reports:

- Kim S. Pedersen: Turbulence in Optical Flow Fields. *Technical Report DIKU-00-3*, DIKU, Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark. January 2001. Technical report version of masters thesis.

- Kim S. Pedersen: Normalisering af billedafledte i skalarum ved hjælp af den fraktale dimension. *Technical Report DIKU-99-2*, Department of Computer Science, University of Copenhagen, Universitetsparken 1, DK-2100 Copenhagen, Denmark, January 1999.

Chapter 1

Introduction

This thesis is divided into three separate major parts: Part I is on natural image statistics, and part II is on motion analysis. Each of the two parts represents an independent track of research in which I have been involved. The title of the thesis, *Statistics of Natural Image Geometry*, reflects the topic of the first part, since it comprises the major part of my research. The last part III is a summary of part I and II and a discussion of open problems.

In this introduction, I will put my research into a wider perspective. The structure of the chapter reflects the overall division of the thesis. The first section 1.1 discusses my research on natural image statistics and its possible applications. The second section 1.2 discuss the research I have done on motion analysis. The last section 1.3 describes the outline of the thesis.

1.1 Part I: Natural Image Statistics

1.1.1 Vision as Bayesian Inference

In computer vision and image analysis the task is to make inference from images about physical scenes. We want to develop algorithms that maps input, consisting of one or more images, into an output that is a description of the physical world the image depicts. A description can for instance be thought of as a parameterized model of the scene. Often such descriptions take the form of a hierarchy of descriptions at an increasing level of abstraction, where each description has a meta-description. From this hierarchy of descriptions, we want to deduce some information that lets us solve a certain task or problem. Examples of such problems could be that of making a robotic car that is able to drive autonomously in heavy traffic or specifying an algorithm that makes a preliminary medical diagnosis based on medical images with some modality, e.g. conventional x-ray or MR imaging.

This process of inference has been formalized by David Marr (Marr, 1982) who suggested in his seminal book on visual perception that vision should be thought of as an information processing task. He introduced three levels at which vision must be understood:

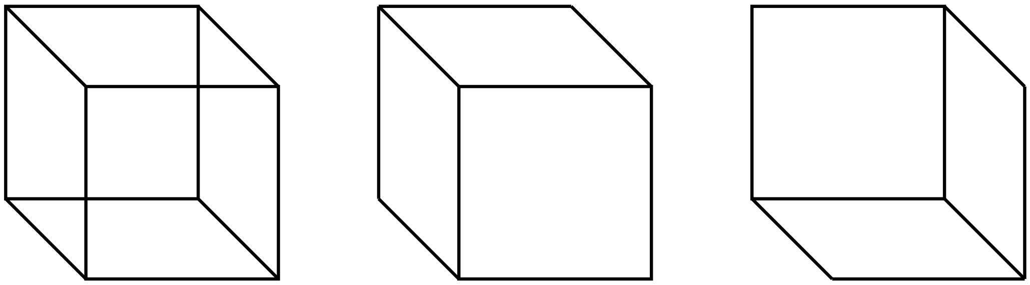
Computational Theory	$\longleftrightarrow$	Representations and Algorithms	$\longleftrightarrow$	Hardware implementation

The *computational theory* describes what is to be computed by the information processing system. The computational theory is an abstract description of the mapping between the input and whatever output we desire. The *representation and algorithm* level is the realization of the computational theory. We first have to choose representations for the input and output, and then construct an algorithm that transforms the input into the output according to the computational theory. The last level, *hardware implementation*, describes how to implement the representation and algorithm in hardware. In the case of computer vision, the hardware can be either a general purpose computer or a specialized machine.

If we build a vision system on a digital computer following these ideas, it is clear that the input is one or more images. An image is an ordered set of measurements. Often we represent images as arrays of numbers that quantizes the measurements. When we use conventional imaging technology such as optical cameras we measure, in a two dimensional array, the light intensity incident on the camera image plane. Other modalities exist such as laser range images and a myriad of medical image modalities, each measuring different properties of the physical world.

It is not obvious what output representation to use. For computer vision Marr suggested a hierarchy of output representations of increasing abstraction leading to a three dimensional description of the physical scene depicted by the input image. The levels of the hierarchy leading up to this 3-D description consist at the lowest level of what he calls the *primal sketch*, which is a representation of local geometric information about the two dimensional image intensity function. At this level he operates with the concept of features such as edges, blobs, and terminations. At the second level he introduces the so-called 2 $\frac{1}{2}$-D sketch, which represents orientation and depth of surfaces in the image. At the top level he has the full 3-D description of the scene.

Various researchers in computer vision have build upon and extended Marr's ideas. An example is Lindeberg's (1994b) scale-space primal sketch, which is based on blob-like image structures. In this thesis I am mainly going to discuss problems of low level vision, i.e. the problem of defining and detecting features, where the idea of a primal sketch is fruitful. In fact,



(a) (b) (c)

Figure 1.1: The Necker cube: An example of the human visual systems multiple hypotheses. Your vision system allows you to flick between the two different views, (b) and (c), of the line drawing of a three dimensional cube (a). Try it out.

the work presented in this thesis can be seen as a first attempt at building a probabilistic primal sketch. To make this statement clear I first have to introduce the following concepts.

The human visual system can have several competing hypotheses about a particular scene. Take for instance the examples in Fig. 1.1 and Fig. 1.2. In my opinion, it seems reasonable that we must have a computational theory that accounts for the concept of ambiguous and competing hypotheses. Bayesian inference[1] has been proposed as a general framework for a computational theory of vision by, among others, Mumford (1996) and Nielsen (1995). The Bayesian approach to inference allows us to handle problems with ambiguous and competing hypotheses.

In Bayesian inference one expresses the ambiguity of scene interpretations by assigning probabilities to each hypothesis or description. Given an image I we assign a conditional probability density $p(D|I)$ to descriptions[2] D. This conditional probability density is called the *a posteriori* (or posterior) probability density. We can make inference about the image I by for instance choosing the most probable description D of the image, i.e. the description D that maximizes the a posteriori probability $p(D|I)$ — the *maximum a posteriori (MAP)* description. There are also other possibilities such as choosing the posterior mean description (see for instance Nielsen (1995) for a discussion of selection criteria).

[1] See e.g. Jaynes (1996) or O'Hagan (1994) for a general discussion of Bayesian inference.

[2] For now, I will assume that both the space of descriptions and the space of images are continuous.

Figure 1.2: Another example of the human visual systems multiple hypotheses. Do you see the old or the young woman?

Bayes theorem tells us that we can obtain the posterior distribution $p(D|I)$ by observing the following relation: The joint probability $p(I, D)$ of the image I and the description D can be written as $p(I, D) = p(I|D)p(D)$ and vice versa $p(I, D) = p(D|I)p(I)$. Setting these two expressions equal to each other and rearranging the terms we get Bayes theorem

$$p(D|I) = \frac{p(I|D)p(D)}{p(I)} \ .$$

The term $p(I|D)$ is called the *likelihood* of the image I given the description D and the term $p(D)$ is known as the *a priori* (or prior) distribution of descriptions. The denominator $p(I)$ is the probability distribution of images, sometimes called the evidence, and it acts as a normalization factor. This can be expressed as marginalizing $p(I, D)$ over D or in a loose sense as integrating out D with respect to the measure dD over the space of descriptions,

$$p(I) = \int p(I|D)p(D)\, dD \ .$$

When the image I is given, the evidence $p(I)$ is of course nothing but a constant.

Determination of the likelihood $p(I|D)$ is problem specific, and it is into this term we introduce the constraints suitable for our particular problem.

The literature is filled with examples of likelihood terms for specific applications (Nielsen, 1995; Yuille and Coughlan, 2000; Sidenbladh and Black, 2001). For instance, Torr (2002) applied the Bayesian approach to model estimation and selection for epipolar geometry and Pece and Worrall (2002) applied the Bayesian approach to tracking combined with empirically based a priori knowledge of the image background statistics.

The prior distribution $p(D)$ is a summary of all the knowledge we have a priori about descriptions D. As we just saw, the distributions $p(I)$ and $p(D)$ are related through the integral equation $p(I) = \int p(I|D)p(D)\,dD$. This means that knowledge about $p(I)$ in principle can give us knowledge about $p(D)$ and vice versa. That is, by studying the distribution of images $p(I)$ with our desired descriptions D in mind, we might get an idea of how to choose the prior $p(D)$ on descriptions. The prior $p(D)$ can be viewed as a model of the background clutter in images as it has been proposed by, among others, Grenander and Srivastava (2001).

Jaynes (1968) has suggested a least committed approach for the construction of prior distributions through the maximum entropy principle[3]. Assume that we have n constraints that we want to apply to the space of possible descriptions, e.g. bounds on or expectation values of model parameters. Jaynes argues that the least committed distribution incorporating our prior information is constructed by choosing the probability distribution that fulfills the n constraints and at the same time maximizes the entropy. Jaynes also argues the importance of the prior to be invariant with respect to certain transformations. For a prior of natural images the natural choices are at least scale and translation invariance. One may also argue for the addition of rotational invariance. If we follow Jaynes' idea we should look for a priori constraints on the space of descriptions and then use the maximum entropy principle to get the prior distribution $p(D)$.

In my research on natural image statistics I focus on finding suitable a priori distributions for natural images. I propose to get this knowledge through empirical studies of the distribution $p(I)$ of natural images. It is important to know what are probable events in the space of natural images, since this can guide us in the choices of constraints to put on the a priori distribution of descriptions $p(D)$. When we know these constraints we can construct $p(D)$ by for instance following the procedure of Jaynes' as discussed above. To be more specific, I propose to study the local geometry of images and develop a probability model of this geometry. The most probable local geometric structures are likely candidates for important features or descriptions D. This can be thought of as a probabilistic model of Marr's primal

[3] see e.g. Jaynes (1957) for an introduction to this principle.

sketch — a probabilistic primal sketch. We will return to this endeavor in Sec. 1.1.4.

1.1.2 Natural Image Statistics: A Historical Overview

Researchers have approached the subject of statistics of natural images from different perspectives. From a neurological and psychophysical point of view, knowing the statistics of the world we perceive through our eyes might help us understand the inner workings of the human visual system, especially why the visual system has developed as it has (Barlow, 1961; Simoncelli and Olshausen, 2001). From the point of view of a computer scientist, knowing the statistics of natural images might be useful in the development of image analysis algorithms. Some examples of the use of natural image statistics in image analysis and computer vision are tracking (Pece and Worrall, 2002; Sidenbladh and Black, 2001), segmentation (Malik et al., 2001), image reconstruction (Nielsen and Lillholm, 2001; Lillholm et al., 2002), compression (Mallat, 1989; Buccigrossi and Simoncelli, 1999), contour completion (Ren and Malik, 2002), and image enhancement (Simoncelli, 1999a).

It seems obvious that the statistics of different classes of images will vary, or at least we cannot assume that it will not. It is therefore important to define the class of images under investigation and be careful not to make generalized statements about images of other classes from results on the investigated class.

What is a natural image then? I define natural images as the class of images of scenes that we might see on our every day walk through life. This includes every thing from images of natural areas such as a forest or mountain terrain to images of indoor scenes such as an office (see Fig. 1.3 for a couple of examples). I define the class of natural images very broadly because I want to be able to characterize the statistics of images in general, independent of specific environments. By doing this we might not capture the statistical variation in a specific set of images, e.g. images of cars, but we hopefully capture the general variation of for instance the background in images of cars.

Besides defining the class of images of interest, one also has the choice between studying ensembles of images or a single image. Some of the results reported below does not hold or varies if we study single images. In my opinion, it is most interesting to study the statistics of ensembles of images, since this would lead to a general understanding of the statistics of natural images.

What follows is a short overview of various results from the research field of natural image statistics that I find important. For an in-depth review of

Figure 1.3: Examples of natural images.

advances in natural image statistics the reader is encouraged to have a look at Srivastava et al. (2003).

In the 1950's studies of television signals (Kretzmer, 1952; Deriugin, 1956) showed that the power spectrum of TV signals follows a power law $S(\xi) = \frac{A}{|\xi|^\alpha}$, where $|\xi|$ is the magnitude of the spatial frequency ξ, A is some constant, and the exponent α is dependent on the class of images and varies around $\alpha = 2$ for broad classes of images. A power spectrum of the form $\frac{1}{|\xi|^2}$ implies scale invariance of the second order statistics of the signal, i.e. the signal energy and covariance structure. This result of scale invariance was rediscovered by Field (1987), this time on a small set of digital images of various natural scenes. Ruderman and Bialek (1994) showed that on an ensemble of images taken in a forest the power spectrum has an exponent around $\alpha = 2 - \eta$ where $\eta = 0.19$, implying self similarity[4] rather than scale invariance. Mumford and Gidas (2001) note that α varies slightly for different classes of images, which illustrates the importance of defining the class of images we are interested in. For image ensembles with a large variation $\alpha \approx 2$ is a good approximation. Other signs of scale invariant statistics of natural images have been reported. For instance Zhu and Mumford (1997) showed that the histogram of some filters applied to natural images are scale invariant. Alvarez et al. (1999c; 1999b) have studied the scaling of object size in natural images.

Another interesting property of natural images is the apparent non-Gaussianity of both marginal and joint statistics. Field (1987) noticed this in his power spectrum experiments. Mallat (1989) showed that wavelet coefficients of natural images have a distribution which can be modeled by the so-called generalized Laplacian distribution. Srivastava et al. (2002) have argued that the so-called Bessel K forms are also a good model of wavelet coefficients and filter responses in general. Both Simoncelli (1999b) and Huang and Mumford (1999) have studied the non-Gaussian joint statistics of pairs and triplets of wavelet coefficients.

If we take the Bayesian approach to vision as outlined in Sec. 1.1.1, we have to clarify what we mean by a probability distribution $p(I)$ on the space of images. Srivastava et al. (2003) identify two views in natural image statistics on the problem of defining probability models on images. The first one is to define probability distributions on the whole space of images and assigning a near zero probability to unlikely images. The space of images in this case can for instance be $\mathbb{R}^{N \times M}$ or a space of functions with some characteristics. The space of functions of so-called bounded variation (BV) has been proposed as a possible space of images, but Alvarez et al. (1999a) argue that only a

[4]See for instance Sec. 5.7 for a definition of self similarity.

small subset of natural images may belong to the space of BV functions and Gousseau and Morel (2001) argue that natural images can not in general be BV functions. It has been argued by Florack (1997) and Mumford and Gidas (2001) that images are not proper functions, but that they rather belong to the space of Schwartz distributions[5] also called generalized functions. The second view on natural image statistics and probability models on images is to study the manifold of natural images embedded in the full space of images and to assign probability distributions to this manifold. One uses the fact that the set of naturally occurring images is quite sparse in the space of images. Here we have the problem of defining the image manifold, or rather constraining what is a natural image and what is not. See discussion of various approaches in Srivastava et al. (2003).

Markov random fields are a common technique for defining probability distributions on images (see e.g. Winkler (1995)). But this method does in general only allow for the modeling of short range dependencies and not long range dependencies. This is the limitation of what has been called the causal MRF models and to some extend their strength, since it is more tractable to specify short range dependencies than long range. Unfortunately, long range dependencies are abundant in natural images. Zhu et al. (1997) have proposed a MRF based texture model which remedies this problem and is able to model large scale patterns. A new approach by Wu et al. (2002) is the so-called texture sketch, where a causal MRF model is used to model the distribution of elements of a texture vocabulary inspired by Marr's (Marr, 1982) primal sketch.

The Brownian image model is another example of a statistical model of images and we will discuss this simple model further in Sec. 1.1.3.

The study of the probability distribution of images $p(I)$ is also important from an information theoretic point of view. Knowledge of $p(I)$ might help us develop efficient image representations and compression schemes. Sparse coding (Olshausen and Field, 1996; Pece, 2002) and ICA (see e.g. (Hyvärinen, 1999)) are examples of methods that tries to take advantage of the structure of the distribution $p(I)$ of natural images.

A probability distribution on images also defines a metric on the space of images. Lillholm et al. (2002) tries to characterize the amount of information carried by features, based on various image norms that reflects simple models of the probability distribution of images. Better knowledge of the probability distribution of images would lead to a better understanding of the amount of information carried by features.

[5] See Friedlander and Joshi (1998) for an introduction to the theory of Schwartz distributions.

I would like to point out that the work by Griffin et al. (2002) on 1D profiles of natural images is related to my work discussed in Sec. 1.1.4 and Chapters 3 and 4. Griffin et al. (2002) base their analysis on the concept of metamerism and show empirically that the most likely explanation of measurements of zeroth and first order derivatives of Gaussians along 1D profiles in natural images corresponds to step edges. That is, this work tries to characterize the variation of natural images found within the so-called local spatial metamery classes, i.e. the class of all images that give the same measurements with the chosen filter bank. My work can be seen as an attempt to characterize the variation of measurements that define metamery classes, contrary to characterizing the variation of images within a metamery class.

1.1.3 The Brownian Image Model and Its Applications

The Brownian image model[6] is an example of a probabilistic model on the space of images and can be seen as a crude model of natural images.

The classical Brownian (or just the Brownian) image model is a scale invariant Gaussian distributed stochastic field (or function) on the plane. The fractal Brownian image model is a generalization of this stochastic field, which is not in general scale invariant but self similar. Furthermore, the fractal Brownian image does not have to be a Gaussian distributed stochastic field, but can in general have other types of distributions.

In chapter 2, I use the fractal Brownian image model as a basis for a method of selecting appropriate scales in feature detection. The scaling of Gaussian scale-space derivatives of fractal Brownian images are analyzed. On the basis of this analysis I propose an extension of Lindeberg's (1998b) method for scale normalization of differential invariants used in conjunction with automatic scale selection. In this method I assume either constant energy over scale or self similar energy scaling. This method also leads to a scheme for estimation of the fractal dimension of images.

In Chapter 5, I will discuss both the Gaussian and Brownian image models in more details. The Brownian image model is a scale invariant Gaussian stochastic model. Natural images have been shown to have a scale invariant covariance structure through the $1/|\xi|^\alpha$ law, $\alpha \approx 2$, of their ensemble power spectrum. The classical Brownian image also has a power spectrum proportional to $1/|\xi|^2$. I will argue that the Brownian image model is the least committed, scale invariant, statistical image model which describes the second order statistics of natural images. The model is least committed in

[6]The two dimensional variant of what is known as the Brownian motion.

the sense that it is Gaussian, hence it is the maximum entropy solution constrained only by the mean and covariance structure. I will discuss various properties of the Brownian image model in relation to linear scale-space theory, and show empirically that the second order statistics of natural images mapped into k-jet space may, within some scale interval, be modeled by the Brownian image model. This is consistent with the $1/|\xi|^2$ law. Furthermore, the distribution of Brownian images mapped into k-jet space is Gaussian and an analytical expression can be derived for the covariance matrix of Brownian images in k-jet space, which is also a good approximation of the covariance matrix of natural images in k-jet space. The consequence of these results is, that we can use the Brownian image model as a least committed model of the covariance structure of the distribution of natural images. This is interesting for the study of the statistics of scale-space features, such as toppoints (Johansen et al., 1986; Johansen, 1997), of natural images. These results can also be seen as a partial justification of our use of the fractal Brownian image model in Chapter 2.

1.1.4 Statistics of Local Image Structures

Recently, there has been a great deal of interest in modeling the non-Gaussian structures of natural images. However, despite the many advances in the direction of sparse coding and multi-resolution analysis, the full probability distribution of local geometry has not yet been described. The work of Chapters 3 and 4 can be seen as an attempt to get knowledge about both the distribution of images $p(I)$ and the distribution of descriptions $p(D)$, both introduced in Sec. 1.1.1. The level of description D, in which I am interested, is David Marr's primal sketch. That is, I operate with the concept of local geometric primitives or features, which Marr proposed as the building blocks of a visual system. Since they are building blocks, they should be part of our prior knowledge of natural images, and we should know how often one sees an edge, a blob, a corner, etc. This information should be used as our a priori knowledge when we develop computer vision algorithms. If our models or descriptions D of images have such geometric primitives as "parameters", we should know what constitutes likely "values" of these parameters. In both chapters I focus on the geometry of edges and investigate how the local geometry of natural images distribute with respect to edges. That is, I settle for estimating the marginal distribution of edges instead of the full probability distribution of images $p(I)$.

The reason for focusing on edges is first of all that edges are important in many applications. Second, experiments have shown (Reinagel and Zador, 1999) that in the human visual system, areas of high contrast or high gradient, such as edges, are visited most often while gazing at an image. This indicates to us that edges might carry a lot of important information about an image.

There are other important perceptual cues besides image intensity geometry, such as color and lighting effects like shadows, but I choose to disregard these for now, keeping in mind that at a later stage they should be included in the study of natural image statistics. It is interesting to notice that Caselles et al. (2002) argue that there are no geometric information hidden in the color channels of an image, which is not present in the corresponding gray value image. They conclude that color is a semantic cue and is not carrying any geometric information.

Chapter 3 is an exploration of the space of data points representing the values of 3×3 pixel high-contrast patches from optical and 3D range images. The results show that the distribution of data is extremely "sparse" with the majority of the data points concentrated in clusters and non-linear low-dimensional manifolds. Furthermore, a detailed study of probability densities allows us to systematically distinguish between images of different modalities (optical versus range), which otherwise display similar marginal distributions. This work indicates the importance of studying the full probability distribution of natural images, not just marginals, and the need to understand the intrinsic dimensionality and nature of the data. I believe that object-like structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes in state space. This study of natural image statistics accounts for local geometries (such as edges) in natural scenes, but does not impose such strong assumptions on the data as independent components or sparse coding by linear change of bases.

In Chapter 4, I investigate the statistics of local geometric structures in natural optical images using the linear scale-space image representation. Previous studies ((Lee et al., 2001; Lee et al., 2002) and Chapter 3) of high-contrast 3×3 natural image patches have shown that, in the state space of these patches, we have a concentration of data points along a low-dimensional non-linear manifold that corresponds to edge structures. In this chapter, the analysis is extended to a filter-based multi-scale image representation, namely the local 3-jet of linear scale-space representations. A new picture of natural image statistics seems to emerge, where primitives (such as edges, blobs, and bars) generate low-dimensional non-linear structures in the state space of image data.

1.2 Part II: Motion Analysis

The second part of the thesis focuses on motion analysis of non-rigid bodies such as the motion of fluids. I have developed an algorithm for computation of the optic flow field of video sequences of non-rigid motion.

Motion analysis is a large topic within the field of computer vision, since knowledge of motion, as perceivable in image sequences, is necessary for various tasks such as object tracking, time-to-contact, and structure from motion. Motion analysis is conducted by associating a vector field of velocities to the image sequence, which describes the rate and direction of change of the intensity values. We define the *optic flow* as the velocity field, which describes the temporal changes of the intensity values of the sequence.

A large variety of methods for estimation of optic flow exists, far more than can be listed here, so the following list of references only includes methods I find important: (Horn and Schunck, 1981; Arnspang, 1988; Otte and Nagel, 1994; Lindeberg, 1995; Florack et al., 1998; Alvarez et al., 2000). See also Barron et al. (1994) for a discussion and evaluation of various methods.

The usual assumptions made in optic flow algorithms of rigidity of the objects in the scene and of affine motion, do not hold in the case of fluid motion. Fluid motion is a process consisting of local deformations and mixing and can only in the simplest case be modeled as affine motion. I have developed a method for the computation of optic flow which takes this non-rigidity and non-affinity into account.

In Chapter 6, I will present a multi-scale method for computation of optic flow fields. The optic flow field is extracted from normal flow, by fitting the normal components of a local polynomial model of the optic flow to the normal flow. This fitting is based on an analytically solvable optimization problem, in which an integration scale-space over the normal flow field regularizes the solution. An automatic local scale selection mechanism is used in order to adapt to the scale of local structure of the flow field. The performance profile of the method is compared with that of existing optic flow techniques and I show that the proposed method performs at least as well as the leading algorithms on the benchmark image sequences proposed by Barron et al. (1994). I also do a performance comparison on a synthetic fire particle sequence and apply the proposed method to a real sequence of smoke circulation in a pigsty. Both consist of highly complex non-rigid motion.

1.3 Outline of the Thesis

Some of the chapters of this thesis are reproductions of papers which have already been published elsewhere. These papers are kept in their original form in order to reflect the evolution of thoughts and ideas of my research. The exceptions are the contents of Chapter 2, based on the paper on scale normalization of differential expressions (Pedersen and Nielsen, 2000), which has been rewritten in order to improve the language, and Chapter 5 which is unpublished work.

The order of chapters in each part of the thesis reflects the chronological order of the papers associated with each chapter. This is done to make clear the development of my research over time.

No attempt has been made to make the mathematical notation consistent across chapters, but this will cause no confusion.

The contents of the chapters and relation to already published work are as follows:

Part I:

- **Chapter 2**: This chapter is on scale normalization of differential expressions for scale selection. It is based on the journal paper (Pedersen and Nielsen, 2000) which I wrote together with Mads Nielsen.

- **Chapter 3**: This chapter is on the statistics of high-contrast 3×3 patches of natural images. It is basically a reproduction of the journal paper (Lee et al., 2002) written together with Ann Lee and David Mumford. This journal paper is an extended version of the workshop paper Lee et al. (2001), which has also been published as part of Lee (2001).

- **Chapter 4**: This is on the statistics of natural images mapped into 3-jet space by linear scale-space derivatives. It is a reproduction of the conference paper (Pedersen and Lee, 2002) which was written together with Ann Lee.

- **Chapter 5**: This chapter includes various properties, thoughts and ideas about Brownian functions as a model of natural images. A revised version of this chapter will be submitted to the Scale-Space 2003 conference.

Part II:

- **Chapter 6**: This chapter presents my work on motion analysis. It is a reproduction of the conference paper (Pedersen and Nielsen, 2001) and was written together with Mads Nielsen.

Part III:

- **Chapter 7**: This last chapter includes a summary of the research presented in parts I and II and a discussion of future research directions.

Part I

Natural Image Statistics and Applications

Chapter 2

Scale Normalization of Derivatives of Natural Images

2.1 Introduction

In this paper, we look at some scaling properties of images of natural scenes and use these for scale normalization of image derivatives. Here we use the term *natural image* to denote any image of a real world scene (see Fig. 2.3 for some examples of natural images).

We use the linear Gaussian scale-space theory (Iijima, 1962; Witkin, 1983; Koenderink, 1984) as the mathematical formalism of the concept of scale. The Gaussian scale-space of an image is a continuous family of images parameterized by the scale. We can compute the scale-space of an image by convolution with the one parameter family of Gaussian functions in which the standard deviation acts as the scale parameter. Under such assumptions as causality, translation and scale invariance and a semi-group property (Weickert et al., 1997) it can be shown that the linear Gaussian scale-space is the least committed approach to the resampling of an image at a larger scale. By using scale-space theory we bypass the problem of differentiability of digital images, because differentiation of the image in scale-space may be obtained by differentiation of the Gaussian prior to convolution.

By the computation of a possibly non-linear function of image derivatives, sometimes called measures of feature strength, it is possible to detect features in images by detecting extrema or zero crossings of the feature strength (Marr, 1982; Koenderink and van Doorn, 1987; Florack et al., 1992; Lindeberg, 1994b). The core idea of scale-space feature detection is that features exist on different scales and we should apply our feature detector at the appropriate scale. In order to choose the optimal scale for a certain feature, we have to be able to compare image derivatives across scales, which requires that these derivatives are scale normalized. The choice of scale normalization is an open question. We can choose to make the derivatives dimensionless with respect to scale, as Florack et al. (1994) have proposed, which means that we should normalize the image derivatives by multiplying with the scale σ, $(\partial/\partial x)_{norm} = \sigma\partial/\partial x$. Lindeberg (1994a; 1996; 1998b; 1998a) has proposed an automatic scale selection method that uses a novel approach to scale normalization. His scale normalization is defined through a scaling exponent γ, which depends on the feature we are interested in detecting. Lindeberg determines this parameter on the basis of analysis of feature models.

In the literature one finds several investigations into the fractal nature and statistics of natural images (Mandelbrot, 1982; Pentland, 1984; Field, 1987; Gårding, 1987; Knill et al., 1990; Ruderman and Bialek, 1994). Field (1987) has shown that natural images exhibit scale invariance and have a so-called self similar[1] power spectrum. Ruderman and Bialek (1994) has shown that the self similar power spectrum varies among different classes of images[2]. Mandelbrot (1982), Pentland (1984), Gårding (1987) and Knill et al. (1990) advocate that to some extend natural images are fractals and should be modeled as such.

The graph of a function is said to be fractal if it has a so-called fractal dimension (Hausdorff dimension[3]) that differ from the topological dimension in a fractional manner. Furthermore, a fractal function is self similar. The fractal dimension of an image intuitively describes the roughness of the image intensity graph, and the fractal dimension of the intensity surface of 2D images must a priori lie between 2 and 3.

Among others Pentland (1984) has proposed the fractal Brownian motion (fBm) as a model for images of natural scenes. Pentland use this model to generate synthetic fractal images of natural objects such as mountains and clouds. The fBm model also captures the self similar properties of natural images reported by Field (1987) and Ruderman and Bialek (1994). By using fractal Brownian motion as an image model we have the freedom to model images that exhibit a self similar power spectrum and have any fractal dimension. The classical Brownian motion is a special case of the fBm. The fBm's are in general continuous, but not differentiable. In the limit $D_H \to 2$, the 2 dimensional fBm's generically become smooth (C^∞), whereas in the

[1] See Appendix 5.7 for a definition of statistical self-similarity.

[2] The studies by Ruderman and Bialek (1994) is based on a series of images taken in a forest.

[3] We will use the Hausdorff dimension as the definition of the fractal dimension. See for instance Ott (1993) for a mathematical definition of the Hausdorff dimension.

limit $D_H \to 3$, the 2 dimensional fBm's become spatially uncorrelated.

Our intuition[4] is that the γ parameter in Lindeberg's (1994a; 1996; 1998b; 1998a) scale normalization approach must reflect the local complexity of the image, and may be modeled through the fractal dimension of the local image. In this paper we conjecture a simple relation between the topological dimension of a feature and the fractal dimension of the local image for determining the scale normalization. We will in this paper assume, that the fBm's constitute a reasonable model of images of natural scenes. Using this model we establish a method of scale-space normalization of derivatives, changing the analytical expression of Lindeberg's γ-normalization. This expression includes the fractal dimension of the image in a neighborhood of the feature we want to detect. Furthermore, we can use this normalization method for estimation of the fractal dimension of images.

The estimation of the fractal dimension of regions of interest in images has different interesting prospects. It has been proposed by various authors (Fazzalari and Parkinson, 1996; Veenland et al., 1996) that the fractal dimension of x-ray images of trabecular bone can give an indication of the micro structure of the bone and thereby the biomechanical strength of the bone. This can be a helpful tool for the research of osteoporosis and other bone diseases. Levy Vehel (1996) has among others proposed to use fractal measures and especially multifractal measures for various image analysis tasks such as segmentation and edge detection. Other uses of the fractal dimension could be as a quality measure of surfaces produced in different kinds of industries, e.g. metal plates, wood etc.

2.2 Fractal Brownian Motion and Natural Images

The fractal Brownian motion is a stochastic function. It was first studied by Mandelbrot and van Ness (1968) who defined it in terms of stochastic integrals. This definition is somewhat cumbersome and fortunately we can also define fBm's in terms of self similarity of the probability distribution function of its increments. In the later form, it is straightforward to state the fBm defined over a N dimensional space (Pentland, 1984). A stochastic function $f_H(\mathbf{x}) : \mathbb{R}^N \mapsto \mathbb{R}$ is called a *N dimensional fractal Brownian motion (fBm)* if for all positions $\mathbf{x} \in \mathbb{R}^N$ and all displacements $\Delta\mathbf{x} \in \mathbb{R}^N$

$$P\left(\frac{f_H(\mathbf{x}+\Delta\mathbf{x}) - f_H(\mathbf{x})}{\|\Delta\mathbf{x}\|^H} < y\right) \equiv F(y) \ ,$$

[4]Developed during discussions with Tony Lindeberg.

where $F(y)$ denotes the cumulative probability distribution function of the increments and $P(\cdot)$ the probability. The scaling exponent $H \in]0;1]$ determines the fractal dimension of the fBm. The fBm turns into the classical Brownian motion when $H = 1/2$ and the probability distribution $F(y)$ of the increments are Gaussian. The two dimensional classical Brownian motion is a scale invariant Gaussian stochastic process (see Chapter 5). The distribution $F(y)$ could in general be non-Gaussian. The definition of the fBm implies that any straight line in the domain of the function $f_H(\mathbf{x})$ is a one dimensional fBm with the same scaling exponent H.

The power spectrum of the N dimensional fBm $f_H(\mathbf{x})$ is proportional to

$$|\tilde{f}_H(\omega)|^2 \propto |\omega|^{-\alpha} , \tag{2.1}$$

where $\tilde{f}_H(\omega)$ is the Fourier transform of the fBm, $\omega \in \mathbb{R}^N$ is the spatial frequency vector, and $\alpha = 2H+1$, which is independent of the dimensionality N (Mandelbrot and van Ness, 1968; Pentland, 1984; Voss, 1985). Pentland (1984) and Voss (1985) note the relation, $D_H = N + (1 - H)$, between the Hausdorff dimension[5] D_H of the graph of a N dimensional fBm $f_H(\mathbf{x})$ and the scaling exponent H. The estimation of α by estimating the power spectrum and using Eq. (2.1) is, together with the relation between D_H and H, a well-known method for estimation of the fractal dimension of images (Veenland et al., 1996).

One can find experimental data to support the claim that we can use the fractal Brownian motion to model certain aspects of natural images (Field, 1987; Knill et al., 1990; Mandelbrot, 1982; Ruderman and Bialek, 1994). Field et al. (Field, 1987; Knill et al., 1990) have shown that natural images in general have a power spectrum proportional to $|\omega|^{-2}$, which implies scale invariance of the power spectrum and autocorrelation function, i.e. the second order spatial statistics. This behavior can be modeled by the classical Brownian motion $H = 1/2$. Ruderman and Bialek (1994) have shown that the power spectrum varies among different types of images and is proportional to $|\omega|^{-\beta}$ where β varies around 2, which implies self similarity of the second order spatial statistics. This behavior can be modeled by the fractal Brownian motion where $H \in]0;1]$. It should be noted that the fractal Brownian motion only models the correlation among pixels and does not capture the full complex dependencies among pixels that are common in natural images. Nevertheless we believe it is a suitable image model for the purpose of scale normalization since it captures relevant scaling properties of natural images.

[5] The Hausdorff dimension of the graph of a function can intuitively be viewed as a scaling exponent of the space filling of that function graph.

Lindeberg (1994a; 1994b; 1998b) argues that if we assume that natural images have a scale invariant power spectrum, as Field (1987) has proposed, and that scale invariance implies that the power spectrum has equal energy at all scale-invariant frequency intervals, we get that the power spectrum of these images must be proportional to $|\omega|^{-N}$. If we use the fBm as a model of natural images, we find that this only coincides with $H = 1/2$, which is the case where the images can be modeled by the classical Brownian motion. From this observation and the results of Ruderman and Bialek (1994) we can deduce that natural images do not necessarily have equal energy at all scales, but do exhibit a self similar energy scaling of the power spectrum, i.e. $|\omega|^{-\alpha}$. In this paper we want to normalize image derivatives such that they exhibit equal energy at all scales.

2.3 Scaling of Derivatives of Fractal Images

In this section we will first give a short introduction to linear Gaussian scale-space theory and scale normalized derivatives. Then we will state our extension of Lindeberg's normalization method, which we base on the assumption that natural images may be modeled by the fractal Brownian motion. Our extension is parameterized by the fractal dimension.

2.3.1 Scale-Space and Normalization

The linear Gaussian scale-space representation was independently introduced by Iijima (1962), Witkin (1983) and Koenderink (1984). The scale-space of an image $f(\mathbf{x}) : \mathbb{R}^N \mapsto \mathbb{R}$ can be defined as

$$L(\mathbf{x};t) = \int_\Omega f(\mathbf{x}')G(\mathbf{x}-\mathbf{x}';t)\,d\mathbf{x}' = f(\mathbf{x}) * G(\mathbf{x};t) \tag{2.2}$$

where $t = \sigma^2$ is the scale parameter, the notation $*$ denotes convolution in $\mathbf{x}$ over the image domain $\Omega \subseteq \mathbb{R}^N$, and $G(\mathbf{x};t) : \mathbb{R}^N \times \mathbb{R} \mapsto \mathbb{R}$ is the N dimensional isotropic Gaussian aperture function

$$G(\mathbf{x};t) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}\right) .$$

It can be shown that $L(\mathbf{x};t)$ defined in Eq. (2.2) is the solution to the heat diffusion equation (Koenderink, 1984).

The nth order partial derivative of a scale-space image with respect to x_i, the ith element of $\mathbf{x}$, can be found by using the following commutation relationship

$$\frac{\partial^n}{\partial x_i^n}(f(\mathbf{x}) * G(\mathbf{x};t)) = (-1)^n \frac{\partial^n f(\mathbf{x})}{\partial x_i^n} * G(\mathbf{x};t) = f(\mathbf{x}) * \frac{\partial^n G(\mathbf{x};t)}{\partial x_i^n} \ .$$

Hence we can compute image derivatives at the scale t by convolving the image with derivatives of the scale-space aperture function. In this paper we will in general use a simplified tensor notation when writing partial derivatives, where

$$L_{i_1 \cdots i_n}(\mathbf{x};t) = \frac{\partial^n}{\partial x_{i_1} \cdots \partial x_{i_n}} L(\mathbf{x};t)$$

and each $i_1, \ldots, i_n \in \{1, \ldots, N\}$ are indices of the N components of the vector $\mathbf{x} \in \mathbb{R}^N$. Later on we can choose a coordinate system, either global or local, thereby fixing the meaning of the indices and allowing us to write partial derivatives in this coordinate system. E.g. in the Cartesian coordinate system we get $L_{xy} = \frac{\partial^2}{\partial x \partial y} L$. Furthermore, we use Einstein's summation convention in which we implicitly assume summation over repeated indices. An example is

$$L_{ij} L_{ik} = \sum_{i=1}^{N} L_{ij} L_{ik} \ .$$

Scale normalization of image derivatives has been proposed by several authors (Lindeberg, 1994a; Lindeberg, 1998b; Elder and Zucker, 1996; Florack et al., 1994). The standard normalization of the nth scale-space derivative based on dimensional analysis (Florack et al., 1994) is

$$L_{i_1 \cdots i_n, norm}(\mathbf{x};t) = t^{n/2} L_{i_1 \cdots i_n}(\mathbf{x};t) \ ,$$

which for the first order derivatives is the same as $(\partial/\partial x)_{norm} = t^{1/2} \partial/\partial x$. Lindeberg (1994a; 1996; 1998b; 1998a) proposes a feature dependent method of scale normalization of image derivatives in connection with an automatic scale selection scheme. He proposes that the nth order derivatives could be normalized by

$$L_{i_1 \cdots i_n, \gamma_n - norm} = t^{\gamma_n} L_{i_1 \cdots i_n}$$

where $\gamma_n = n\gamma/2$ and γ is a free normalization parameter. In conjunction with feature detection Lindeberg has determined γ by an analysis of model patterns reflecting the features under consideration. In this analysis the parameter varies in the interval $[\frac{1}{2}; 1]$ (see table 2.1).

2.3.2 Scale-Space Normalization Using the Fractal Dimension

We propose that the normalization parameter γ_n can be stated as a relation of α for the image or feature (i.e. the scaling parameter H), the topological dimension N of the image, and the order of differentiation n. This is based on the assumption that images of natural scenes may be modeled by fBm's and that normalized derivatives must have equal energy at all scales.

We will investigate quadratic differential image invariants on the form

$$I^{(n)}(\mathbf{x};t) = L_{i_1 \cdots i_n}(\mathbf{x};t) \cdot L_{i_1 \cdots i_n}(\mathbf{x};t)\ . \tag{2.3}$$

We say that this kind of invariants are of the nth order of differentiation. In the following we examine the L_1-norm of such invariants, $\|I^{(n)}\|_1$, which corresponds to looking at the L_2-norm of image derivatives $L_{i_1 \cdots i_n}$. That is, we examine scaling of the energy of image derivatives. Note furthermore that the L_1-norm of any other invariant quadratic in $L(\mathbf{x};t)$ of total order of derivation $2n$ is equivalent to $\|I^{(n)}\|_1$ (see Nielsen et al. (1997)).

Theorem 2.1
If $f_H(\mathbf{x}) : \mathbb{R}^N \mapsto \mathbb{R}$ is a N dimensional fBm and $L(\mathbf{x};t) : \mathbb{R}^N \times \mathbb{R} \mapsto \mathbb{R}$ is the scale-space of $f_H(\mathbf{x})$, then the nth order invariants $I^{(n)}(\mathbf{x};t)$ in this scale-space can be normalized to equal energy on all scales by the following relation

$$I^{(n)}_{norm}(\mathbf{x};t) = t^{\gamma_n} I^{(n)}(\mathbf{x};t)$$

where $\gamma_n = -\alpha/2 + n + N/2$, and $\alpha = 2H + 1$.

Proof: The proof is inspired by a similar analysis of the power spectrum of natural images by Lindeberg (1994a; 1994b; 1998b). We compute the L_1-norm of $I^{(n)}(\mathbf{x};t)$ which is equivalent to the total energy of the power spectrum and use Parseval's identity

$$\begin{aligned}
\|I^{(n)}\|_1 &= \int_{\mathbf{x}\in\mathbb{R}^N} |L^2_{i_1\cdots i_n}(\mathbf{x};t)|\, d\mathbf{x} \\
&= \int_{\omega\in\mathbb{R}^N} |\tilde{G}^2_{i_1\cdots i_n}(\omega;t)\tilde{f}^2_H(\omega)|\, d\omega \\
&= \int_{\omega\in\mathbb{R}^N} \left| \imath^n |\omega|^{2n} e^{-|\omega|^2 t} \tilde{f}^2_H(\omega) \right| d\omega \\
&= \int_{\omega\in\mathbb{R}^N} e^{-|\omega|^2 t} |\omega|^{-\alpha+2n}\, d\omega,
\end{aligned}$$

where $\imath = \sqrt{-1}$, and $\tilde{f}_\alpha(\omega)$ and $\tilde{G}_{i_1 \cdots i_n}(\omega; t)$ are the Fourier transform of the image and the nth order derivative of the Gaussian, respectively. Using the relation

$$\int_0^\infty x^m e^{-ax^2} dx = \frac{\Gamma\left((m+1)/2\right)}{2a^{(m+1)/2}}$$

and introducing N dimensional spherical coordinates, we find

$$\begin{aligned} \|I^{(n)}\|_1 &= \int_{\rho \in [0,\infty[; \varphi_1, \ldots, \varphi_{N-1} \in [0,2\pi]} \rho^{-\alpha+2n+N-1} e^{-\rho^2 t} \cdot d\rho d\varphi_1 \cdots d\varphi_{N-1} \\ &= (2\pi)^{N-1} \frac{\Gamma(-\frac{\alpha}{2} + n + \frac{N}{2})}{2t^{-\frac{\alpha}{2}+n+\frac{N}{2}}} = K \cdot t^{\frac{\alpha}{2}-n-\frac{N}{2}} = K \cdot t^{-\gamma_n}, \end{aligned}$$

where K is an arbitrary constant, and hereby we see that we must have $\gamma_n = -\alpha/2 + n + N/2$ in order to have equal energy at all scales. □

The normalization relation of Theorem 2.1 implies a special case for the 0'th order derivatives, meaning the case of the undifferentiated scale-space image. According to Theorem 2.1 we should in this case scale normalize the scale-space $L(\mathbf{x}; t)$ by an exponent $\gamma_0 = -\alpha/2 + N/2$ introduced by the fractal dimension of the original image. Normalization of the nth order derivatives can then be seen as just the normalization based on dimensional analysis, because $\gamma_n = \gamma_0 + n$. We can get an intuitive explanation for this special case of the 0th order derivatives if we think of the fBm as the fractional derivative (or integral) of the classical Brownian motion (Mandelbrot and van Ness, 1968), i.e. γ_0 is the normalization coming from the fractional derivative of the classical Brownian motion. In the case of the classical Brownian motion ($\alpha = 2$) and $N = 2$, we see that our scale normalization reduces to normalization based on dimensional analysis $\gamma_n = n$.

In conjunction with feature detection, we must use the fractal dimension of the image in a neighborhood of the feature of interest. This suggests a simple relation between the topological dimension of the feature and a suitable choice of fractal dimension. In Table 2.1 we have listed Lindeberg's (Lindeberg, 1996) suggested normalized measures of feature strength. For each feature we have calculated the values of H and D_H which correspond to his suggested γ values. It is interesting to notice that corners and blobs have a fractal dimension of 2.5 and edges and ridges only have a fractal dimension of 2. The topological dimension of corners and blobs is 0, while edges and ridges have a topological dimension of 1. Round a corner or a blob we would expect the void hypothesis of $H = \frac{1}{2}$. This is not expected to be true in a neighborhood of 1D features owned to the spatial extend and we

Feature type	Normalized feature strength	γ	H	D_H	T
Edge	$t^{\gamma/2} L_v$	1/2	1	2	1
Ridge	$t^{2\gamma}(L_{pp} - L_{qq})^2$	3/4	1	2	1
Corner	$t^{2\gamma} L_v^2 L_{uu}$	1	1/2	2.5	0
Blob	$t^{\gamma} \nabla^2 L$	1	1/2	2.5	0

Table 2.1: This table shows the γ-normalized measures of feature strength used by Lindeberg for feature detection with automatic scale selection. We have calculated the corresponding values of H and D_H using our extension of γ. This table is a reproduction of a table from (Lindeberg, 1996) extended with columns for H, D_H and the topological dimension T of the features. Note that a simple relation exists between the fractal dimension D_H and the topological dimension T. The relation between γ and T is not as straight forward.

see that Lindeberg's choice of γ leads us to the hypothesis of $H = 1$ for both 1D features. Hence for 1D features the hypothesis is that the pixels of the feature are highly correlated.

A benefit of the proposed normalization method is that the normalization relation can be used for estimation of the fractal dimension of images. This can be done by calculating the L_1-norm of a collection of differential invariants having the form given in Eq. (2.3) and then fit the graph of the logarithm of the norm values to a straight line. The slope of this line is an estimate of γ_n which then gives us an estimate of the fractal dimension D_H of the image intensity surface. We use this method in Sec. 2.4 to estimate the fractal dimension of synthetic and real images. We will not conduct a comparative study of this method and other methods for estimation of the fractal dimension of images (see (Veenland et al., 1996) for a study of other methods), but merely point out the existence of the method.

2.4 Experiments

We have conducted several experiments on synthetic and real 2D images in order to study the normalization of natural images. We can, as stated earlier, use the normalization method to find the fractal dimension of images by computing unnormalized derivatives in the scale-space of the considered image. We can estimate the value of γ_n and calculate the Hausdorff dimension D_H from these unnormalized scale-space derivatives. In the same manner one can get estimates of the local fractal dimension at a point in the original image. The fractal dimension of a point could be viewed as a contradiction in terms, but it is never the less possible to give meaning to this concept

α	Estimated γ_n	Correct γ_n	Relative error
	a) Values for a 2D image differentiated $n = 1$ times ($L_i L_i$)		
1	1.51	1.50	0.6 %
2	1.03	1.00	3.0 %
2.5	0.81	0.75	8.0 %
3	0.61	0.50	22.0 %
	b) Values for a 2D image differentiated $n = 2$ times ($L_{ij} L_{ij}$)		
1	2.50	2.50	0.0 %
2	2.00	2.00	0.0 %
2.5	1.75	1.75	0.0 %
3	1.51	1.50	0.6 %

Table 2.2: These tables show estimated γ_n values for synthetic N=2 dimensional fBm images with $\alpha = 1, 2, 2.5, 3$. The top table a) shows γ_n values estimated by computing $L_i L_i$ ($n = 1$) together with the correct γ_n values and corresponding relative error. The bottom table b) shows γ_n values estimated by computing $L_{ij} L_{ji}$ ($n = 2$). The fBm images and the corresponding graphs of the L_1-norm of $L_i L_i$ and $L_{ij} L_{ji}$ are depicted in Fig. 2.1. The γ_n values are estimated by sampling the scale-space of the two differential invariants applied to the synthetic images at 10 different scales and computing the L_1-norm of these invariants. The value of γ_n is estimated as the slope of a line fitted to the log-log graph of the L_1-norm as a function of scale t.

due to the intrinsic property of scale-space: A point in scale-space correspond to a neighborhood in the underlying image. It is the authors opinion that in principle all theory of fractal measures may be reformulated in the inherently well-posed framework of linear scale-space theory, thereby easing operationalization of fractal measures.

In Table 2.2 we show some results for synthetic fBm images (see the images in Fig. 2.1). The fBm images used for these experiments were constructed in the frequency domain and were given a power spectrum proportional to $|\omega|^2$ and a random phase. We have calculated the L_1-norm of different images from two scale-space differential invariants $L_i L_i$ and $L_{ij} L_{ji}$. On this basis we have estimated the γ_n values for the fBm images and compared them to the theoretical values from the continuous domain theory.

The method of estimating the fractal dimension that we propose is fairly accurate on synthetic images of known fractal dimensions. From Table 2.2 we see that the inaccuracy of the γ_n estimate increases with α. The reason for this inaccuracy is that when the α value is increased, the synthetic image will have structure on an increasing scale and when the α value of the image becomes *large enough* the outer scale of the large structures will exceed the

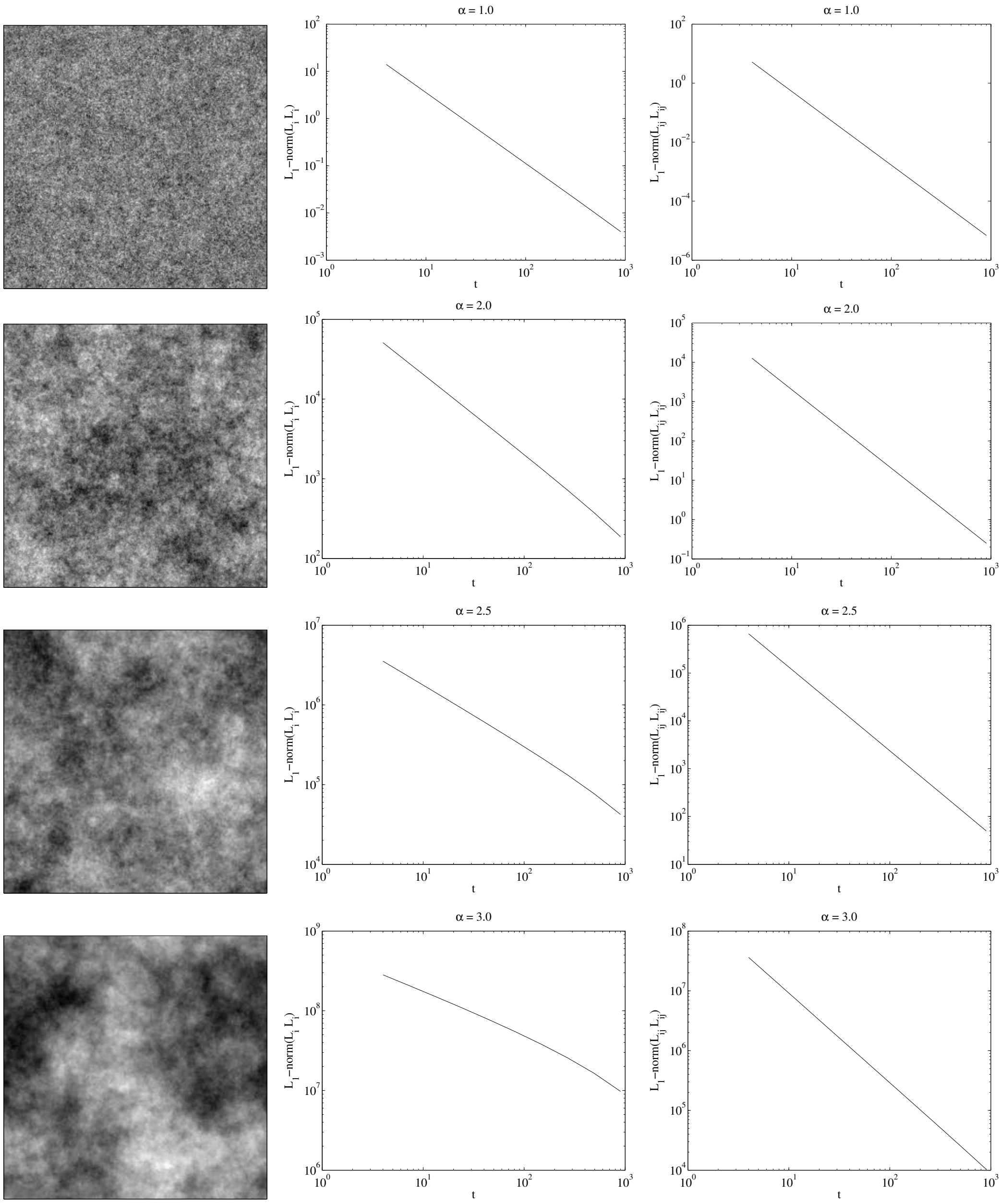


Figure 2.1: In this figure we show images of fBm's with different $\alpha = 1, 2, 2.5, 3$ values and the corresponding log-log graphs of the L_1-norm of $L_i L_i$ and $L_{ij} L_{ji}$, which we use for estimation of the γ_n values in Table 2.2. Note that $\alpha = 2$ corresponds to the classical Brownian motion. The L_1-norm graphs were produced by sampling the scale-space of the two differential invariants at 10 different scales between $t = 4$ and $t = 900$ with exponential growing increments. The synthetic images all have 256×256 pixels.

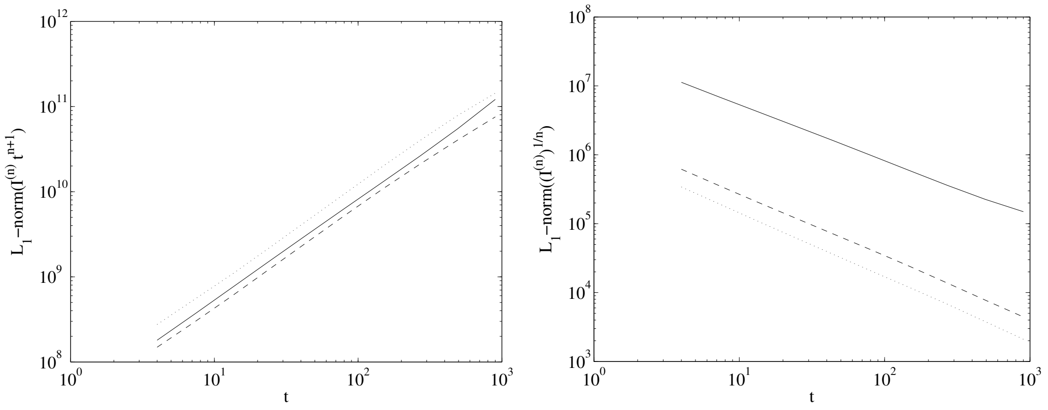


Figure 2.2: We have computed the scale-space differential invariants L_iL_i (solid line), $L_{ij}L_{ji}$ (dashed line) and $L_{ijk}L_{kji}$ (dotted line) of the tree image from Fig. 2.3. These graphs show the L_1-norm of $I^{(n)}t^{n+1}$ and $(I^{(n)})^{1/n}$ of the three differential invariants. The estimated slopes are for $I^{(n)}t^{n+1}$, $\gamma_0 = 1.19, 1.16, 1.17$ ($n = 1, 2, 3$), and for $(I^{(n)})^{1/n}$, $\gamma_0 = -0.81, -0.91, -0.93$ ($n = 1, 2, 3$).

outer scale of the image thereby misleading our method. This can also be seen on the graphs of Fig. 2.1 where the estimate of the L_1-norm becomes inaccurate at high scales, i.e. the graph bends downward, and this effect increases with α. It can also be seen from Table 2.2 that when we increase the order of differentiation we also increase the accuracy of the method. The reason for this is that when we differentiate our image we enhance the fine structure of the image by effectively looking at a scale interval, which has been moved towards smaller scales. In real examples image noise from the capture device will exhibit another structure than the random process of the scene. In general this is more uncorrelated noise and a scale interval of smaller scales will exhibit structure merely from the capture device. That is, we must choose an appropriate scale if we wish to measure scale characteristics.

We expect a logarithmic relation between the scale and $\|I^{(n)}\|_1$. From Table 2.2 and Fig. 2.1 we can see that the method for normalization proposed here is quite reasonable for synthetic images. In order to examine our method on real images, we have calculated the L_1-norm of invariants of increasing order of differentiation of the tree image from Fig. 2.3. We have normalized the computed invariants by $I^{(n)}t^{n+1}$, which corresponds to our normalization method and $(I^{(n)})^{1/n}$, which corresponds to the standard normalization method based on dimensional analysis, in order to examine the scaling property of the image independently of the order of differentiation. The slope of the logarithmic plot corresponds to γ_0 and we would expect that this slope should be approximately the same for all orders of differentiation only for our normalization method $I^{(n)}t^{n+1}$. The results can be viewed in Fig. 2.2. From this figure it can be concluded that our normalization

Title	Estimated γ_2	H	D_H
Garden	-1.90	0.60	2.40
X-rayed bone	-1.62	0.88	2.12
Water Lilly	-1.53	0.97	2.03
Sea weed	-1.98	0.52	2.48
Grains of sand	-2.09	0.41	2.59
Satellite clouds	-1.89	0.61	2.39
Landscape	-1.75	0.75	2.25
Tree	-1.82	0.68	2.32

Table 2.3: This table shows the estimated γ_2 values and the corresponding scaling exponent H and Hausdorff dimension D_H for the images in Fig. 2.3. For each image we compute $L_{ij}L_{ji}$ ($n = 2$) and estimate the γ_2 values in the same fashion as in table 2.2.

method seems as a reasonable choice, but we can also see that the γ of the standard normalization method for this image is fairly independent of the order of differentiation. If the analyzed image has a power spectrum proportional to $|\omega|^{-2}$, i.e. the power spectrum of the classical Brownian motion ($H = 1/2$), our normalization procedure will reduce to the standard normalization $\gamma_n = n$. This inconclusive experiment therefore calls for a thorough evaluation of the scaling properties of a large ensemble of images of natural scenes.

We have also tried to estimate the Hausdorff dimension of some 2 dimensional images of natural scenes. The results can be viewed in table 2.3. The estimated values of H and D_H indicates the same results as found by Ruderman and Bialek (1994): The Hausdorff dimension of images of natural scenes are not necessarily close to $D_H = 2.5$. Unfortunately we have no way of determining the error on the results in this table.

Strictly speaking, the x-rayed bone and satellite images in Fig. 2.3 do not fit within our definition of natural images due to their modalities. We include them to emphasize that the proposed method for normalization and estimation of the fractal dimension is independent of the class of images as long as the power spectrum of this class can be modeled by the fractal Brownian motion model.

2.5 Conclusions

We have related Lindeberg's (1994a; 1996; 1998b; 1998a) scale-space normalization method to the notion of fractal dimension, assuming that images of

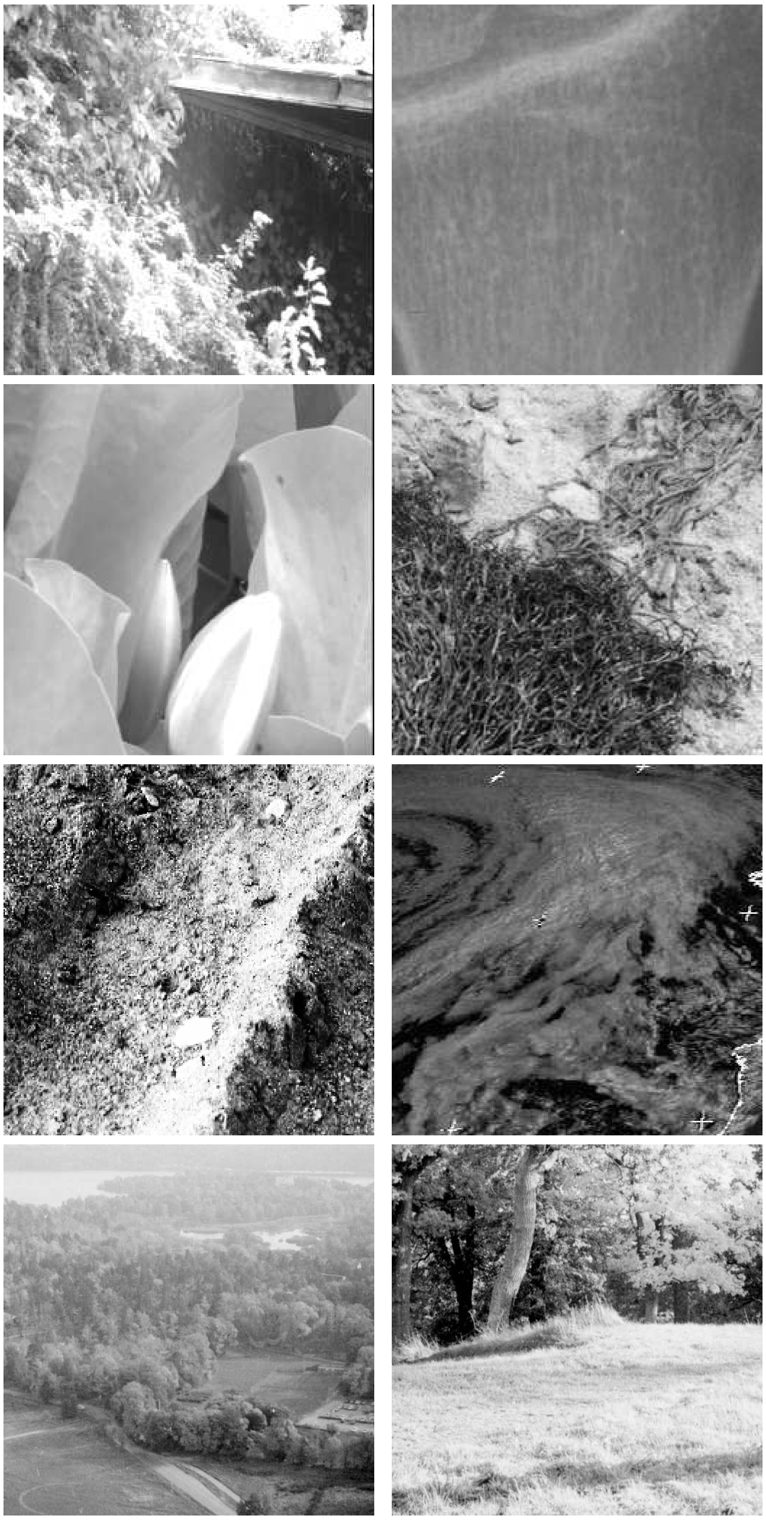


Figure 2.3: Here we show examples of natural images. The first six images all have 256×256 pixels and the last two have 512×512 pixels. We name the images from the top left corner going in the reading direction; Garden, X-rayed bone, Water Lilly, Sea weed, Grains of sand, Satellite clouds, Landscape and Tree.

natural scenes can be modeled by fractal Brownian motion. We propose that feature strength measures are normalized using an expression involving the fractal dimension of the local image.

We have found a normalization expression that has the Hausdorff dimension as a parameter. This expression reveals a possible relation between the topological dimension of the feature of interest and the fractal dimension of the local image around the feature (see Table 2.1). We conjecture (for future experimental testing):

> *The topological dimension of the feature uniquely determines the scale-space normalization parameter.*

We propose a further investigation into the relationship between different features and their Hausdorff dimension. It would be interesting to see whether it is possible to generalize the results described in table 2.1 and further establish a general relationship between the topological dimension of features and the fractal dimension locally in the image. Furthermore, we suggest a thorough investigation of the scaling properties of images of natural scenes using a large ensemble of images. Finally, it could be interesting to do a comparison of the performance of our normalization method and the one Lindeberg originally proposed for his automatic scale selection method.

Acknowledgments

We would like to thank Professor Tony Lindeberg for inspiring us to do this work. Furthermore, we thank Professor Peter Johansen for his comments on some of the theory of this paper.

Chapter 3

The Nonlinear Statistics of High-Contrast Patches in Natural Images

3.1 Introduction

A number of recent attempts have been made to describe the non-Gaussian statistics of natural images (Field, 1987; Ruderman and Bialek, 1994; Olshausen and Field, 1996; Huang and Mumford, 1999; Simoncelli, 1999b; Grenander and Srivastava, 2001). The interest for these studies in the computer vision community has been motivated by the search for more realistic priors for applications as diverse as object localization (Sullivan et al., 1999), segmentation (Malik et al., 2001; Tu et al., 2001), image reconstruction (Nielsen and Lillholm, 2001), denoising (Zhu and Mumford, 1998; Simoncelli, 1999a) and compression (Buccigrossi and Simoncelli, 1999).

The research in natural image statistics can roughly be divided into two related directions. Some studies involve analyzing 1D or 2D marginal statistics with respect to some fixed linear basis. Grenander and Srivastava (2001), for example, have shown that one can use a family of Bessel functions to model the 1D marginals of band-pass filtered data from a variety of different types of images. In Wegmann and Zetzsche (1990) and Simoncelli (1999b), the authors use a wavelet basis to uncover complex dependencies between pairs of wavelet coefficients at nearby spatial positions, orientations, and scales. In the other direction, there are studies of image statistics which try to find an "optimal" set of linear projections or basis functions in the state space defined by the image data (8×8 patches, for example, define a distribution in $\mathbb{R}^{64}$). The directions in state space are usually chosen according to some higher-order statistical measure reflecting the non-Gaussianity or multi-modality of the projected data density; see e.g. projection pursuit (Huber, 1985; Friedman, 1987), sparse coding (Olshausen and Field, 1996) and ICA (see Hyvärinen (1999) for a survey of ICA and related methods).

Despite the many advances in sparse coding and multi-resolution analysis, we are still short of a description of the *full* probability distribution (as opposed to marginal distributions) of pixels in a neighborhood. Furthermore, so far, there have been few attempts to make the connection between the object structure in the world and the probability distribution of natural images precise.

In this paper, we analyze the state space of local patterns of pixel values. More precisely, we examine the empirical probability distribution of 3×3 patches of optical and range images. These two types of images reflect different aspects of generators (or objects) in the world, as well as differences in image sensor properties. After preprocessing (consisting of subtracting the mean of each patch and whitening the data), the extracted 3×3 image patches define a distribution on a 7-dimensional sphere. We address the following questions: "How is the data distributed in this state space?" and "Are there any clear qualitative differences between the distributions of data from images of different modalities, e.g. optical versus range images?".

To develop statistically efficient image representations, it is important to understand how natural data is distributed in higher-dimensional state spaces. Without this knowledge of natural images we are not able to fully exploit the sparseness of the state space of the data. From a sparse coding point of view, high-density clusters and low-dimensional manifolds are especially interesting. These types of structures greatly reduce the dimensionality of the problem.

In ICA and related methods, one assumes that a *linear* change of basis (into independent components) that sparsifies the image data exists. We believe that an analysis of the probability distribution of natural images should be free from such strong assumptions as independent components, or linear decompositions of an image into a few dominant basis images. In reality, the most common rule for image formation is occlusion — which is non-linear — and the state space of image patches is rather complex with many more high-density directions than the dimension of the state space. The complexity of the data can partially be seen in the Haar wavelet statistics of natural images (Huang and Mumford, 1999; Huang et al., 2000). Take, for example, the 3D joint distribution of horizontal, vertical, and diagonal wavelet coefficients of natural images. Fig. 3.1 shows that the equi-probable surface of this distribution has several "hot spots" with 6 vertices along the axes $y = z = 0$, $x = z = 0$, $x = y = 0$ and $x = \pm y = \pm z$; and 8 local maxima

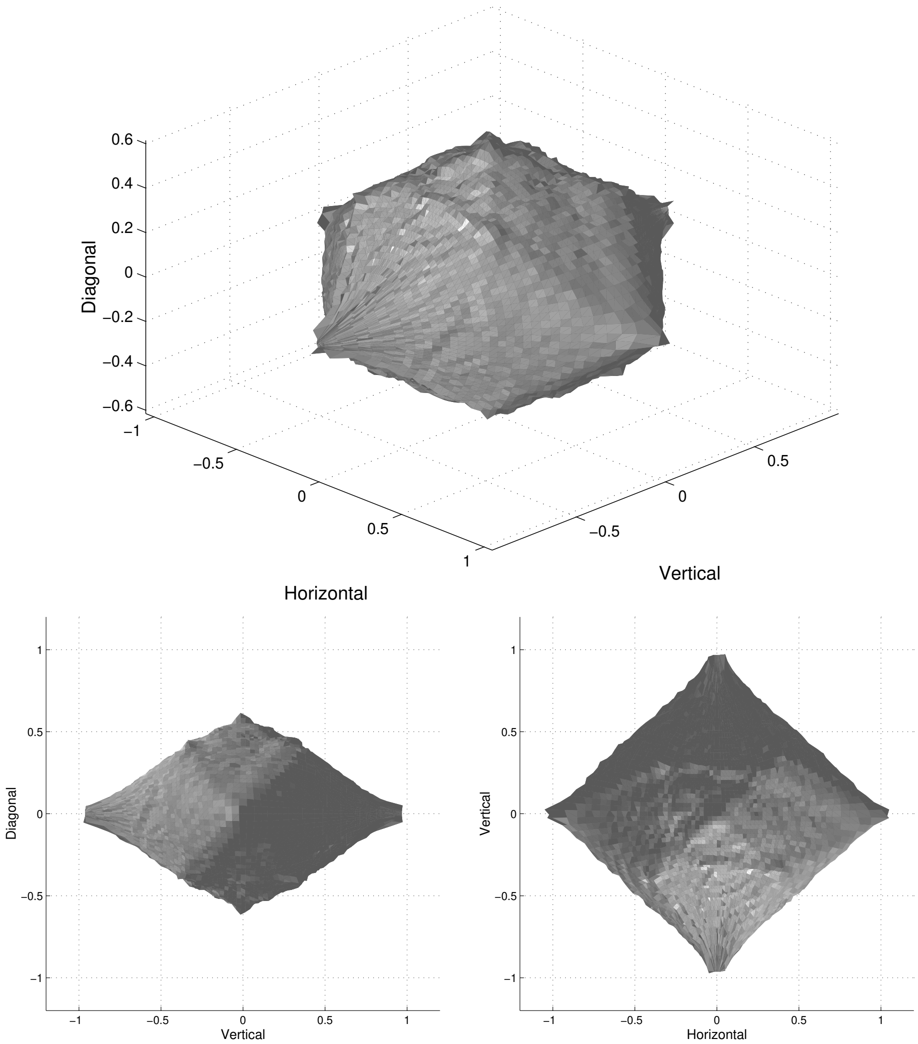


Figure 3.1: An equi-probable surface of the joint distribution of horizontal, vertical, and diagonal wavelet coefficients in optical images, viewed from three different angles.

around the shoulders $x = \pm y = \pm z$. These cusps are even more striking for range images (Fig. 3.2).

The observed cusps in Fig. 3.1 and Fig. 3.2 show very clearly frequently occurring *local geometric patterns* in pixelated natural images. A simple calculation [1] of the patterns corresponding to these cusps gives the following 2×2 blocks and their rotations:

$$\begin{pmatrix} a & a \\ b & b \end{pmatrix}, \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \begin{pmatrix} a & b \\ b & b \end{pmatrix}$$

A similar analysis can be done empirically for Haar wavelet coefficients at adjacent spatial locations in the same subband. Such a study of the 2D joint histogram of these so called wavelet "brothers" will reveal frequent occurrences in 2×4 patches of more complex local geometries best described as blobs, T-junctions, edges and bars (Huang et al., 2000).

We believe that object-like structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes (manifolds or clusters) in state space. We want to get a better understanding of how edges and other image "primitives" (see David Marr's primal sketch (Marr, 1982)) are represented geometrically in the state space of image data. Furthermore, we want to study how empirical data from natural images are distributed statistically with respect to the predicted clusters and manifolds. In other words, we are searching both for a *geometric* and *probabilistic* model in state space of the *basic* primitives of generic images.

In this study, we focus on high-contrast [2] data. It is commonly believed that image regions with high contrast carry the most important content of a scene. Reinagel and Zador (1999) have shown, for natural images with a variety of cognitive content, that humans tend to focus their eye movements around high-contrast regions — thus significantly biasing the effective input that reaches the early stages of the visual system towards these types of image

[1] The joint distribution of the 3 Haar wavelet responses

$$\begin{aligned} cH &= \frac{1}{2}(a_{00} + a_{01} - a_{10} - a_{11}) \\ cV &= \frac{1}{2}(a_{00} - a_{01} + a_{10} - a_{11}) \\ cD &= \frac{1}{2}(a_{00} - a_{01} - a_{10} + a_{11}) \end{aligned} \tag{3.1}$$

is a sufficient statistic for 2×2 blocks $\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$ modulo mean.

[2] We here choose the top 20 percent highest-contrast image patches. The qualitative results of our study, however, seem very robust to the exact choice of the cut-off.

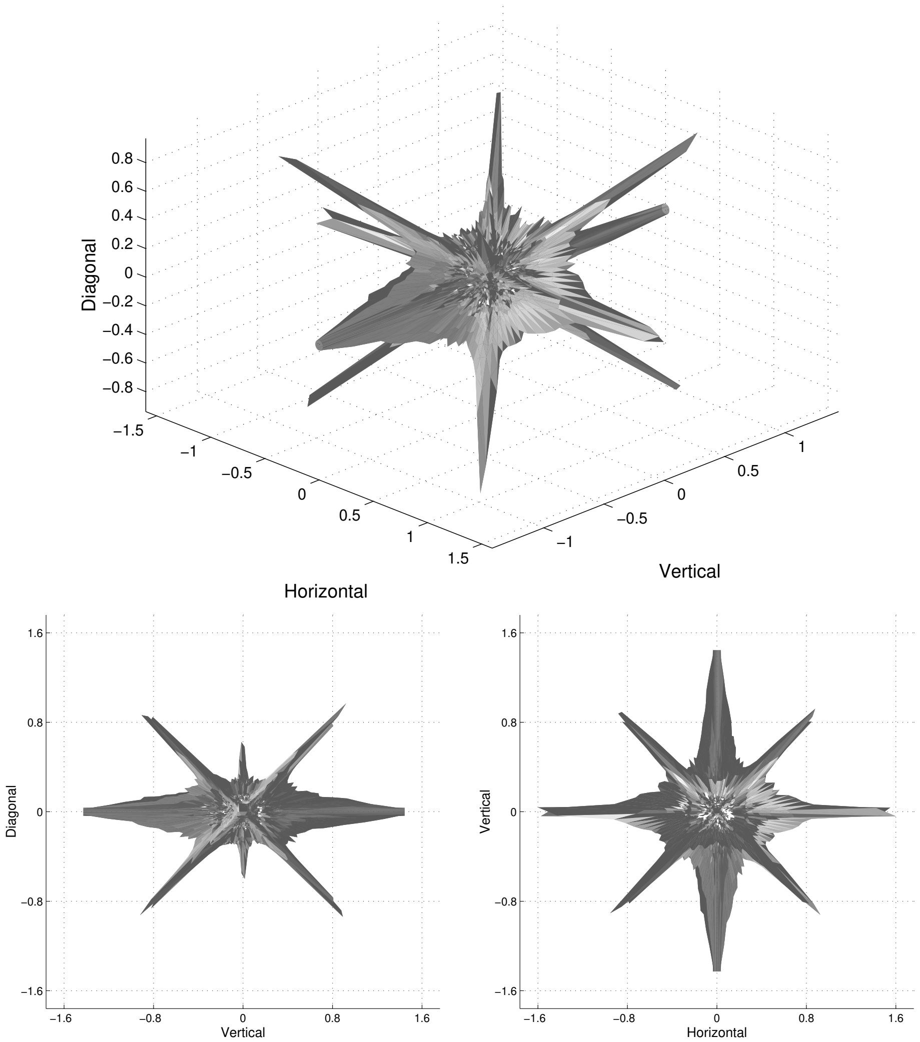


Figure 3.2: An equi-probable surface of the joint distribution of horizontal, vertical, and diagonal wavelet coefficients in range images, viewed from three different angles.

regions. Furthermore, we tend to believe that high-contrast and low-contrast regions follow qualitatively different distributions and should be modeled separately. The equi-probable contours mentioned above (see Fig. 3.1 and Fig. 3.2) are highly irregular and star-shaped in the regions far from the origin of the plot. This clearly indicates the non-Gaussian statistics of high-contrast data. The contours near the center part of the plot look different. These contours are more ellipsoidal, which suggest that fluctuations around low-contrast image regions may be Gaussian in nature.

This study deals with local patterns of pixel values, although it should be noted that the more general results (regarding the intrinsic dimension and shape of structures in state spaces) generalize to larger image patches and collections of filter responses (see the discussion in Sec. 3.7). Deriving *pixel-level* models — for example, through an iterative coarse-to-fine scheme using 3×3 or 5×5 patch structures — is also interesting by itself. What makes denoising and many computer vision applications difficult is that a natural image often contains many irrelevant, often partially resolved, objects. This type of "noise" is highly non-Gaussian and sometimes referred to as "clutter". To develop better image enhancement algorithms that can deal with structured noise, we need explicit models for the many regularities and geometries seen in local pixel patterns.

This work has some similarities to work by Geman and Koloydenko (1999). The latter study also concerns geometrical patterns of 3×3 patches. The authors quantize 3×3 blocks according to a modified order statistic and define "equivalence classes" based on photometry, complexity and geometry in image space. One of their goals is data classification for object recognition applications.

The organization of the paper is as follows: In Sec. 3.2 we describe the two data sets extracted from an optical image database by van Hateren and van der Schaaf (1998), and a range image database by Huang et al. (2000). In Sec. 3.3 we describe the preprocessing of the data sets. Our analysis is divided into three parts: In Sec. 3.4 we study the distribution of our data with respect to a Voronoi tessellation of the space of data points. This first part is a *model-free* first exploration of the state space of contrast-normalized patches. We proceed in Sec. 3.5 with a study of the probability density of high-contrast optical image patches around a 2D manifold. The manifold represents the loci in state space of blurred step edges of different orientations and positions. Finally, in Sec. 3.6 we analyze the probability density of range data around clusters corresponding to binary patches. The results are discussed in Sec. 3.7.

3.2 Optical and Range Data Sets

We extract two data sets of high-contrast 3×3 patches from optical and range images, respectively. These patches are then preprocessed as discussed in Sec. 3.3.

- The *optical* data set contains about $4.2 \cdot 10^6$ high-contrast log-intensity patches. These are extracted from van Hateren's still image collection (van Hateren and van der Schaaf, 1998) of 4167 calibrated 1020×1532 images; see Fig. 3.3 for samples. The pixel values in these images are approximately linearly proportional to the scene luminance. From each image in the database, we randomly select 5000 3×3 patches, and keep the top 20 percent, i.e. 1000 patches, with the highest contrast in log-intensities (see Sec. 3.3 for a definition of contrast or the "D-norm").

- The *range* data set contains about $7.9 \cdot 10^5$ high-contrast log-range patches. These are extracted from the Brown database [3] by Huang and Lee (Huang et al., 2000) of around 200 444×1440 range images with mixed outdoor and indoor scenes; see Fig. 3.4 for samples. We divide each image into disjoint 3×3 patches, and discard all patches with out-of-range data. From the remaining patches in each image, we randomly select 20000 patches and keep the top 20 percent with highest contrast in log-range values.

The optical and range images have quite different sub-resolution properties. In optical images, the subpixel details are *averaged* by the point-spread function of the camera. The pixel values in a range image, on the other hand, usually correspond to the *minimum* of the sub-resolution details. The field of view of the scanner is 80° vertically and 259° horizontally. The beam divergence of the laser range finder is approximately 3 mrad. If the laser footprint hits two targets with a range difference larger than 3 meters, the returned value is the range to the nearest target. If the range difference is less than 3 meters, the returned value is roughly a weighted average of both ranges where the weight depends on the reflectivity of the two targets.

3.3 Preprocessing

We want to compare and group image patches based on their *geometrical structure*. In a natural scene, the reflectance and the shape of a surface are

[3] Available at http://www.dam.brown.edu/ptg/brid/

Figure 3.3: Samples from the van Hateren optical image database. The gray values code for log-intensity values.

usually fixed quantities, while the absolute distance to, and the illumination of, a point in a scene can vary widely. We thus work with the *logarithm*, rather than the absolute values, of intensity [4] or range. Furthermore, before analyzing the data, we subtract the mean and contrast-normalize each image patch.

Let $\mathbf{x} = [x_1, \ldots, x_9]^T = [I_{11}, I_{21}, I_{31}, I_{12}, \ldots, I_{33}]^T \in \mathbb{R}^9$ be a non-constant vector with the log-values of the original patch. Subtracting the mean and contrast normalizing lead to a new vector

$$\mathbf{y} = \frac{\mathbf{x} - \frac{1}{9}\sum_{i=1}^{9} x_i}{||\mathbf{x} - \frac{1}{9}\sum_{i=1}^{9} x_i||_D} \ . \tag{3.2}$$

The contrast $||\mathbf{x}||_D$, or "D-norm", is here calculated by summing the differences between 4-connected neighbors ($i \sim j$) in a 3×3 patch and then taking

[4] The logarithmic compression of intensities is consistent with the inverse relationship between ambient illumination and human sensitivity. According to Weber's law, the ratio $\Delta L/L$ of the just noticeable difference ΔL and the ambient luminance L, is constant for a wide range of luminances. It is argued that the human visual system uses this adaptation scheme to preserve the relative reflectances (brightness ordering) of an array of surfaces in a scene.

Figure 3.4: Samples from the Brown range image database by Huang and Lee. The gray values code for log-range values.

the square root, i.e.

$$||\mathbf{x}||_D = \sqrt{\sum_{i\sim j}(x_i - x_j)^2}. \tag{3.3}$$

In matrix form, we have

$$||\mathbf{x}||_D = \sqrt{\mathbf{x}^T D\mathbf{x}}, \tag{3.4}$$

where

$$D = \begin{bmatrix} 2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}. \tag{3.5}$$

The preprocessed data points lie on a 7-dimensional ellipsoid $\tilde{\mathrm{S}}^7 \subset \mathbb{R}^9$, where

$$\tilde{\mathrm{S}}^7 = \{\mathbf{y} \in \mathbb{R}^9 \; : \; \sum_{i=1}^{9} y_i = 0 \; , \; \mathbf{y}^T D\mathbf{y} = 1\} \; . \tag{3.6}$$

For convenience, we make a change of basis to a coordinate system where the data points lie on a Euclidean sphere. In the case of scale-invariant images, this is exactly equivalent to whitening the data [5]. The 2-dimensional Discrete Cosine Transform (DCT) basis of a 3×3 image patch diagonalizes the matrix D. In vector form, we write the 8 non-constant DCT basis vectors as

$$\begin{aligned}
\mathbf{e}_1 &= \tfrac{1}{\sqrt{6}}[\;1, \; 0, \; -1, \; 1, \; 0, \; -1, \; 1, \; 0, \; -1]^T \\
\mathbf{e}_2 &= \tfrac{1}{\sqrt{6}}[\;1, \; 1, \; 1, \; 0, \; 0, \; 0, \; -1, \; -1, \; -1]^T \\
\mathbf{e}_3 &= \tfrac{1}{\sqrt{54}}[\;1, \; -2, \; 1, \; 1, \; -2, \; 1, \; 1, \; -2, \; 1]^T \\
\mathbf{e}_4 &= \tfrac{1}{\sqrt{54}}[\;1, \; 1, \; 1, \; -2, \; -2, \; -2, \; 1, \; 1, \; 1]^T \\
\mathbf{e}_5 &= \tfrac{1}{\sqrt{8}}[\;1, \; 0, \; -1, \; 0, \; 0, \; 0, \; -1, \; 0, \; 1]^T \\
\mathbf{e}_6 &= \tfrac{1}{\sqrt{48}}[\;1, \; 0, \; -1, \; -2, \; 0, \; 2, \; 1, \; 0, \; -1]^T \\
\mathbf{e}_7 &= \tfrac{1}{\sqrt{48}}[\;1, \; -2, \; 1, \; 0, \; 0, \; 0, \; -1, \; 2, \; -1]^T \\
\mathbf{e}_8 &= \tfrac{1}{\sqrt{216}}[\;1, \; -2, \; 1, \; -2, \; 4, \; -2, \; 1, \; -2, \; 1]^T
\end{aligned} \tag{3.7}$$

where the normalization is chosen such that $||\mathbf{e}_1||_D = \ldots ||\mathbf{e}_8||_D = 1$.

[5]The "D-norm" $||I||_D^2 = \sum_{i\sim j}(I_i - I_j)^2$ is a finite version of the unique scale-invariant norm $\int\int ||\nabla I||^2 dxdy \propto \int\int(\xi^2+\eta^2)||\hat{I}||^2 d\xi d\eta$ on images $I(x,y)$ (with Fourier transforms $\hat{I}(\xi,\eta)$). See also Chapter 5.

Let the DCT basis vectors above be the columns of a 9×8-matrix $A = [\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_8]$, and introduce a diagonal 8×8-matrix Λ with the diagonal elements equal to $1/||\mathbf{e}_1||^2$, $1/||\mathbf{e}_2||^2$, …, $1/||\mathbf{e}_8||^2$. A change of basis taking $\mathbf{e}_i$ to be the unit vectors according to

$$\mathbf{v} = \Lambda A^T \mathbf{y} \ , \tag{3.8}$$

or equivalently, $\mathbf{y} = A\mathbf{v}$, will then transform the ellipsoid $\tilde{\mathrm{S}}^7$ in Eq. (3.6) to a 7-dimensional Euclidean sphere

$$\mathrm{S}^7 = \{\mathbf{v} \in \mathbb{R}^8 \ : \ \sum_{i=1}^{8} v_i = 0 \ , ||\mathbf{v}|| = 1\} \ . \tag{3.9}$$

The 7-sphere S^7 is the *state space* of our preprocessed data points.

We would like to be able to measure the distance between two 3×3 image patches P_1 and P_2. Since the contrast-normalized data is on a sphere, we simply calculate the *angular distance* between the corresponding two points $\mathbf{v}_1, \mathbf{v}_2 \in \mathrm{S}^7 \subset \mathbb{R}^8$ on the sphere. In other words, our distance measure is given by

$$\begin{aligned} \mathrm{dist}(P_1, P_2) &= \arccos(\mathbf{v}_1 \cdot \mathbf{v}_2) \\ &= \arccos\left(\frac{B\mathbf{y}_1 \cdot B\mathbf{y}_2}{||B\mathbf{y}_1|| \ ||B\mathbf{y}_2||} \right) \ , \end{aligned} \tag{3.10}$$

where the matrix $B = \Lambda A^T$, and the vectors $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^9$ represent the centered 3×3 image patches P_1 and P_2.

3.4 A First Exploration of the 7-Sphere

As a first exploration of our two data sets, we divide the 7-sphere $\mathrm{S}^7 \subset \mathbb{R}^8$ into Voronoi cells, and analyze how the data points are distributed with respect to the tessellation. This is a model-free exploration of the state space, where we derive non-parametric probability distributions.

Assume a discrete collection of sampling points $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ in S^7. A Voronoi cell Ω_i around a sampling point $\mathbf{P}_i$ is defined as the set of all points $\mathbf{x} \in \mathrm{S}^7$ that are at least as close to $\mathbf{P}_i$ as to any other sampling point $\mathbf{P}_j$; that is,

$$\Omega_i = \{\mathbf{x} \in \mathrm{S}^7 \mid \mathrm{dist}(\mathbf{x}, \mathbf{P}_i) \leq \mathrm{dist}(\mathbf{x}, \mathbf{P}_j) \text{ for any } \mathbf{P}_j \in \mathcal{P}\}, \tag{3.11}$$

where $\mathrm{dist}(\cdot, \cdot)$ is the angular distance, as defined in Eq. (3.10), between two points on the sphere.

The problem of choosing a dense set of sampling points on a sphere $S^7 \subset \mathbb{R}^8$ is analogous to the problem of packing spheres in $\mathbb{R}^8$ itself. For a fixed number of sampling points, we seek a set of points such that equal non-overlapping spheres centered at the points cover the sphere "efficiently", in the sense that the space not covered by these spheres is minimal. This is a non-trivial problem in the general n-dimensional case (see Conway and A. Sloane (1988) for an in depth exposé of sphere packing in higher dimensions). In the case of 8-dimensional lattices, however, an optimal solution given by the so called E_8 lattice exists. There are several possible coordinate systems for E_8. Using the "even" coordinate system, we obtain

$$E_8 = \{(x_1, \ldots, x_8) \mid \text{all } x_i \in \mathbb{Z} \text{ or all } x_i \in \mathbb{Z} + 1/2, \sum x_i \equiv 0 \pmod 2\}.$$

Suppose that there are N points in the E_8 lattice at a distance u from the origin. Then these points, when rescaled by dividing them by u, form a dense set of sampling points $\mathcal{P} = \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_N\}$ of S^7. The first spherical shell with $u = \sqrt{2}$ and $N = 240$ is the unique solution to the "kissing number" problem in $\mathbb{R}^8$, where one wants to arrange the maximum number of non-overlapping spheres of radius 1 so that they all touch the unit sphere. For our Voronoi sampling, we have chosen to take the 4:th spherical shell of E_8 with $u = \sqrt{8}$. After normalization, this gives us a total of 17520 Voronoi cells on the 7-sphere with roughly the same size. The sampling points $\mathbf{P}_i$ are given by the permutations and sign changes of the following five 8-vectors:

1. The 112 permutations and sign changes of $[2, 2, 0, 0, 0, 0, 0, 0]^T/\sqrt{8}$.
2. The 8960 permutations and sign changes of $[2, 1, 1, 1, 1, 0, 0, 0]^T/\sqrt{8}$.
3. The 256 permutations and sign changes of $[1, 1, 1, 1, 1, 1, 1, 1]^T/\sqrt{8}$.
4. The 7168 permutations and sign changes with the constraint that the number of minus signs is an odd number;
   $\left[\frac{3}{2}, \frac{3}{2}, \frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]^T/\sqrt{8}$.
5. The 1024 permutations and sign changes with the constraint that the number of minus signs is an even number;
   $\left[\frac{5}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right]^T/\sqrt{8}$.

From a Monte Carlo simulation (with 5 million random points on the 7-dimensional unit sphere), we get that the volumes of the 5 types of Voronoi cells above are approximately $6.3 \cdot 10^{-3}$, $1.8 \cdot 10^{-3}$, $4.1 \cdot 10^{-3}$, $1.8 \cdot 10^{-3}$, and $1.8 \cdot 10^{-3}$, respectively.

We bin our high-contrast optical and range patches into the 17520 Voronoi cells using the definition given in Eq. (3.11). We define the density of data points in the Voronoi cell Ω_i around sample point $\mathbf{P}_i$ as

$$\rho(\Omega_i) = \frac{N(\Omega_i)/\sum_i N(\Omega_i)}{\mathrm{vol}(\Omega_i)/\mathrm{vol}(\mathrm{S}^7)}, \tag{3.12}$$

where $N(\Omega_i)$ is the number of patches in the Voronoi cell Ω_i, $\mathrm{vol}(\Omega_i)$ is the volume of that cell, and $\mathrm{vol}(\mathrm{S}^7) = \sum_i \mathrm{vol}(\Omega_i) = \pi^4/3$ is the total volume of the 7-sphere.

In Fig. 3.5 and 3.6 we show the density ρ of the Voronoi cells for high-contrast optical and range patches, together with the percentage of volume occupied by the percentage of patches. We find that the distribution of data on S^7 is extremely "sparse", with the majority of data points concentrated in a few high-density regions on the sphere. For both the optical and range data, half of the patches can be divided into an optimal set of Voronoi cells that occupies less than 6% of the total volume of the 7-sphere.

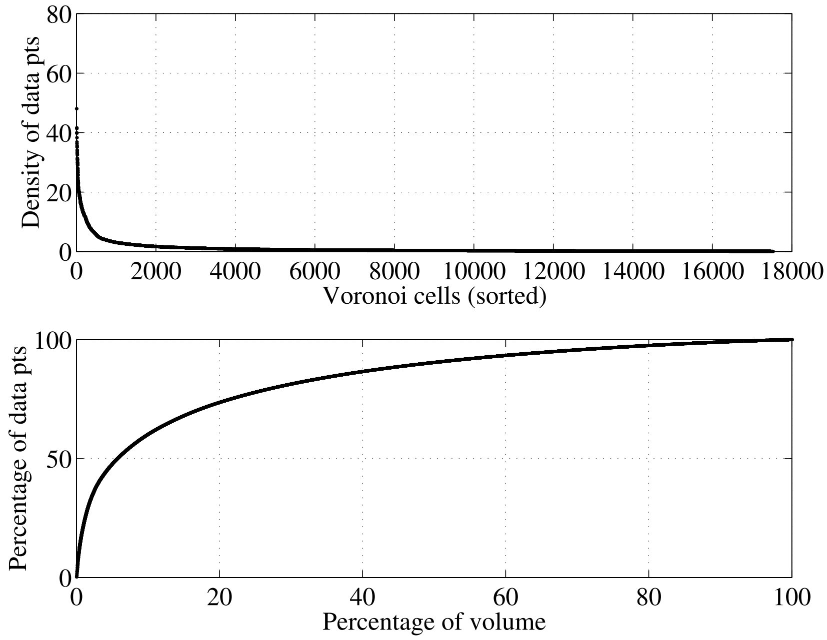


Figure 3.5: (Top) Density $\rho(\Omega_i)$ of high-contrast optical patches in Voronoi cells that are sorted according to decreasing density. (Bottom) Cumulative percentage of optical patches in the Voronoi cells above versus the cumulative percentage of volume in S^7 that are occupied by these cells.

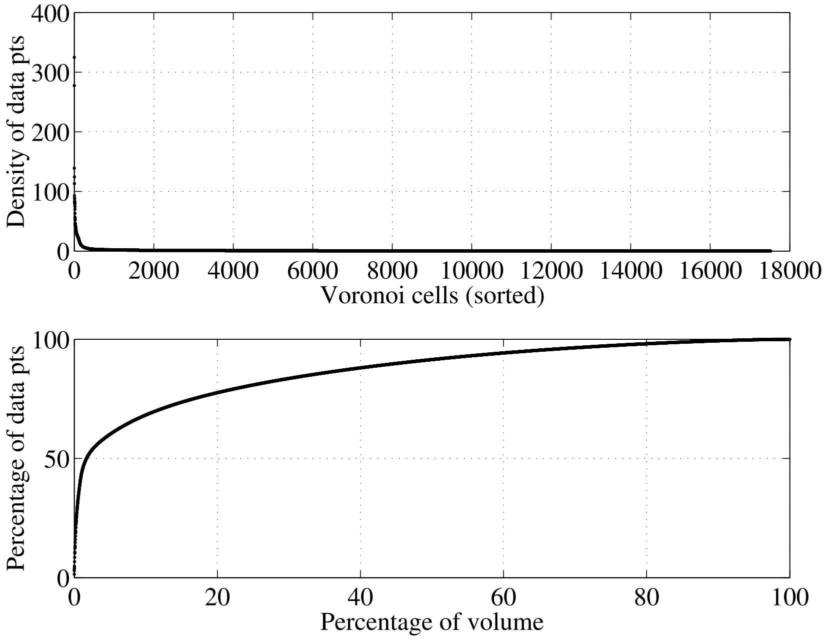


Figure 3.6: (Top) Density $\rho(\Omega_i)$ of high-contrast range patches in Voronoi cells that are sorted according to decreasing density. (Bottom) Cumulative percentage of range patches in the Voronoi cells above versus the cumulative percentage of volume in S^7 that are occupied by these cells.

We can use the Kullback-Leibler distance or relative entropy to get an information-theoretic measure of the deviation of the probability distributions of our data from a uniform distribution. Note that a Gaussian assumption on natural images corresponds to a uniform distribution in state space after whitening.

We estimate the probability density functions $p_{\mathrm{o}}(\Omega_i)$ and $p_{\mathrm{r}}(\Omega_i)$ for optical and range data, respectively, by calculating

$$p(\Omega_i) = \frac{N(\Omega_i)}{\sum_i N(\Omega_i)} \tag{3.13}$$

for $i = 1, \ldots, 17520$. As before $N(\Omega_i)$ is the number of optical or range data points in Voronoi cell Ω_i. The corresponding probability density function for a uniform distribution is defined as

$$q_{\mathrm{u}} = \frac{\mathrm{vol}(\Omega_i)}{\mathrm{vol}(\mathrm{S}^7)} , \tag{3.14}$$

where $\mathrm{vol}(\Omega_i)$ is the volume of Voronoi cell Ω_i, and $\mathrm{vol}(\mathrm{S}^7) = \pi^4/3$ is the total volume of the 7-sphere. The KL-distance between the empirical probability density of the data and a uniform density on the 7-sphere is given by

$$D(p\|q_{\mathrm{u}}) = \sum_i p(\Omega_i) \log_2 \left(\frac{p(\Omega_i)}{q_u(\Omega_i)} \right) . \tag{3.15}$$

This gives us a measure of the number of excess bits we incur by coding the data points distributed by $p_{\mathrm{o}}(\Omega_i)$ or $p_{\mathrm{r}}(\Omega_i)$ with a code book based on the uniform distribution $q_{\mathrm{u}}(\Omega_i)$ (Cover and Thomas, 1991). For our two datasets, we have that $D(p_{\mathrm{o}}\|q_{\mathrm{u}}) = 1.62$ bits and $D(p_{\mathrm{r}}\|q_{\mathrm{u}}) = 2.59$ bits. These numbers indicate that the distribution of optical or range data points in the state space of the contrast-normalized data is *highly* non-uniform.

In Fig. 3.7 and Fig. 3.8, we display the first 25 sampling points $\mathbf{P}_i$ of the Voronoi cells ordered after their densities $\rho(\Omega_i)$ (defined according to Eq. (3.12)) for optical and range data, respectively. The pixel patterns in these two figures depend of course on the exact choice of sampling scheme, and can look very different if one were to choose basis functions from a different lattice. In Fig. 3.7 for optical patches, the centers of the Voronoi cells with highest densities are close to blurred step edges (see Sec. 3.5.1). For high-contrast range patches, the cells with highest densities resemble binary patches (compare Fig. 3.8 with Fig. 3.18). Note in particular that some of the first 25 Voronoi cells here are similar to binary symmetry classes 1', 2', and 5'.

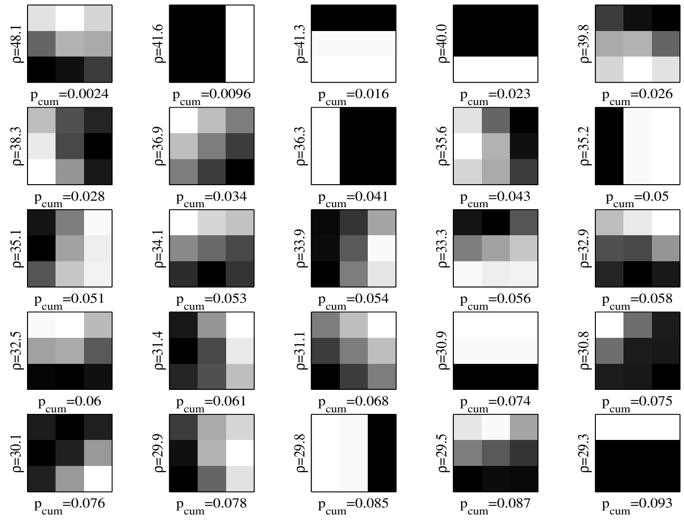


Figure 3.7: The first 25 Voronoi patches ordered after their densities $\rho(\Omega_i)$ in Eq. (3.12) for optical images. For each Voronoi cell, we display the 3×3 patch corresponding to the sample point $\mathbf{P}_i$. The cumulative sum of $p(\Omega_i)$ (Eq. (3.13)) over the ordered patches is shown as p_{cum}.

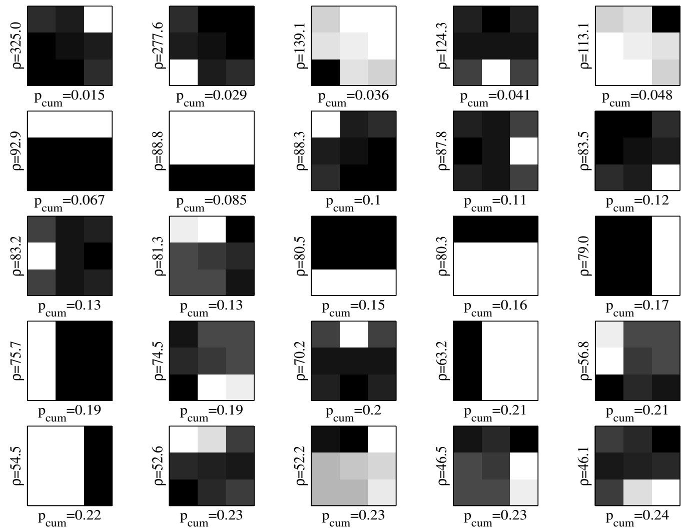


Figure 3.8: The first 25 Voronoi patches ordered after their densities $\rho(\Omega_i)$ in Eq. (3.12) for range images. For each Voronoi cell, we display the 3×3 patch corresponding to the sample point $\mathbf{P}_i$. The cumulative sum of $p(\Omega_i)$ (Eq. (3.13)) over the ordered patches is shown as p_{cum}.

3.5 Optical Data: Probability Density Around a 2D Manifold of Step Edges

The analysis of Voronoi cells on the 7-sphere indicates that blurred step edges are common high-contrast patterns in optical patches. In Sec. 3.5.1, we present an ideal model for edges in optical images that takes into account the *averaging* effects of the optics of the camera. The model predicts a 2-dimensional continuous manifold in state space, parametrized by the orientation α and position l of an edge. In Sections 3.5.2 and 3.5.3, we test the model with optical data from natural images. Finally, in Sec. 3.5.4 we apply this model to range data and discuss the differences in the results.

3.5.1 The Ideal Manifold of Edges

We represent the 3×3 image patch by a square $SQ = \{(x, y) : -3/2 \leq x, y \leq 3/2\}$. The pixels in the patch are given by

$$S_{ij} = \{(x, y) : \ (j-3/2) \leq x \leq (j-1/2), \ (1/2-i) \leq y \leq (3/2-i)\} \ , \quad (3.16)$$

where $i, j = 0, 1, 2$.

The pixel value $I(i, j)$ in an optical image is approximately an average of the underlying scene $\phi(x, y)$ recorded at each pixel S_{ij}, i.e. the pixel intensity

$$I(i, j) = \int_{S_{ij}} \phi(x, y) dx dy \quad (3.17)$$

where $\phi(x, y)$ is the scene luminance. For an ideal step edge,

$$\phi_{\alpha,l}(x, y) = \begin{cases} a & \text{if } -x \sin\alpha + y \cos\alpha > l \\ b & \text{if } -x \sin\alpha + y \cos\alpha < l \end{cases} \quad (3.18)$$

where $a > b$. The parameter $\alpha \in [0, 2\pi)$ is the angle that the direction perpendicular to the edge makes with the y-axis, and the parameter $l \in (-3/2, 3/2)$ is the displacement of the edge from the origin (Fig. 3.9). Thus, pixels in $I_{\alpha,l}(i, j)$ strictly above the edge have intensities a, pixels strictly below the edge have intensities b, and pixels along the edge are a weighted average of a and b. After subtracting the mean and contrast normalizing each edge patch (see Sec. 3.3), we arrive at a set of points $\mathbf{v}_{\alpha,l} \in \mathrm{S}^7 \subset \mathbb{R}^8$. It can be shown that the loci of these points, with $\alpha \in [0, 2\pi)$ and $l \in (-3/2, 3/2)$, define a C^1 2-dimensional manifold, M^2, embedded in the 7-dimensional sphere.

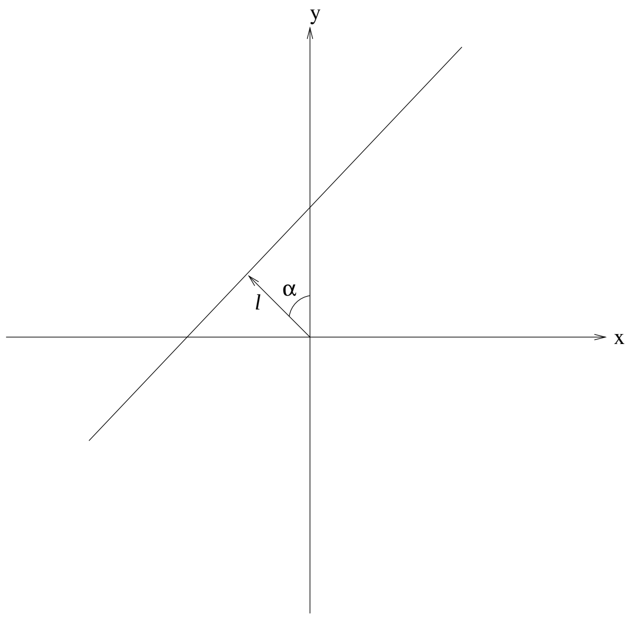


Figure 3.9: The parameter $\alpha \in [0, 2\pi)$ is the angle that the direction perpendicular to the edge makes with the y-axis, and the parameter $l \in (-3/2, 3/2)$ is the displacement of the edge from the origin.

Because of the centering and the contrast normalization, some (α, l)-values are *degenerate*, i.e. they lead to the same patch or point $\mathbf{v}_{\alpha,l} \in \mathrm{M}^2$ after preprocessing. This situation occurs for patches with "glancing edges". Assume for example that the surface parameter $l > 0$. Then consider all edges that connect the points $(-1.5, y)$ and $(x, 1.5)$ on the border of the patch, where $x \in (-1.5, \infty)$ is *fixed* and y can take *any* value in the interval $[0.5, 1.5)$. Simple trigonometry shows that for a fixed value of x, this defines a set of (α, l)-values that correspond to the *same* contrast-normalized $\mathbf{v}_{\alpha,l}$-patch. Rotations by $\pi/2$, reflections and contrast sign inversions (see the 16 symmetries in Eq. (3.28)) give the full family of equivalent edge patches. Each line in Fig. 3.10 represents one set of equivalent (α, l)-values in this family. There are two special cases of the example above with edges through the points $(-1.5, y)$ and $(x, 1.5)$. One special case is when $x \in (-1.5, -0.5]$: *All* edges with $-1.5 < x \leq -0.5$ and $0.5 \leq y < 1.5$ lead to the same contrast-normalized patch. The light shaded regions in Fig. 3.10 represent these "corner patches" and their symmetries. Another special case is when $x \to \infty$: This limit case and its symmetries "converge" to the set of equivalent (α, l)-values where $0.5 \leq l < 1.5$ or $-1.5 < l \leq -0.5$ and $\alpha = 0, \pi/2, \pi$ or $3\pi/2$ (horizontal and vertical edges). *Non-degenerate* edge patches are given

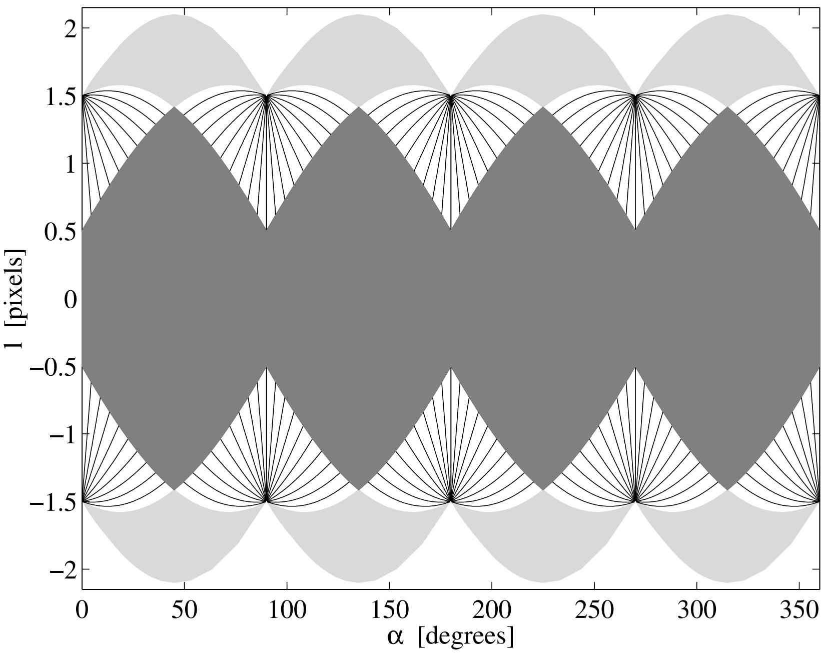


Figure 3.10: Because of the contrast normalization, some (α, l)-values are degenerate, i.e. they correspond to the same point $\mathbf{v}_{\alpha,l}$ on the 7-dimensional unit sphere S^7. Each curve in the figure is an example of a set of (α, l)-values that lead to the same contrast-normalized 3×3 patch. The light shaded region corresponds to degenerate values for corner edges. The dark shaded region in the interior of the graph shows all (α, l)-values that are well-defined, i.e. non-degenerate.

by $0 \leq \alpha \leq \pi/4$ and $0 \leq l < (1.5 \sin\alpha + 0.5 \cos\alpha)$, and its 16 symmetries. In the figure, these (α, l)-values are represented by the dark shaded region.

Fig. 3.11 shows a few examples of ideal edge patches that correspond to different (α, l)-values between $0 \leq \alpha \leq 180$ degrees and $1.5 < l < -1.5$. The step edges are here chosen on a triangular grid with the spacing $\Delta\alpha = 15$ degrees and $\Delta l = 1/4$ pixel units. Numbers between patches represent the angular distances in degrees between nearest neighbors. Although we use an even sampling of grid points in the (α, l)-coordinate system, the distances between the nearest neighbors vary widely.

Fig. 3.12 shows the geometry of the surface M^2 of step edges of different orientations and positions more clearly. Here we estimate the surface metric $f(\alpha, l) = \frac{dA}{d\alpha\, dl}$ numerically with a triangulated mesh that is much finer spaced than in the example above: We first divide the surface into rectangular bins with widths $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixels, hereafter we discard all bins that are completely outside the interior region of the

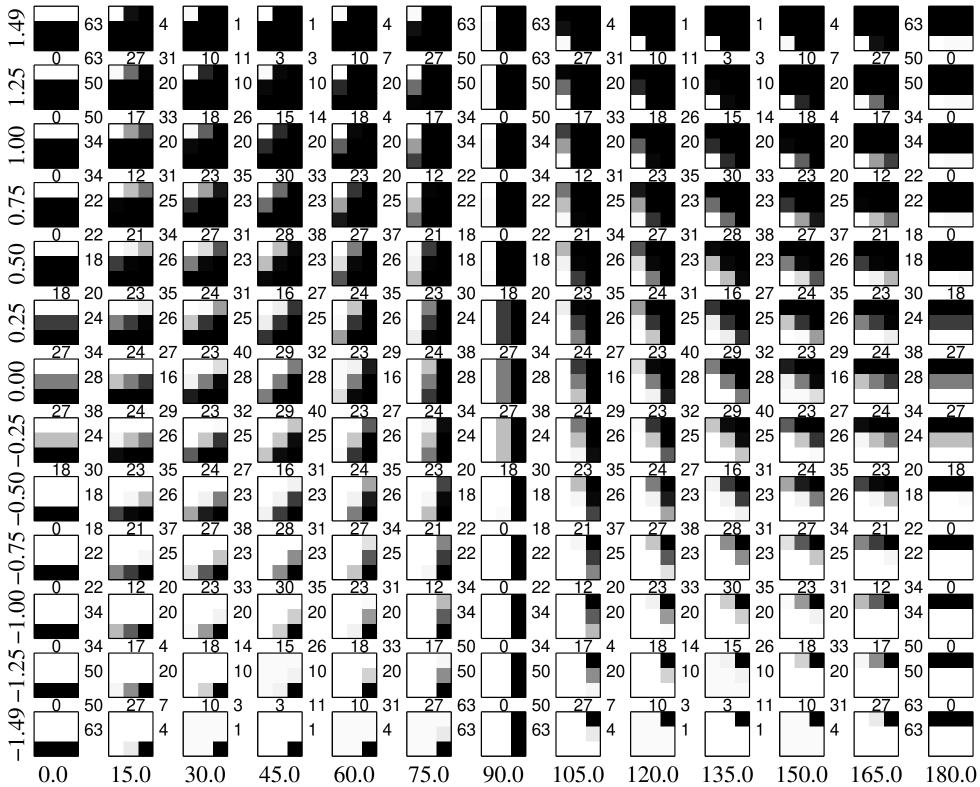


Figure 3.11: Examples of ideal step edges between $0 \leq \alpha \leq 180$ degrees (horizontal axis) and $1.5 < l < -1.5$ (vertical axis). The step edges are here chosen on a triangular grid with sides $\Delta\alpha = 15$ degrees and $\Delta l = 1/4$ pixel units. Numbers between patches represent the angular distances in degrees between neighbors in the horizontal, vertical and (lower left - upper right) diagonal directions in the (α, l)-grid.

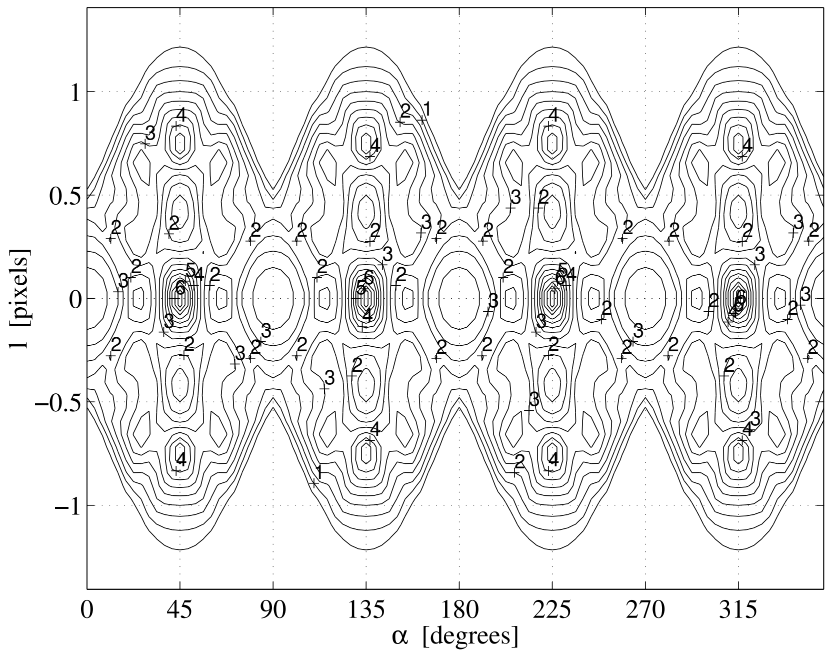


Figure 3.12: Contour plot of the metric $f(\alpha, l) = \frac{dA}{d\alpha\, dl}$ for a surface of ideal step edges. The rectangular bins in the figure have widths $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixels. For the calculation of the area A of a bin, we add up the areas of 4 or more spherical triangles inside the bin (see text).

(α, l)-plot (see Fig. 3.10), that is, bins with degenerate points. Each of the remaining bins is then split into 4 spherical triangles, where the vertices represent blurred step edges in M^2. The mesh is finally successively refined, where needed, until the distance between any two vertices in a triangle is less than 8 degrees. The final mesh contains 14376 spherical triangles. The area of a rectangular bin in Fig. 3.12 corresponds to the sum of the areas of spherical triangles [6] inside the bin.

3.5.2 Density of Optical Data as a Function of Distance to the Ideal Edge Manifold

We now try to get a numerical estimate of the probability density of high-contrast optical data around the surface of step edges. In the following

[6]The area of a spherical triangle with radius R is equal to $A = R^2\epsilon$, where $\epsilon = \alpha + \beta + \gamma - \pi$ is the so called spherical excess. The three angles α, β and γ can be directly related to the lengths of the sides (great circular arcs) of the triangle (Bronshtein and Semendyayev, 1998).

experiment we use the "optical dataset" with $N_{\mathrm{tot}} = 4.2 \cdot 10^6$ high-contrast patches described in Sec. 3.2. As above, we model the surface with a mesh consisting of about 14000 spherical triangles, where the vertices of the triangles are blurred step edges in M^2.

For each optical data point $\mathbf{x}_n$ ($n = 1, 2, \ldots, N_{\mathrm{tot}}$), we calculate the distances to the centers of the spherical triangles in the mesh. We assume that the distance to the ideal edge manifold M^2 is approximately the same as the distance to the closest triangle center $\mathbf{v}_{\alpha_n, l_n}$, i.e. we assume that

$$\theta = \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathrm{M}^2) \approx \min_{\mathbf{v}_{\alpha_\mathrm{n},\mathrm{l_n}}} \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathbf{v}_{\alpha_\mathrm{n},\mathrm{l_n}}). \tag{3.19}$$

The error is largest for very small θ, where we sometimes get an overestimate of θ due to the finite grid spacing.

Fig. 3.13(a), top, shows a normalized histogram of the number of data points as a function of the estimated distance θ. Let

$$N(\theta) = \# \left\{ n : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2} \right\} , \tag{3.20}$$

where $\Delta\theta$ is the bin width of the histogram. Linear regression (Fig. 3.13(b), top) gives that $N(\theta) \propto \theta^{1.4}$ for small θ.

For the density estimate, we also need to calculate the volume of the set

$$B_\theta = \left\{ \mathbf{x} \in \mathrm{S}^7 : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2} \right\} . \tag{3.21}$$

Fig. 3.13(a), bottom, and Fig. 3.13(b), bottom, show the results from a Monte Carlo simulation with $V_{\mathrm{tot}} = 10^7$ sample points, $\mathbf{s}_n$ ($n = 1, 2, \ldots, V_{\mathrm{tot}}$), that are uniformly randomly distributed on a 7-dimensional unit sphere. The number of sample points in B_θ, i.e.

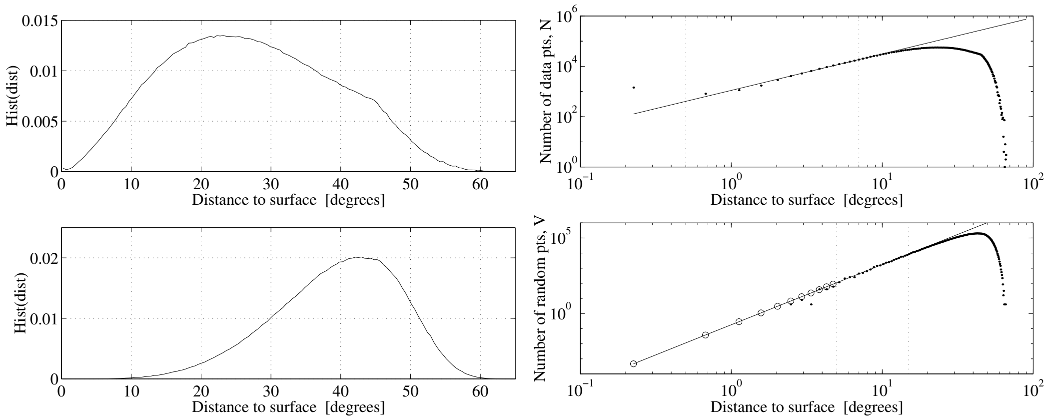
$$V(\theta) = \# \left\{ n : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{s}_n, \mathrm{M}^2) < \theta + \frac{\Delta\theta}{2} \right\} , \tag{3.22}$$

is directly proportional to the *volume* of B_θ. As expected[7], the number of random points V increases approximately as $V(\theta) \propto \theta^4$ for small θ. The

[7]For a single point $\mathbf{x}_0 \in \mathrm{S}^7$, the volume

$$\mathrm{vol}\{\mathbf{x} \in \mathrm{S}^7 : \theta \leq \mathrm{dist}(\mathbf{x}, \mathbf{x}_0) \leq \theta + d\theta\} = \mathrm{C}_6 \, (\sin\theta)^6 d\theta \ ,$$

where C_6 is the volume (surface area) of a 6-dimensional unit sphere. A 2-dimensional surface in S^7 (such as the manifold $\mathrm{M}^2 \subset \mathrm{S}^7$ or the envelop of overlapping spherical caps around the centers of the triangles in the mesh) has 5 normal directions. Hence, for smaller θ, where the sub-manifold or surface is locally linear, $\mathrm{vol}(X_\theta) \propto (\sin\theta)^4 \, d\theta$.



(a) (b)

Figure 3.13: (a) *Top*: Normalized histogram of the number of optical high-contrast patches, N, versus the distance, θ, to the surface. *Bottom*: Normalized histogram of the number of random Monte Carlo samples, V, versus θ. (b) *Top*: Log-log plot of N versus θ. Linear regression in the interval between 0.5 and 7 degrees gives $N \sim \theta^{1.4}$ (solid line). *Bottom*: Log-log plot of V versus θ. Linear regression in the interval between 5 and 15 degrees gives $V \sim \theta^{4.0}$ (solid line). The circles represent extrapolated values of V for $\theta \leq 5$ degrees. The total number of optical data points is $N_{\text{tot}} \approx 4 \cdot 10^6$. The total number of random samples is $V_{\text{tot}} = 10^7$. The random samples are uniformly distributed on a 7-dimensional unit sphere, and give a Monte Carlo estimate of the volume of the space occupied by the histogram bins.

curve for $V(\theta)$ has a maximum around $\theta = 43$ degrees, after which it drops. The drop may indicate folds in the surface where part of the 7-sphere are at equal distance to different points of the surface. Furthermore, the plot shows that all points in S^7 are within approximately 60 degrees of the edge manifold M^2. If the surface was flat, we would expect the corresponding distance to be 90 degrees (for the two antipodal points on the sphere).

Fig. 3.14 shows the density function

$$p(\theta) = \frac{N(\theta)/N_{\text{tot}}}{V(\theta)/V_{\text{tot}}} \ . \tag{3.23}$$

For $\theta \lesssim 10$ degrees,

$$p(\theta) \sim \theta^{-2.5} \ . \tag{3.24}$$

This result strongly supports the idea that there exists a 2-dimensional manifold in the 7-sphere where the data points are concentrated. In fact, we

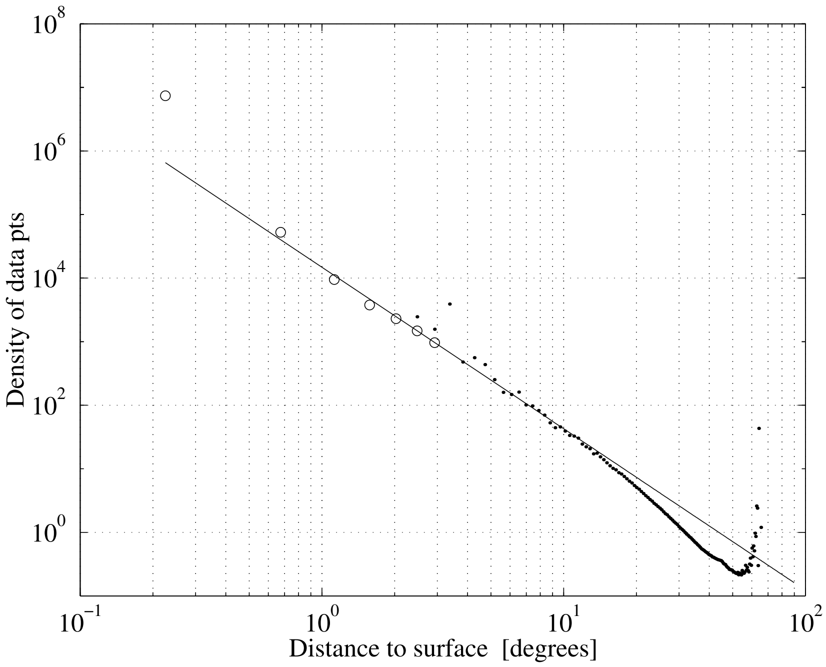


Figure 3.14: Log-log plot of the average density function $p(\theta)$ for optical data. For $\theta \lesssim 10$ degrees, $p(\theta) \sim \theta^{-2.5}$ (solid line). The circles represent the values of the ratio $\frac{N(\theta)/N_{\rm tot}}{V(\theta)/V_{\rm tot}}$ for extrapolated values of $V(\theta)$ when θ is small (Fig. 3.13b, bottom).

here see evidence of an *infinite* probability density at the ideal edge manifold (where $\theta = 0$).

In Fig. 3.15 (Top) we have plotted the percentage of data points that are within a certain distance θ of the surface. The curve shows that about 50% of the data points are within a tubular neighborhood

$$K_\theta = \{\mathbf{x} \in \mathrm{S}^7 : \mathrm{dist}(\mathbf{x}, \mathrm{M}^2) \leq \theta\} \tag{3.25}$$

of the surface with width $\theta = 26$ degrees. This neighborhood corresponds to only 9% of the total volume of S^7 (Fig. 3.15, bottom).

3.5.3 Density of Optical Data as a Function of the Surface Parameters

We next study how the high-contrast optical data are spread out *along* the surface of step edges. For the position calculation, we only include data points that are very close to the surface. For each data point $\mathbf{x}_n$ in our data set, we find the closest center point $\mathbf{v}_{\alpha_n,l_n}$ in the triangulated mesh. We

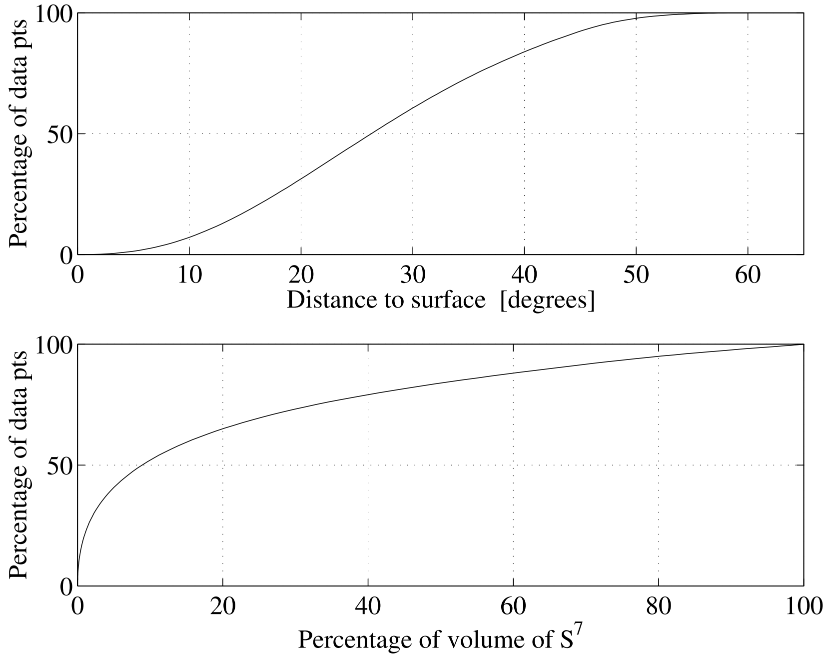
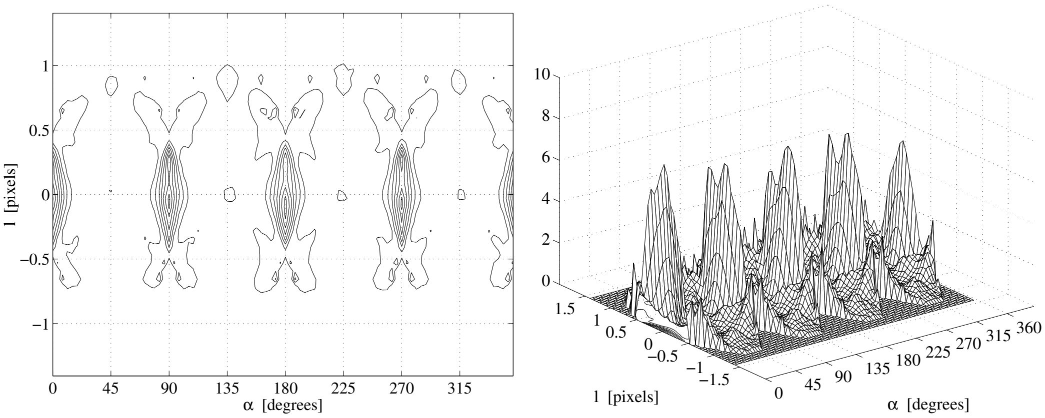


Figure 3.15: (a) Percentage of optical data points that are within a tubular neighborhood K_θ of the surface with width θ ("distance to surface"). (b) Same curve plotted versus the ratio $\frac{100\cdot \mathrm{vol}(K_\theta)}{\mathrm{vol}(\mathrm{S}^7)}$ ("percentage of volume of S^7").



(a) (b)

Figure 3.16: (a) Two-dimensional contour plot and (b) three-dimensional mesh of the density $p(\alpha, l)$ of high-contrast optical data along the surface of step edges. In the density calculation, we only include data points that are, at a position (α, l) where $f(\alpha, l) > 0.5$, and in a tubular neighborhood of the surface with width $\theta_{\max} = 20$ degrees. The bin widths are $\Delta\alpha = \pi/48$ (3.75 degrees) and $\Delta l = 1/16$ pixel units.

compute the 2D histogram

$$\begin{aligned} N(\alpha, l) = \quad & \#\{n : \mathrm{dist}(\mathbf{x}_n, \mathbf{v}_{\alpha_n, l_n}) \leq \theta_{\max}, \\ & \alpha_n \in \left[\alpha - \tfrac{\Delta\alpha}{2}, \alpha + \tfrac{\Delta\alpha}{2}\right), \; l_n \in \left[l - \tfrac{\Delta l}{2}, l + \tfrac{\Delta l}{2}\right)\} \;, \end{aligned} \tag{3.26}$$

where $\theta_{\max} = 20$ degrees, and the bin widths $\Delta\alpha = \pi/48$ radians and $\Delta l = 1/16$ pixels.

We define the density $p(\alpha, l)$ of data points along the surface as

$$p(\alpha, l) = \frac{N(\alpha, l) / \sum_{\alpha, l} N(\alpha, l)}{f(\alpha, l) / \sum_{\alpha, l} f(\alpha, l)} \;, \tag{3.27}$$

where the sum $\sum_{\alpha, l} N(\alpha, l) \approx 1.3 \cdot 10^6$, and $f(\alpha, l)$ is given in Sec. 3.5.1.

Fig. 3.16 shows the results for regions where $f(\alpha, l) > 0.5$ (see Fig. 3.12 for the surface metric). Although the data points are spread out along the whole surface, there is a clear concentration of data points around $\alpha = 0$, 90, 180, and 270 degrees (vertical and horizontal edges) and the (α, l)-values near the border of degenerate edges.

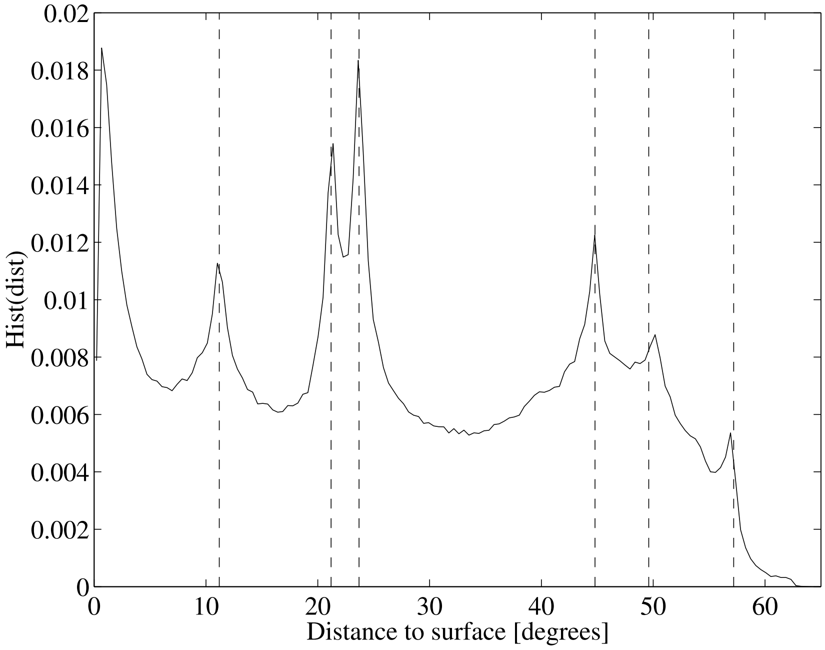


Figure 3.17: Normalized histogram of $N(\theta)$ for range patches versus the distance θ to the surface of ideal step edges. Dashed lines corresponds to binary symmetry classes $3'$, $4'$, $5'$, $14'/18'$, $25'$, and $42'$ (see Fig. 3.18).

3.5.4 Range Data Comparison: Probability Density as a Function of the Distance to the Edge Manifold

A probability density estimate of high-contrast *range* data leads to very different results around the surface of blurred step edges.

In Fig. 3.17 we show a normalized histogram $N(\theta)$ of range patches as a function of the distance θ to the surface; compare with Fig. 3.13(a), (Top), for optical patches. The histogram for range data has sharp peaks at $\theta \approx 0, 11, 22, 24, \ldots$ degrees. These peaks indicate the presence of *high-density clusters* of data points in S^7. A more detailed analysis shows that the positions of the local maxima correspond closely to the distances between binary patches and the edge manifold.

For 3×3 patches, there are 510 binary patches. These can be divided into 50 equivalence classes with respect to the 16 elements in the product group [8]

$$G = \{-1\} \times C_{4v} \ , \tag{3.28}$$

[8] The statistics of optical patches from natural images are, to a first approximation, invariant under the operations in G (Geman and Koloydenko, 1999).

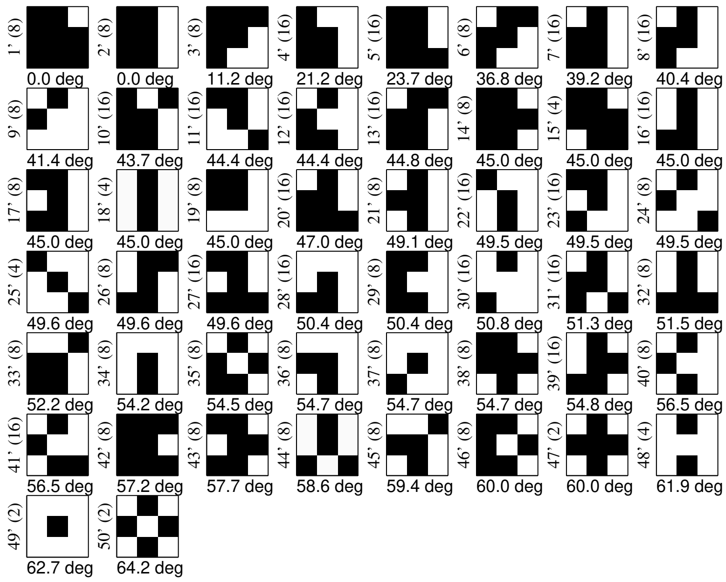


Figure 3.18: The 50 symmetry classes of binary patches arranged according to increasing θ-values, where θ is the angular distance to the closest point on the surface of ideal step edges. The number below each displayed patch represents this distance. The number in parenthesis left of each patch represents the number of binary patches in each equivalence class.

where $\{-1\}$ represents sign inversion, and the point group C_{4v} (Schönflies notation (Elliott and Dawber, 1979)) is generated by rotations through $\pi/2$ around the center pixel, and reflections across a plane containing the rotation axis.

We denote the set of 510 binary patches in $S^7 \subset \mathbb{R}^8$ by

$$\mathrm{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{510}\} \tag{3.29}$$

and the set of 50 distinct equivalence classes of the binary patches with respect to the symmetry group G by

$$\mathcal{E} = \{[\mathbf{b}_{1'}]_G, [\mathbf{b}_{2'}]_G \ldots, [\mathbf{b}_{50'}]_G\} \ . \tag{3.30}$$

We use primed indices to denote binary patches that are grouped into equivalence classes, and unprimed indices to denote the 510 original binary patches.

In Fig. 3.18, we have sorted the 50 symmetry equivalence classes according to their distances to the surface of step edges. Note that the binary patches in $[\mathbf{b}_{1'}]$ and $[\mathbf{b}_{2'}]$ are exactly on the surface. Patches in $[\mathbf{b}_{3'}]$ are at the same distance from the surface as the second local maximum at $\theta \approx 11$ (Fig. 3.17), patches in $[\mathbf{b}_{4'}]$ and $[\mathbf{b}_{5'}]$ correspond to the third and fourth local maxima at $\theta \approx 22$ and $\theta \approx 24$, respectively. The peak at $\theta \approx 45$ may be due to, for example, the symmetry classes $[\mathbf{b}_{14'}]$ (blobs) and $[\mathbf{b}_{18'}]$ (horizontal and vertical bars). The peaks at $\theta \approx 50$ and $\theta \approx 57$ could be signs of patterns similar to $[\mathbf{b}_{25'}]$ (diagonal bars) and $[\mathbf{b}_{42'}]$ (single dots), respectively. The displacement of the first peak from 0 in Fig. 3.17 (compare $[\mathbf{b}_{1'}]$ and $[\mathbf{b}_{2'}]$) may be due to the overestimate of the distance θ for points that are very close to the surface.

3.6 Range Data: Probability Density Around Binary Patches

In the previous section we saw that high-contrast range image patches are concentrated in high-density clusters both on and around the surface of step edges. Furthermore, these clusters appear to be centered around binary patches. This motivates us to investigate the density of high-contrast range patches around the 510 possible binary patches.

3.6.1 Density as a Function of Distance to Nearest Binary Patch

We start our analysis by calculating the density of the $N_{\text{tot}} = 7.9 \cdot 10^5$ high-contrast range patches as a function of the angular distance to the nearest binary patch:

For each high-contrast range patch we compute the angular distances to each of the 510 binary patches in the set $\mathrm{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{510}\}$, and find the *nearest* binary patch $\mathbf{b}_k \in B$. Let

$$N_k(\theta) = \# \left\{ n : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}_\mathrm{n}, \mathbf{b}_\mathrm{k}) < \theta + \frac{\Delta\theta}{2} \right\} , \tag{3.31}$$

where $k = 1, 2, \ldots, 510$, be the histogram of range patches $\mathbf{x}_n$ ($n = 1, 2, \ldots, N_{\text{tot}}$) that are closest to, and at a distance $\theta - \Delta\theta/2 < \phi \leq \theta + \Delta\theta/2$ from, the binary patch $\mathbf{b}_k$. The volume

$$V(\theta) = \mathrm{vol}\{\mathbf{x} \in \mathrm{S}^7 : \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{x}, \mathbf{b}_\mathrm{k}) < \theta + \frac{\Delta\theta}{2}\} \tag{3.32}$$

is given by

$$V(\theta) = \frac{16\pi^3}{15} \int_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} \sin^6(\phi)\, d\phi. \tag{3.33}$$

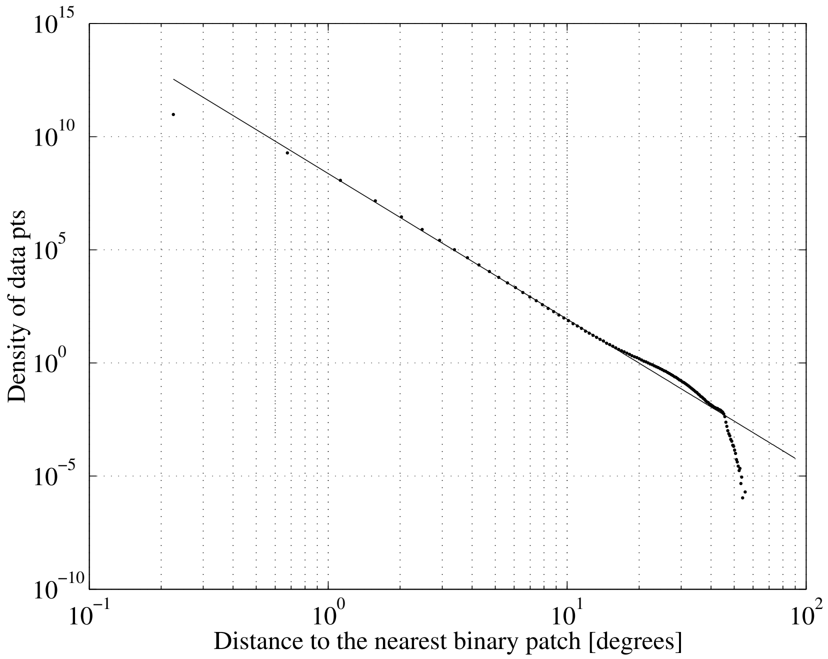


Figure 3.19: Log-log plot of the density $p(\theta)$ (Eq. (3.34)) of high-contrast range patches with respect to θ, the distance to the nearest binary patch. Linear regression in the interval 0.6 and 10 degrees gives $p(\theta) \sim \theta^{-6.4}$ (see solid line).

We define the average density of high-contrast range patches as a function of the angular distance θ to the nearest binary patch as

$$p(\theta) = \frac{\sum_{k=1}^{510} N_k(\theta)/N_{\mathrm{tot}}}{510 \cdot V(\theta)/\mathrm{vol}(\mathrm{S}^7)}, \tag{3.34}$$

where $\mathrm{vol}(\mathrm{S}^7) = \pi^4/3$. In Fig. 3.19 we show a log-log plot of $p(\theta)$. The graph is almost straight for more than a decade of distances, from 0.6 to 10 degrees. Linear regression gives that

$$p(\theta) \sim \theta^{-6.4}. \tag{3.35}$$

In Fig. 3.20 we show the cumulative percentage

$$P_{\mathrm{cum}}(\theta) = \sum_{k,\ \phi \leq \theta} \frac{N_k(\phi)}{N_{\mathrm{tot}}} \tag{3.36}$$

of the number of patches with respect to the distance θ to the nearest binary patch, as well as the cumulative volume

$$V_{\mathrm{cum}}(\theta) = \sum_{k,\ \phi \leq \theta} \frac{V(\phi)}{\mathrm{vol}(\mathrm{S}^7)} \tag{3.37}$$

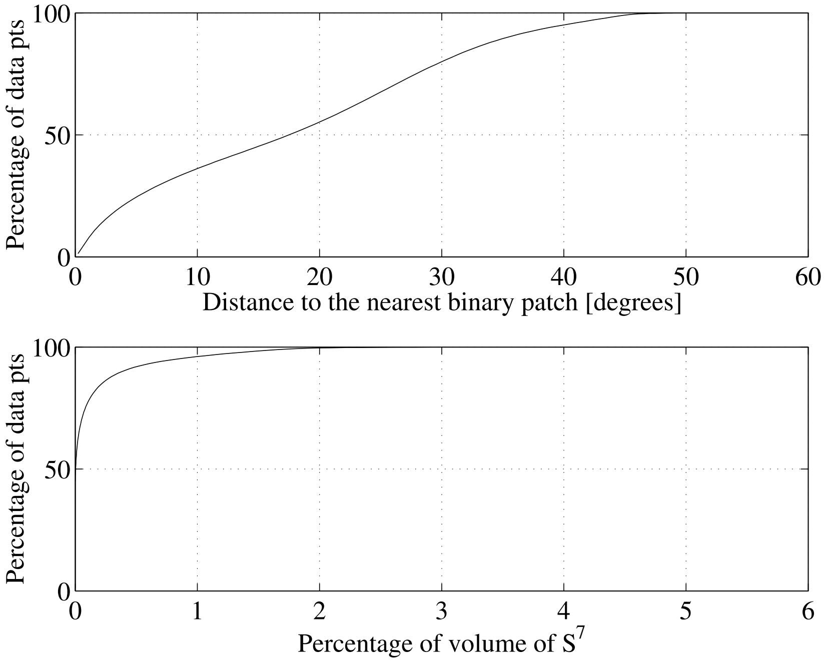


Figure 3.20: (Top) The cumulative percentage $P_{\mathrm{cum}}(\theta)$ of the number of patches $N(\theta)$ with respect to the distance θ to the nearest binary patch. (Bottom) The cumulative volume $V_{\mathrm{cum}}(\theta)$ versus $P_{\mathrm{cum}}(\theta)$.

versus $P_{\mathrm{cum}}(\theta)$. We see that 80% of the high-contrast range patches are within a spherical neighborhood of 30 degrees from one of the 510 binary patches. The neighborhoods of these 80% patches occupy only 0.14% of the total volume of S^7.

These results show that 3×3 high-contrast range patches are densely clustered around the 510 binary patches, and that the probability density is infinite at the positions of these binary patches.

3.6.2 Distribution of Range Patches Across the 50 Binary Symmetry Classes

We end this paper by studying how the range patches are distributed among the 50 symmetry classes for binary patches defined in Sec. 3.5.4. This will give us an idea of the typical geometrical patterns for high-contrast range patches in image space.

As before, we find the binary patches that are nearest to the data points $\mathbf{x}_1 \ldots \mathbf{x}_{N_{\mathrm{tot}}}$. We then group the data points together depending on which equivalence classes (with respect to the symmetry group G) the closest binary patches belong to. We define the number of range patches in symmetry class $[\mathbf{b}_{k'}]_G$ (as a function of the distance θ to the nearest binary patch) by

$$N_{k'}(\theta) = \sum_{j \text{ s.t. } \mathbf{b}_j \in [\mathbf{b}_{k'}]_G} N_j(\theta) \ , \tag{3.38}$$

where N_j is given in Eq. (3.31). Furthermore, the density of range patches in the symmetry class $[\mathbf{b}_{k'}]_G$ is defined by

$$p_{k'}(\theta) = \frac{N_{k'}(\theta)/N_{\text{tot}}}{\text{size}([\mathbf{b}_{k'}]_G) \cdot V(\theta)/\text{vol}(\mathrm{S}^7)} \ , \tag{3.39}$$

where $V(\theta)$ is given by Eq. (3.33), and size$([\mathbf{b}_{k'}]_G)$ is the number of equivalent patches in the class $[\mathbf{b}_{k'}]_G$.

Fig. 3.21 displays the 50 symmetry classes of binary patches ordered after the percentage of range patches $P_{k'} = \sum_\theta N_{k'}(\theta)/N_{\text{tot}}$ in each class. For each symmetry class $[\mathbf{b}_{k'}]_G$, we show the binary patch $\mathbf{b}_j \in [\mathbf{b}_{k'}]_G$ that is most common among the range patches. The figure also shows the cumulative percentage $P_{k',\text{cum}} = \sum_{j' \leq k'} P_{k'}$. From these numbers we conclude that most high-contrast range patches cluster around binary patches that belong to only a few of the 50 symmetry classes; in fact, 70% of the patches are closest to the first 7 symmetry classes. The most common structures among high-contrast patches are horizontal and vertical edges followed by slanted edges, corner- and T-junction–like structures. The least probable structures are checkerboard and cross patches. Fig. 3.21 agrees with our Voronoi results for range patches (Fig. 3.8), as the patterns of the 25 most frequent Voronoi cells resemble the patterns of the patches in the 5 most frequent binary symmetry classes.

The graphs of the densities $p_{k'}(\theta)$ (Eq. (3.39)) for the 50 symmetry classes are similar in appearance to Fig. 3.19. In Fig. 3.22, we show the slopes obtained by linear regression in a log-log plot of $p_{k'}(\theta)$ ordered after decreasing cumulative percentage $P_{k',\text{cum}}$. The most frequent symmetry groups have very steep density curves, and there is a gradual decrease in the slopes of the curves for the less frequent symmetry classes. This is consistent with the result that most patches are close to binary patches which belong to the 7 most frequent classes.

3.7 Summary and Conclusions

In this work, we have taken a somewhat different approach to natural image statistics. Most of the work in image statistics focuses either on modeling

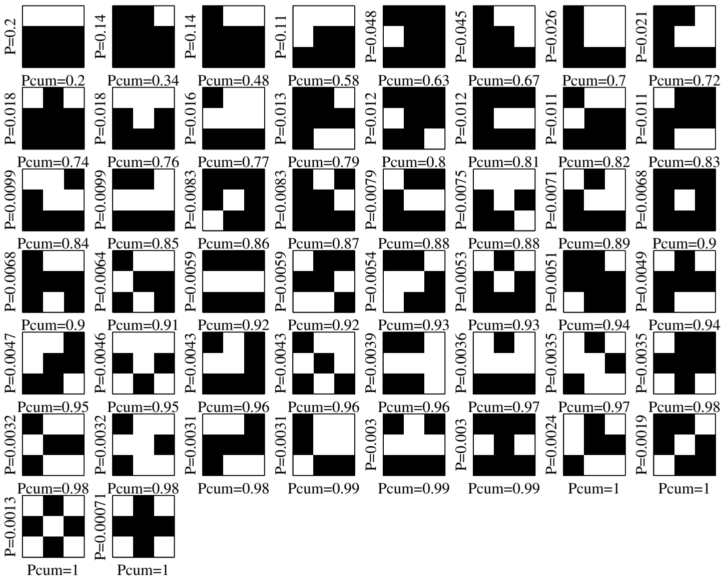


Figure 3.21: The 50 symmetry classes of binary patches ordered after the percentage $P_{k'}$ of high-contrast range patches closest to one of the binary patches in the equivalence class $[\mathbf{b}_{k'}]_G$. The cumulative percentage $P_{k',\mathrm{cum}}$ is also indicated. For each class, we display the binary patch which is most common among the high-contrast range patches.

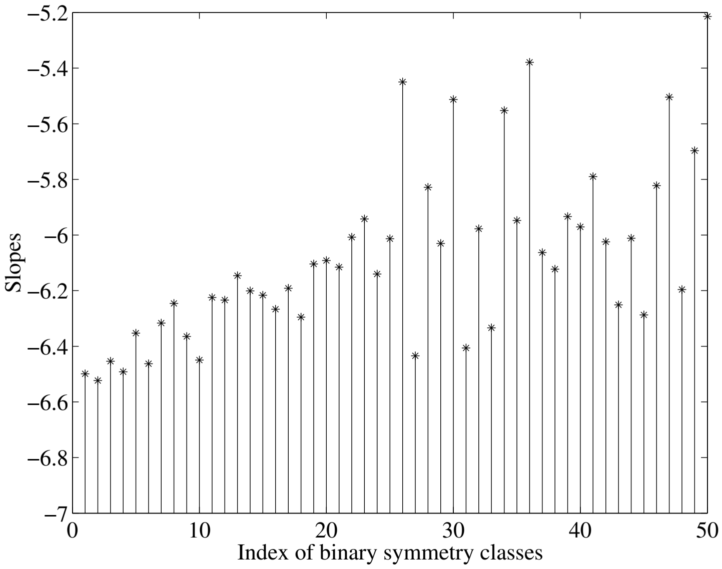


Figure 3.22: Slopes in a log-log plot of the densities $p_{k'}(\theta)$ for each of the 50 binary symmetry classes $[\mathbf{b}_{k'}]_G$ ordered after decreasing cumulative percentage $P_{k',\mathrm{cum}}$.

1D or 2D histograms of linear filter reactions or on finding a linear change of basis that sparsifies the data. Few attempts have been made to understand the *full* probability distribution of natural images and the *intrinsic dimension* of the state space of generic image data. It seems that we cannot take full advantage of the sparseness of the data state space without this knowledge of natural images.

In this study, we have analyzed the local geometric patterns seen in *generic* images. We believe that simple geometric structures in the world and the sensor properties of the probing device generate observations that are concentrated along predictable shapes (manifolds or clusters) in state space. The basic vocabulary of images (with edges, bars, blobs, terminations etc.) seems to be the same — whether one studies all types of natural images as here, or specific classes of images; e.g. medical images or images of just trees, indoor scenes etc.

Optical and range images measure different aspects of generators (objects) in the world; scene luminance versus distances to the nearest objects in a natural scene, respectively. Thus, there are bound to be differences in the statistics of these two types of images. However, we believe that the main qualitative differences between optical and range images are due to differences in sub-resolution properties. It seems that basic primitives (such as edges) in the world and *morphological* or ordering filters (such as median and mean filters) lead to *compact clusters* in state space. On the other hand, the same primitives and *averaging* filters lead to *continuous submanifolds* in state space.

In this paper, we have analyzed the probability distribution of 3×3 high-contrast patches from natural images of different modalities (optical versus 3D range images). In the preprocessing stage, we subtracted the mean and contrast-normalized the log-values of each image patch. The state space of the preprocessed image data (from optical or range images) is a 7-dimensional unit sphere in $\mathbb{R}^8$.

As a first exploration, we examined how the data distribute with respect to an approximately uniform Voronoi tessellation of the 7-sphere. The analysis showed that both optical and range patches occupy a very small amount of the total surface area (volume) of the state space: In both cases, half of the data can be divided into a set of Voronoi cells with a total volume of less than 6% of the volume of the 7-sphere. For optical patches, the centers of the most densely populated Voronoi cells resemble blurred step edges. For range patches, they resemble binary patterns.

A more detailed analysis showed clear differences in the probability distributions of optical and range patches. The majority of high-contrast optical patches are concentrated around a 2-dimensional C^1 *submanifold* embedded in the 7-dimensional sphere. This surface is highly *non-linear* and corresponds to ideal step edges blurred by the optics of the camera. A density calculation showed that the probability density of optical patches is infinite on this ideal surface. About 50% of the optical data points are within a tubular neighborhood of the surface with a width that corresponds to only 9% of the total volume of the state space.

The majority of range patches, on the other hand, are concentrated in *compact clusters*, rather than on a smooth manifold. The centers of these clusters seem to correspond to binary patches, i.e. patches with only two range values. A density calculation around the 510 possible binary patches indicated an infinite probability density at these "hot spots" of the image space. About 80% of the high-contrast patches are in a neighborhood of these spots that correspond to only 0.14% of the total volume of the 7-dimensional unit sphere. The most frequent binary patches are horizontal edges followed by slanted edges and corner- and T-junction–like structures.

Although the analysis in the current paper only deals with 3×3-patches, we believe that the more general results apply to *larger* patches and even *general* filter responses. The picture that seems to emerge is that basic image primitives — such as edges, blobs, and bars — generate low-dimensional and (in general) non-linear structures in the state space of image data. Therefore, while the dimension of the state space, determined by the number of filters or pixels in the analysis, is usually very large, the intrinsic dimension of the manifolds generated by different primitives is *fixed* and determined by the complexity of the primitives only. The edge manifold we found for optical data is continuous and 2-dimensional. This is because an ideal edge can be characterized by two parameters: the orientation α and the position l of the edge. For optical data and bar structures which can be parameterized by 3 parameters (the orientation, position, and width of a bar), we would expect a 3-dimensional submanifold in state space, regardless of the dimension of the state space.

More generally, we believe that one can define a dictionary of probability models on representations of *general* primitives parameterized by $\Phi = \{\phi_1, \phi_2, \ldots\}$ for *any* set of filters $f_1, \ldots, f_N$. In the N-dimensional state space of the filter-based image representations, the image primitives will define manifolds of the general form

$$M(\Phi) = [f_1(\cdot) * I(\cdot; \Phi), \ldots, f_N(\cdot) * I(\cdot; \Phi)]^T ,$$

where $*$ denotes a convolution.

Our empirical results for the edge manifold and binary patches of optical and range data, respectively, show that, when studying natural image statistics, important geometric and probabilistic structures emerge only after abandoning assumptions such as independent components or sparse coding by a linear change of basis. Furthermore, when looking at the full probability distribution of small patches, clear differences appear between different types of image modalities (optical versus range) that otherwise have seemingly similar statistics. Analyzing 3×3 patches could thus offer a systematic and precise way of distinguishing and comparing image data. A complete description of the probability distribution of natural image patches, however, also requires modeling low-contrast patches, and high-density regions that lie outside the ideal manifold of step edges (for optical patches) and binary clusters (for range patches). The Voronoi tessellation in Sec. 3.4 offers an automatic way of characterizing the full state space of 3×3 pixel data, while the "geometry-based" methods in Sec. 3.5 and Sec. 3.6 may lead to a better understanding and parametric probability models of natural image data.

Acknowledgments

The authors would like to thank Professor Peter Johansen and Professor Mads Nielsen for their valuable comments. This work was supported by ARO Muri grant DAAH04-96-1-0445.

Chapter 4

Toward a Full Probability Model of Edges in Natural Images

4.1 Introduction

The study of natural image statistics is an active research area and different approaches have been taken in this field (Field, 1987; Huang et al., 2000; Huang and Mumford, 1999; Lee et al., 2001; Lee et al., 2002; Ruderman and Bialek, 1994; Simoncelli, 1999b; Wainwright and Simoncelli, 2000; Zetzsche et al., 1993). It has previously been shown that natural image statistics — such as the marginal distributions of image intensity I and the gradient magnitude $\|\nabla I\|$ — are highly non-Gaussian, and approximately invariant to changes in scale (Field, 1987; Mallat, 1989; Ruderman and Bialek, 1994). Roughly speaking, the research in natural image statistics can be divided into two related directions: Researchers such as Zetzsche et al. (1993), Simoncelli (1999b), and Huang et al. (2000; 1999) have looked at the 1D marginal and 2D joint statistics of filter responses for a fixed wavelet basis. They have, for example, explored complex dependencies between pairs of wavelet coefficients at nearby spatial positions, orientations and scales. Others have looked at the state space of image data and tried to find a set of directions (or projections of the data) that lead to an optimal image representation in some sense; see e.g. sparse coding (Olshausen and Field, 1996) and ICA (Bell and Sejnowski, 1997; Hyvärinen, 1999).

In this paper, we take a different approach to natural image statistics. We believe that in order to fully understand the statistics of natural images one needs to explore the *full probability distribution* of the salient structures of local image patches. The analysis is free from such restrictive assumptions as independent components, or even linear decompositions of an image into basis images.

Our work is inspired by David Marr (Marr, 1982) who proposed that the structure of images can be described by primitives such as edges, bars, blobs and terminations — the so-called "primal sketch". The basic questions we ask are: *What are the probability distributions of Marr's primitives and how are these primitives represented geometrically in the state space of image data? Can we develop models that will tell us how likely we are to observe a local geometric structure (such as an edge, ridge, blob, corner etc.) of a certain spatial extension in an image?* Not much work has been done in this direction to our knowledge. Such probability models would, however, be useful as a priori knowledge in image processing and computer vision applications as diverse as feature detection (Desolneux et al., 2001), segmentation (Malik et al., 2001) and enhancement (Simoncelli, 1999a; Zhu and Mumford, 1998). In particular, probability models on features could be used as prior distributions on features in image coding and reconstruction. Nielsen and Lillholm (2001) have, for example, suggested that image reconstruction can be done by solving a variational optimization problem constrained by localized feature measurements and a prior distribution on features.

To study the distribution of local geometric image structures, one first has to choose a representation that captures the image geometry in a neighborhood of some fiducial point. We will use a representation based on a set of local measurements of the luminance captured through a set of sensors — a sensorium. The concept of a sensorium makes sense both from a biological vision point of view (the receptive fields in the early visual system have been compared to feature detectors (Koenderink and van Doorn, 1987)) and from a mathematical point of view: Florack (1997) and Mumford and Gidas (2001) have both argued that an image I is not a function or a point-wise estimate $I(x, y)$ but a Schwartz distribution that can only be probed by averaging (or measuring), $\int\int I(\xi, \eta)\phi_j(x-\xi, y-\eta)d\xi d\eta$, through smooth "test" functions or sensors ϕ_j.

In our previous study (Chapter 3 and Lee et al. (2001; 2002)) of natural image statistics, the sensorium was defined by the sensors in the CCD camera used to collect the images. More specifically, we studied the joint statistics of the *pixel* intensity values in high-contrast 3×3 natural image patches. We found (for optical images) that the state space of the patch data is extremely sparse, with most of the data concentrated around a continuous non-linear manifold in state space. This manifold corresponds to edges of different orientations α and positions l.

In this work, we investigate whether our previous results (such as the existence of an ideal edge manifold in state space and the concentration of natural image data around this manifold) generalize to different scales and more general image representations.

In the context of linear Gaussian scale-space theory (Koenderink, 1984), Koenderink and van Doorn (Koenderink and van Doorn, 1987) proposed the so-called *local jet* of an image as a biologically plausible representation of local image geometry. In this setting, the sensorium consists of partial derivatives of the Gaussian kernel function. Convolving an image with these kernels is equivalent to measuring the partial derivatives of a coarse-grained representation of the image — the so-called scale-space image. The jet space captures the local geometry in a neighborhood of a point in the image, where the size of the neighborhood is determined by the standard deviation or *scale* of the Gaussian kernel.

In this work, we choose an image representation defined by Gaussian scale-space image derivatives up to third order — *the 3-jet.* This sensorium can distinguish between image structures such as edges, ridges, blobs and corners (Koenderink and van Doorn, 1987), but is blind to structures that require the descriptive power of image derivatives of order higher than 3.

We believe that such a representation of image data has certain advantages compared to many other types of multi-scale representations. First of all, the Gaussian scale-space representation gives us a sensible way of defining image derivatives — the scale-space image derivatives. With these derivatives, we can use the language of differential geometry to define and interpret local features in images. The Gaussian kernel and its derivatives are furthermore similar to the receptive fields found in the human visual system (Koenderink, 1984; Koenderink and van Doorn, 1987). Both the human receptive fields and the Gaussian scale-space derivatives are tuned to structures at different scales, orientations and spatial frequencies.

As in Lee et al. (2001; 2002), we focus our analysis on edge structures. We first define a model of an ideal edge in scale-space, and show that the 3-jet representations of edges define a 2-dimensional differentiable manifold in jet space. We then study how empirical data, extracted from a large database of natural images, are distributed in 3-jet space with respect to this manifold. We find, in accordance with previous results in Lee et al. (2001; 2002), that the natural image data are densely distributed around the edge manifold with a probability density function of the form $\theta^{-\gamma}$, where θ is the distance to the edge manifold and γ is close to 0.7. Furthermore, we show that the results are approximately invariant to a change of scale.

This work is an attempt to develop a *full probability model of edges* in natural images that is universal, i.e. independent of scale and image representation. In the future, we plan to extend the analysis to representations of other image primitives (such as bars, blobs, and T-junctions).

The organization of the paper is as follows: In Sec. 4.2 we provide the necessary background on jet space and linear Gaussian scale-space theory. We introduce the Gaussian edge model in Sec. 4.3, and in Sec. 4.4, we describe our image data set, the whitening and contrast normalization of this data, and the results of our analysis. Finally, we finish with concluding remarks (Sec. 4.5).

4.2 Multi-scale Local Jet

Linear Gaussian scale-space was proposed among others by Koenderink (1984) as a sound theoretical framework for doing multi-scale image analysis. The Gaussian scale-space $L : \Omega \mapsto \mathbb{R}$ of an image $I : \Omega \mapsto \mathbb{R}$ (where $\Omega \subseteq \mathbb{R}^2$) can be defined as the solution to the heat diffusion equation

$$\frac{\partial L}{\partial t} = \frac{1}{2}\nabla^2 L \tag{4.1}$$

under the constraint $L(x, y; s = 0) = I(x, y)$. The *scale* $s \geq 0$ is related to t by $t = s^2$. The Gaussian scale-space representation of an image is thus a convolution

$$L(x, y; s) = \iint_\Omega I(\xi, \eta)\phi(x - \xi, y - \eta; s)\, d\xi\, d\eta \tag{4.2}$$

with a Gaussian kernel function $\phi : \mathbb{R}^2 \mapsto \mathbb{R}$ where

$$\phi(x, y; s) = \frac{1}{2\pi s^2} e^{-\frac{(x^2+y^2)}{2s^2}} \ . \tag{4.3}$$

We interpret the parameter s as the measurement scale of the scale-space image $L(x, y; s)$ as it corresponds to the width (or standard deviation) of the "smoothing filter" $\phi(\cdot)$.

An image is in general not a differentiable function, but in scale-space we obtain a family of smoothed versions of the image which are C^∞-differentiable. We can compute partial derivatives $\partial_{x^n}\partial_{y^m} (\equiv \frac{\partial^{n+m}}{\partial x^n \partial y^m})$ of the scale-space representation by convolving the image $I(\cdot)$ with partial derivatives of the Gaussian kernel function $\phi(\cdot; s)$, as

$$L_{x^n y^m}(\cdot; s) \equiv \partial_{x^n}\partial_{y^m}(I * \phi) = I * (\partial_{x^n}\partial_{y^m}\phi) \ . \tag{4.4}$$

Note that the scale-space derivatives $L_{x^n y^m}$ constitute a scale-space as they also satisfy the heat diffusion equation Eq. (4.1).

Various approaches exist for discretization of scale-space representations (see e.g. Lindeberg (1994b)). Here we evaluate the convolution in Eq. (4.2) by multiplying the discrete Fourier transform of the discrete image with a discretized Fourier transform of the Gaussian derivatives $\partial_{x^n}\partial_{y^m}\phi$.

To be able to compare scale-space derivatives at different scales, it is convenient to use dimensionless coordinates (the so-called natural coordinates) and scale-normalized differential operators (Lindeberg, 1994b). In the (x, y)-coordinate system the dimensionless coordinates are given by

$$(x', y') = \left(\frac{x}{s}, \frac{y}{s}\right) \tag{4.5}$$

and the scale-normalized partial derivatives are

$$L_{x'^n y'^m}(x, y; s) = \partial_{x'^n}\partial_{y'^m}L(x, y; s) = s^{n+m}\partial_{x^n}\partial_{y^m}L(x, y; s)\ . \tag{4.6}$$

In the rest of this paper we will assume that *all* scale-space derivatives $L_{x^n y^m}$ are scale normalized.

We can describe the local geometry of an image by the so-called local jet (Florack et al., 1996; Koenderink and van Doorn, 1987). Since the scale-space image $L(x, y; s)$ is a smoothed, differentiable version of the image, we can use a Taylor series to describe the geometry of the image intensity surface around a point (x_0, y_0). For $(x_0, y_0) = (0, 0)$, for example, we have

$$\begin{aligned} L(x, y; s) = L + L_x x + L_y y + \frac{1}{2}(L_{xx}x^2 + 2L_{xy}xy + L_{yy}y^2) \\ + \frac{1}{6}(L_{xxx}x^3 + 3L_{xxy}x^2y + 3L_{xyy}xy^2 + L_{yyy}y^3) + \ldots \end{aligned} \tag{4.7}$$

where the scale-space derivatives $L, L_x, L_y, \ldots$ are evaluated at $(x_0, y_0) = (0, 0)$. Consider now the truncated Taylor expansion of degree k. The so-called *local k-jet* (of $L(x, y; s)$ at (x_0, y_0)) is an equivalence class of smooth functions with respect to the map $j^kL : \mathbb{R}^2 \mapsto \mathcal{J}^k(\mathbb{R}^2 \mapsto \mathbb{R}) \subset \mathbb{R}^N$, $N = (2 + k)!/(2k!)$, where

$$j^kL(x, y; s) = (L(x, y; s), L_x(x, y; s), L_y(x, y; s), \ldots, L_{x^n y^m}(x, y; s))^T \tag{4.8}$$

and $n + m = k$. The space $\mathcal{J}^k(\mathbb{R}^2 \mapsto \mathbb{R})$ of all k-jets of functions $\mathbb{R}^2 \mapsto \mathbb{R}$ is sometimes called a *k-jet space*. Local images that belong to the same k-jet (i.e. the same point in $\mathcal{J}^k(\mathbb{R}^2 \mapsto \mathbb{R})$) "look" the same, up to order k, in the sense that we can not distinguish between them by only looking at scale-space derivatives up to order k. Koenderink and van Doorn (1996) named this class a metamer inspired by the terminology of Schrödinger's theory of colorimetry.

We limit our analysis to the partial derivatives parameterizing the 3-jet, as the 3-jet captures the characteristics of common geometric structures such as edges, ridges, blobs, and corners (Koenderink and van Doorn, 1987; Koenderink and van Doorn, 1996). In the following, we will study the statistics of images mapped into 3-jet space by $\tilde{j}^3 L : \mathbb{R}^2 \mapsto \tilde{\mathcal{J}}^3(\mathbb{R}^2 \mapsto \mathbb{R})$, where

$$\tilde{j}^3 L(x,y;s) = (L_x, L_y, L_{xx}, L_{xy}, L_{yy}, L_{xxx}, L_{xxy}, L_{xyy}, L_{yyy})^T \tag{4.9}$$

and the measurement scale $s > 0$. The scale-space derivatives $L_x, L_y, L_{xx}, \ldots$ are evaluated at $(x, y; s)$, and are *scale normalized* according to Eq. (4.6), i.e. $L_{x^n y^m}(x,y;s) = s^{n+m} \partial_{x^n} \partial_{y^m} L(x,y;s)$. In the above 3-jet representation, we have excluded the intensity $L(x,y;s)$ of the blurred image, as we are only interested in variations in the local image geometry. The tilde notation is to indicate that $\tilde{\mathcal{J}}^3(\mathbb{R}^2 \mapsto \mathbb{R}) \subset \mathcal{J}^3(\mathbb{R}^2 \mapsto \mathbb{R})$.

4.3 The Edge Manifold

We will investigate a simple model of edges mapped into jet space. We model edges by the scale-space of an ideal step edge (step edges blurred with a Gaussian — a Gaussian edge). In this section, we show that the 3-jet representations of blurred step edges of different orientations, positions and scales trace out a differentiable 2D submanifold in the jet space $\tilde{\mathcal{J}}^3$.

For convenience, we define the edge model in the local orthonormal (u, v)-coordinate system where the v-axis has the direction of the local gradient at any point P_0 and the u-axis is perpendicular, i.e. we define unit vectors

$$e_v = (\cos\alpha, \sin\alpha)^T = \frac{1}{\sqrt{L_x^2 + L_y^2}} (L_x, L_y)^T \Big|_{P_0} \tag{4.10}$$

and

$$e_u = (\sin\alpha, -\cos\alpha)^T \; . \tag{4.11}$$

In this coordinate system, an ideal step edge (defined on $\mathbb{R}^2$) has the form

$$f(u,v;l) = \begin{cases} 1 & \text{if } v \geq l \\ 0 & \text{if } v < l \end{cases} \tag{4.12}$$

where $l \in \mathbb{R}$ is the displacement of the edge from the origin in the v-direction. The scale-space representation of the ideal step edge is (according to Eq. (4.2)) given by

$$G(u,v;l,s) = f(u,v;l) * \phi(u,v;s) = \int_{v'=-\infty}^{v} \psi(v';l,s)\, dv' \tag{4.13}$$

where $\psi(v; l, s) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(v-l)^2}{2s^2}}$ is a one-dimensional Gaussian kernel centered at l. The scale-normalized partial derivatives of the edge model $G(u, v; l, s)$ along the u and v-axes are

$$\begin{aligned} G_{u^n}(u, v; l, s) &= 0 \\ G_{v^n}(u, v; l, s) &= s^n \partial_{v^{n-1}} \psi(v; l, s) \end{aligned} \tag{4.14}$$

for $n \geq 1$.

We now map the edge model $G(u, v; \alpha, l, s)$ into the jet space $\tilde{\mathcal{J}}^3$ by computing the nine components of the map $\tilde{j}^3 G(0, 0; \alpha, l, s)$ defined by Eq. (4.9). Since

$$\begin{aligned} \partial_x &= \cos\alpha\, \partial_v + \sin\alpha\, \partial_u \\ \partial_y &= \sin\alpha\, \partial_v - \cos\alpha\, \partial_u \ , \end{aligned} \tag{4.15}$$

we get that

$$\begin{aligned} G_{x^m, y^n}(0, 0; \alpha, l, s) &= \cos^m \alpha \ \sin^n \alpha\, G_{v^{m+n}}(u, v; l, s)\Big|_{(u,v)=(0,0)} \\ &= s^{m+n} \cos^m \alpha \ \sin^n \alpha\, \partial_{v^{m+n-1}} \psi(v; l, s)\Big|_{v=0} \ . \end{aligned} \tag{4.16}$$

Denote the map that takes the edge model to the 3-jet space $\tilde{\mathcal{J}}^3$ by $\mathcal{E} : [0, 2\pi) \times \mathbb{R} \times \mathbb{R}_+ \backslash \{0\} \mapsto \tilde{\mathcal{J}}^3$, where

$$\begin{aligned} \mathcal{E}(\alpha, l, s) = (&G_x(0, 0; \alpha, l, s), G_y(0, 0; \alpha, l, s), G_{xx}(0, 0; \alpha, l, s), \\ &G_{xy}(0, 0; \alpha, l, s), G_{yy}(0, 0; \alpha, l, s), G_{xxx}(0, 0; \alpha, l, s), \\ &G_{xxy}(0, 0; \alpha, l, s), G_{xyy}(0, 0; \alpha, l, s), G_{yyy}(0, 0; \alpha, l, s))^T \ . \end{aligned} \tag{4.17}$$

Although the edge map $\mathcal{E}$ is a function of three variables (the angle α, the displacement l and the scale s), the loci of all points $\mathcal{E}(\alpha, l, s)$ trace out a *2-dimensional* C^∞ differentiable manifold in $\mathbb{R}^9$ that only depends on α and the ratio l/s (see Appendix 4.6). Note that the edge manifold is periodic in α for fixed l/s ratio.

4.4 Statistics of Edge Structures

4.4.1 The Empirical Data Set

In our experiments, we use the van Hateren and van der Schaaf (1998) still image collection consisting of 4167 1020×1532 pixels gray-scale images[1] (see

[1] We use the raw image set (.iml) where the intensity values have been linearized by the camera's lookup table.

Figure 4.1: Sample images from the van Hateren still image collection. We show the log-transformed intensity values, $\log(I(x,y)+1)$.

Fig. 4.1 for samples from the database). Before doing any processing of the images $I(x,y)$ in the database, we compress the intensity range by taking the logarithm $\log(I(x,y)+1)$ of the intensity.

From each scale-space image $L^{(i)}(x,y;s)$ at a fixed scale s, where $s = 1,2,4,8,16,32$ and $i = 1,\ldots,4167$, we extract a random set of 1000 spatial coordinates $X^{(i,s)} \subseteq \Omega$ ($\Omega \subseteq \mathbb{R}^2$ denotes the image domain). At these spatial coordinates, we compute the 3-jet representation defined according to Eq. (4.9). This gives us data sets

$$J_s = \left\{ \tilde{j}^3 L^{(i)}(x,y;s) \subseteq \tilde{\mathcal{J}}^3 \,\middle|\, (x,y) \in X^{(i,s)}; i = 1,\ldots,4167 \right\} \tag{4.18}$$

where $s = 1,2,4,8,16,32$. Elements in each set J_s are points in $\tilde{\mathcal{J}}^3$ that have been sampled from different spatial positions and different images at fixed scale s. The total number of data points[2] in each set J_s is $|J_s| \approx 4.1 \cdot 10^6$.

4.4.2 Whitening and Contrast Normalization

The lighting conditions may vary across and between images. We are interested in variations in the local *geometry* of the image and would like to disregard variations caused by changing lighting.

Before contrast-normalizing we first whiten the data. This will lead to a vector representation of the 3-jets where the elements are uncorrelated and of the same order of magnitude.

[2]To prevent numerical problems during contrast normalization (see Sec. 4.4.2 and Eq. (4.20)), we discard data points $\mathbf{y}$ with a norm $\|\mathbf{y}\|$ that is close to zero after whitening. This corresponds to 1% of all data points.

Assume that $\mathbf{x} \in J_s$ where J_s is our data set (Eq. (4.18)). The covariance or correlation matrix $C =< \mathbf{x}\mathbf{x}^T >$ is scale invariant (see Chapter 5 for details), so we can get a robust estimate of C from the joined data set $\mathbf{x} \in \bigcup_s J_s$ where s denotes the scale. The mean $< \mathbf{x} >= \mathbf{0}$ due to the convolution with zero mean scale-space filters.

The first step in the data preprocessing is to define transformed input variables

$$\mathbf{y} = \Lambda^{-1/2} U^T \mathbf{x} , \tag{4.19}$$

where U is a 9×9-matrix with the normalized eigenvectors of C as columns, and Λ is a diagonal 9×9-matrix with the corresponding eigenvalues of C as diagonal elements. The transformed data $\mathbf{y}$ is "white" in the sense that the covariance matrix $< \mathbf{y}\mathbf{y}^T >= \mathbf{1}$.

The second step is to contrast-normalize the data according to

$$\hat{\mathbf{p}} = \frac{\mathbf{y}}{\|\mathbf{y}\|} \tag{4.20}$$

so that scale-space images of similar geometric structure have the same representation. The whitened and contrast-normalized data points $\hat{\mathbf{p}}$ all lie on a 8-dimensional unit sphere

$$\mathrm{S}^8 \equiv \{\hat{\mathbf{p}} | \; \|\hat{\mathbf{p}}\| = 1\} \subset \mathbb{R}^9 . \tag{4.21}$$

The 8-sphere S^8 is the *state space of whitened and contrast-normalized* 3-jet representations. The whitened and contrast-normalized data set J_s at a fixed scale s is denoted by $\hat{J}_s \in \mathrm{S}^8$.

Similarly, we define the map $\hat{\mathcal{E}} : [0, 2\pi) \times \mathbb{R} \mapsto \mathrm{S}^8$ that takes the edge map $\mathcal{E}(\alpha, l, s)$ (Eq. (4.17)) to the state space of whitened and contrast normalized 3-jet representations by

$$\hat{\mathcal{E}}(\alpha, l/s) = \frac{\Lambda^{-1/2} U^T \mathcal{E}(\alpha, l/s, 1)}{\|\Lambda^{-1/2} U^T \mathcal{E}(\alpha, l/s, 1)\|} . \tag{4.22}$$

We measure the *distance* between two data points $\hat{\mathbf{p}}_0, \hat{\mathbf{p}}_1 \in \mathrm{S}^8$ on the 8-sphere by their angular separation, i.e.

$$\mathrm{dist}(\hat{\mathbf{p}}_0, \hat{\mathbf{p}}_1) \equiv \arccos(\hat{\mathbf{p}}_0^T \hat{\mathbf{p}}_1) . \tag{4.23}$$

4.4.3 Empirical Density Results

In Sec. 4.3, we described the theoretical manifold of edges in 3-jet space. In this section, we verify that the empirical data from natural images (i.e.

the whitened and contrast-normalized data in sets $\hat{J}_s$ from Sec. 4.4.2) are really densely distributed around the manifold of edges. By putting parallel bins around the manifold and computing histograms of the data, we can get an estimate of the functional form of the probability density around the manifold.

First we divide the whitened and contrast-normalized edge manifold $\hat{\mathcal{E}}(\alpha, l/s)$ of Eq. (4.22) into a mesh of spherical triangles in the same fashion as described in Chapter 3 and Lee et al. (2001). We sample the edge manifold with parameters $l \in [-4s, 4s]$ and $\alpha \in [0, 2\pi)$. We refine the mesh of triangles until no vertices in a triangle are more than 11 degrees apart. This gives us a triangulated mesh with a total of 22944 triangles. We use the triangulated mesh of the manifold to estimate the distances between the data points of $\hat{J}_s$ and the manifold of edges $\hat{\mathcal{E}}(\alpha, l/s)$ on the 8-sphere S^8. The distance $\mathrm{dist}(\mathbf{x}, \hat{\mathcal{E}}(\alpha, l/s))$ between a data point $\mathbf{x} \in \hat{J}_s$ and the edge manifold $\hat{\mathcal{E}}(\alpha, l/s)$ is approximated by the distance to the center point of the closest triangle in the mesh.

Fig. 4.2 (top) shows a normalized histogram of the number of whitened and contrast-normalized data points $\hat{\mathbf{p}}^s_n \in \hat{J}_s$ ($n = 1, \ldots, |\hat{J}_s|$) versus the distance θ to the edge manifold $\hat{\mathcal{E}}(\alpha, l/s) \subset \mathrm{S}^8$. Let

$$N(\theta; s) = \# \left\{ n \,\middle|\, \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\hat{\mathbf{p}}^s_n, \hat{\mathcal{E}}(\alpha, l/s)) < \theta + \frac{\Delta\theta}{2} \right\} \tag{4.24}$$

where $\Delta\theta$ is the histogram bin width and $\mathrm{dist}(\hat{\mathbf{p}}^s_n, \hat{\mathcal{E}}(\alpha, l/s))$ is the angular distance (Eq. (4.23)) from the data point $\hat{\mathbf{p}}^s_n$ to the closest point on the triangulated mesh of the edge manifold $\hat{\mathcal{E}}(\alpha, l/s)$.

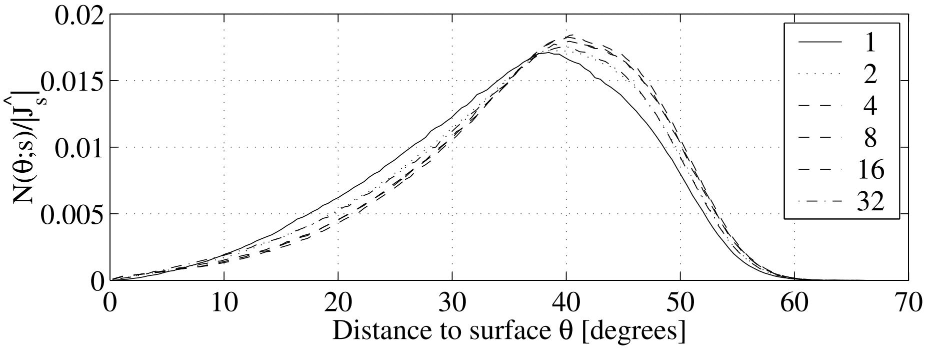
To get an estimate of the probability density of points around the edge manifold, we also need to calculate the volume of the bins $[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2})$ in the state space S^8. We here estimate the bin volume by sampling $V_{\mathrm{tot}} = 10^7$ uniformly randomly distributed points $\mathbf{v}_n$ ($n = 1, \ldots, V_{\mathrm{tot}}$) on the 8-sphere. The histogram

$$V(\theta) = \# \left\{ n \,\middle|\, \theta - \frac{\Delta\theta}{2} \leq \mathrm{dist}(\mathbf{v}_n, \hat{\mathcal{E}}(\alpha, l/s)) < \theta + \frac{\Delta\theta}{2} \right\} \tag{4.25}$$

of the number of samples versus the distance θ to the surface of edges is a Monte Carlo estimate of the volume of the histogram bins. Fig. 4.2 (bottom) shows the normalized histogram $V(\theta)/V_{\mathrm{tot}}$.

We define the empirical *density* of data points around the edge manifold as

$$\rho(\theta; s) = \frac{N(\theta; s)/|\hat{J}_s|}{V(\theta)/V_{\mathrm{tot}}} \,. \tag{4.26}$$



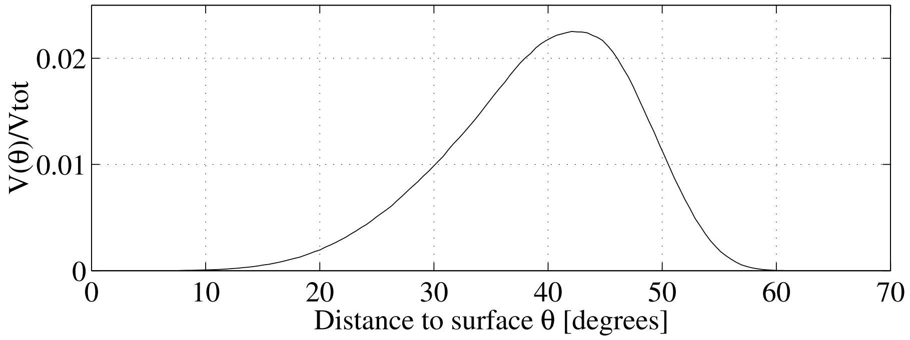


Figure 4.2: (Top) The normalized histograms $N(\theta; s)/|\hat{J}_s|$ of the data sets $\hat{J}_s$ ($|\hat{J}_s| \approx 4.1 \cdot 10^6$ points), where $s = 1, 2, 4, 8, 16, 32$ (see legend) and θ is the distance to the edge manifold $\hat{\mathcal{E}}(\alpha, l/s) \subset \mathrm{S}^8$. (Bottom) Normalized histogram $V(\theta)/V_{\text{tot}}$, which corresponds to the Monte Carlo estimated volume on S^8 of the histogram bins of $N(\theta; s)$ ($V_{\text{tot}} = 10^7$ points).

Fig. 4.3 shows the calculated density for the data sets $\hat{J}_s$, where $s = 1, 2, 4, 8, 16, 32$. These results indicate that the probability distribution of data points in jet space has an *infinite* density at the manifold of blurred step edges (where $\theta = 0$). This is consistent with the results on high-contrast 3×3 pixel image patches in Lee et al. (2001). Furthermore, the density function $\rho(\theta; s)$ is approximately *scale invariant* and seems to converge towards the functional form $\rho(\theta; s) \sim \theta^{-0.7}$ as the scale s increases. The latter results are consistent with many of the previous empirical findings on scale invariance of natural image statistics; see e.g. Field (1987) and Ruderman and Bialek (1994).

In Fig. 4.4 (top), we calculate the cumulative sum $\sum_{\beta \leq \theta} N(\beta; s)/|\hat{J}_s|$ (in percent) of the number of data points as a function of the distance θ to the manifold of edges. For all scales s, we get that 20% of all data points are within 29 degrees of the manifold of edges, which corresponds to less than

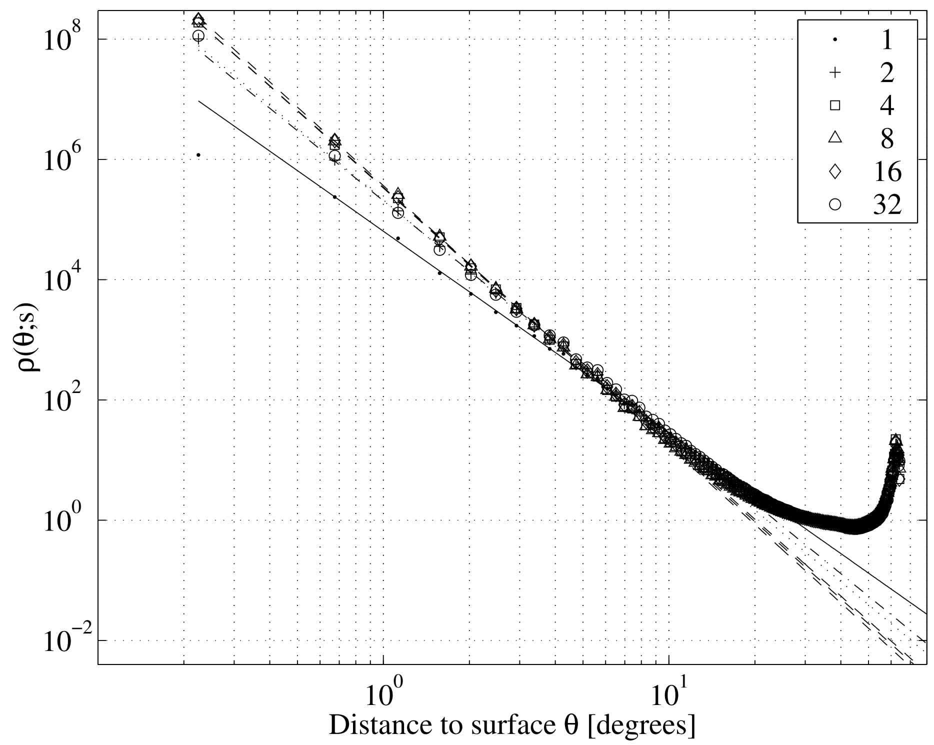


Figure 4.3: Density $\rho(\theta; s)$ versus the distance θ to the edge manifold $\hat{\mathcal{E}}(\alpha, l/s) \subset \mathrm{S}^8$ for data points in $\hat{J}_s$. Each graph represents the density $\rho(\theta; s)$ at a fixed scale s, where $s = 1, 2, 4, 8, 16, 32$ (see legend). By linear regression for $\theta < 9$ degrees we get that $\rho_{\text{fit}}(\theta; s) \sim \theta^{-\gamma_s}$, where $\gamma_s = 1.7, 1.0, 0.7, 0.7, 0.7, 1.1$ $(s = 1, 2, 4, 8, 16, 32)$.

12% of the total surface area of the 8-sphere S^8 (see Fig. 4.4 (bottom)). In other words, points in these subsets of $\hat{J}_s$ are densely clustered around the low-dimensional manifold of edges.

To better illustrate the connection between the density function $\rho(\theta)$ and the image space, we end this section by computing $\rho(\theta)$ for pixels in the classical "Lena" image. Fig. 4.5 shows both scale-space images of Lena and the corresponding log-densities $\log(\rho(\theta))$ for different scales s. In the density calculation, we first map the pixels in the scale-space of Lena into the jet space $\tilde{\mathcal{J}}^3$ given by Eq. (4.9) and then whiten and contrast normalize according to Sec. 4.4.2. We subsequently compute the distance θ of these points to the edge manifold $\hat{\mathcal{E}}(\alpha, l/s) \subset \mathrm{S}^8$ and, finally, we look up the density values by $\rho(\theta) = \theta^{-0.7}$ corresponding to the computed θ-values. We choose the exponent $\gamma = 0.7$ because it corresponds to the exponent for the apparent stable range of scales $s \in [4; 16]$ (see Fig. 4.3). The gray values in the second and fourth columns of Fig. 4.5 code for the magnitude of $\log(\rho(\theta))$ for different

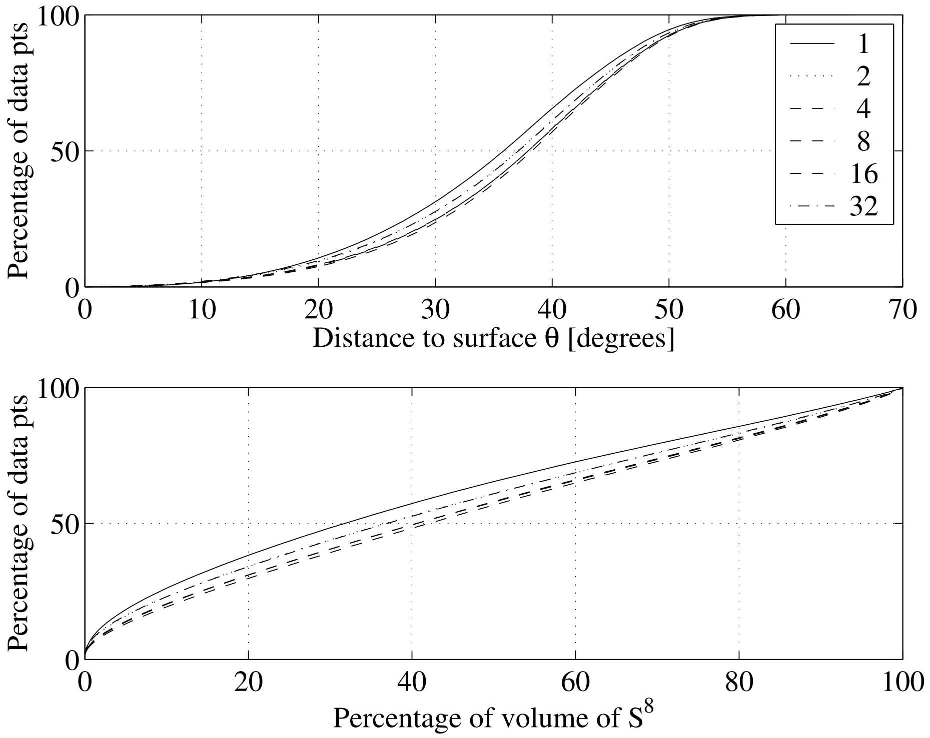


Figure 4.4: (Top) Cumulative sum $\sum_{\beta \leq \theta} N(\beta; s)/|\hat{J}_s|$ (in percent) of the number of data points in $\hat{J}_s$ as a function of the distance to the manifold of edges. (Bottom) Cumulative volume versus cumulative number of data points for data sets $\hat{J}_s$.

scales $s \in [1; 16]$. The first and third columns show the corresponding scale-space images.

4.5 Conclusions

We have extended the results of Lee et al. (2001; 2002) and Chapter 3 from a pixel-based image representation to the jet space representation of linear Gaussian scale-space. The goal of this work is to investigate whether our previous findings on small image patches generalize to larger scales and general filter-based image representations.

In this work, we analyze Gaussian scale-space derivatives computed at randomly chosen points in (a large database of) natural images. At each chosen location, we compute the 3-jet representation (a 9-dimensional vector of up to 3rd order scale-space image derivatives) at different fixed scales. After whitening and contrast-normalizing, the data is on the surface of a

Figure 4.5: The first and third columns show scale-space images of "Lena" (226×226 pixels) for exponentially increasing scales $s \in [1; 16]$. The second and fourth columns show the corresponding log-densities $\log(\rho(\theta))$ in jet space (a bright pixel corresponds to a high density). The densities $\rho(\theta)$ are estimated at each pixel by computing the distance θ to the manifold of blurred step edges and looking up the density values by $\rho(\theta) = \theta^{-0.7}$.

unit 8-sphere in $\mathbb{R}^9$ centered at the origin.

Analysis shows that the probability distribution of empirical data has an infinite density at a 2-dimensional C^∞-differentiable manifold in the 8-sphere (the state space of whitened and contrast-normalized 3-jet representations). This non-linear surface corresponds to the loci in jet space of Gaussian blurred step edges of different orientations α, positions l and scales s. Our results are approximately invariant to a change of scale. In fact, for increasing scales s, the density around the manifold seems to converge towards the functional form $\rho(\theta) \sim \theta^{-0.7}$, where θ is the distance to the edge manifold. For all scales, we find that 20% of the randomly chosen image points have a 3-jet representation that are within 29 degrees of the edge manifold. This region around the manifold corresponds to less than 12% of the total surface volume of the 8-sphere.

The results above are consistent with our earlier findings in Chapter 3 and Lee et al. (2001; 2002) for 3×3 natural image patches. In this work, we have studied the manifold of Gaussian blurred step edges parametrized by the orientation α and the scale-normalized position $l' = l/s$. More generally, we believe that one can define a dictionary of probability models on representations of *general* primitives (edges, bars, blobs, T-junctions) parametrized by $\Phi = \{\phi_1, \phi_2, \ldots\}$ for *any* set of filters $f_1, \ldots, f_N$ in a sensorium. In the N-dimensional state space of the filter-based image representations, the image primitives will define manifolds of the general form

$$M(\Phi) = [f_1(\cdot) * I(\cdot; \Phi), \ldots, f_N(\cdot) * I(\cdot; \Phi)]^T . \tag{4.27}$$

The picture that seems to emerge is that natural images are extremely sparse with most of the data in state space concentrated along these low-dimensional structures that correspond to edges, blobs, bars etc. One has to realize that these manifolds are in general highly *non-linear* — this makes our approach fundamentally different from, for example, ICA and sparse coding where one studies linear projections in state space. It should also be noted that the dimension of the state space of the image data is determined by the number of filters in the analysis (which is usually very large), while the dimension of the manifolds of image primitives is fixed and determined by the complexity of the primitives only. Because of the *low dimensionality* of the primitive manifolds (2 for edges, 3 for bars, etc), a "probabilistic primal sketch" of natural images may have important implications on the information-theoretic bounds one can put on compression of these images.

Acknowledgments

We would like to thank Professor Mads Nielsen, Professor Peter Johansen and Professor David Mumford for their valuable comments.

4.6 Appendix: The Edge Manifold

Theorem 4.1
The loci of all points $\mathcal{E}(\alpha, l, s)$ (Eq. (4.17)) trace out a 2-dimensional C^∞-differentiable manifold in the jet space $\tilde{\mathfrak{J}}^3 \subset \mathbb{R}^9$. This manifold of edge representations is parametrized by the angle α and the ratio l/s between the displacement l and the scale s.

Proof. The edge map $\mathcal{E}$ (Eq. (4.17)) is *infinitely differentiable*, as

$$G_{x^m,y^n}(x, y; \alpha, l, s) = s^{m+n} \cos^m \alpha \ \sin^n \alpha \, \partial_{v^{m+n-1}} \psi(v; l, s) \tag{4.28}$$

and the functions $\cos\alpha$, $\sin\alpha$ and $\psi(v; l, s)$ are C^∞-differentiable with respect to α, l and s (for $s > 0$).

Furthermore, if we introduce the dimensionless variables

$$u' = \frac{u}{s} \ , \quad v' = \frac{v}{s} \ , \quad l' = \frac{l}{s} \ , \quad s' = 1 \tag{4.29}$$

and assume scale-normalized derivatives as in Eq. (4.14), we get that

$$G_{v^n}(u, v; l, s) = G_{v'^n}(u', v'; l', s') \tag{4.30}$$

for $n \geq 1$. Eq. (4.30) follows from the scaling properties of the Gaussian function $\psi(v; l, s)$: We have that

$$\psi(v'; l', s') = s\psi(v; l, s) \tag{4.31}$$

and (Arfken, 1985)

$$\partial_{v^k} \psi(v; l, s) = \left(\frac{-1}{s\sqrt{2}} \right)^k H_k \left(\frac{v - l}{s\sqrt{2}} \right) \psi(v; l, s) \tag{4.32}$$

for $k \geq 0$. Hence,

$$\begin{aligned} \partial_{v^k} \psi(v; l, s) &= \frac{1}{s^k} \left(\frac{-1}{s'\sqrt{2}} \right)^k H_k \left(\frac{v' - l'}{s'\sqrt{2}} \right) \frac{1}{s} \psi(v'; l', s') \\ &= \frac{1}{s^{k+1}} \, \partial_{v'^k} \psi(v'; l', s') \ , \end{aligned} \tag{4.33}$$

which is equivalent to (Eq. (4.30))

$$\underbrace{s^{k+1} \, \partial_{v^k} \psi(v; l, s)}_{G_{v^n}(u,v;l,s)} = \underbrace{s'^{k+1} \, \partial_{v'^k} \psi(v'; l', s')}_{G_{v'^n}(u',v';l',s')} \tag{4.34}$$

where $n = k + 1 \geq 1$.

Although the edge map $\mathcal{E}$ in Eq. (4.17) is a function of three variables (the angle α, the displacement l and the scale s), the loci of all points $\mathcal{E}(\alpha, l, s)$ trace out a *2-dimensional* manifold (in the jet space) which only depends on α and the dimensionless ratio l/s.

Chapter 5

Gaussian Image Models

5.1 Introduction and Notation

This chapter is a discussion of the Gaussian and Brownian image models, with a focus on their properties in linear scale-space. The class of Gaussian image models, to which the Brownian model belongs, are interesting because of their fundamental character and simplicity. I will discuss the statistical properties that these models have in common with natural images. Especially, I will show through an empirical study that the Brownian image model captures the second order statistics of natural images mapped into the so-called jet space. A control experiment of the results of Chapter 4 will be carried out on images generated from the Brownian image model, and the results will be discussed in relation to the results on natural images. Let us begin by defining the notation and terminology that are going to be used throughout this chapter.

By a discrete 2-dimensional image I mean an array or matrix $I \in \mathbb{R}^{N \times M}$. I only discuss gray value images in this chapter and I assume for simplicity that images may take values in the full range of $\mathbb{R}$, ignoring that in a digital computer pixel values are represented by a discrete set of gray values and color images by vectors of discretized color components. Instead of regarding the space of images as the space of matrices $\mathbb{R}^{N \times M}$, an image I can also be viewed as a vector in a finite dimensional Euclidean space $\mathbf{I} \in \mathbb{R}^{NM}$. A pixel is denoted by $I_\alpha = I_{i,j}$ $i, j \in \mathbb{Z}$ indexed by $\alpha \in \{1, \ldots, NM\}$. One can view discrete images as realizations of either a discrete stochastic field $I \in \mathbb{R}^{N \times M}$ or a stochastic vector $\mathbf{I} \in \mathbb{R}^{NM}$. I will use both viewpoints interchangeably in this chapter and I want to define a probability model $p(I)$ or $p(\mathbf{I})$ on the finite dimensional state space of images.

In the continuous setting, an image $f(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$ can be viewed as a realization of a continuous stochastic function (or field) on the plane. In this case I seek a probability model on an infinite dimensional vector space of functions (see e.g. Lifshits (1995) or Huang and Yan (2000)).

If we just view images as arrays of numbers — as discrete images — or maybe continuous functions on the plane, we disregard an important aspect of images: The fact that an image is a set of measurements of the light that goes through the camera optics and hits the image plane. Each number in the image array quantifies the amount of light that hit the corresponding spot on the photographic film, the rods and cones of your eye, or the CCD chip in your digital camera. The light that hits this spot does not come from a single point in the physical world, but is a weighted integral of light originating ideally from every point in the physical world. The optics of the camera (or eye) has a certain aperture or response function that dictates the weighting of the light. Furthermore, many physical scenes that give rise to the same set of measurements exist. This phenomenon has been dubbed metamerism by Koenderink and van Doorn (1996). This metamerism must be kept in mind when analyzing images. Therefore, an image should not be considered as a function in the normal sense of the word, but as an entity constructed through a set of response or test functions. Florack (1997) and Mumford et al. (Mumford, 2001; Mumford and Gidas, 2001) have argued independently for this view, and both propose that images should not be considered as functions, but rather as generalized functions — the so-called Schwartz distributions[1] (see e.g. Friedlander and Joshi (1998) for an introduction to the theory of Schwartz distributions).

A generalized function on a subset of the plane $\Omega \subset \mathbb{R}^2$ is a linear functional that maps test functions $\phi(\mathbf{x}) \in \mathcal{D}(\Omega) \subset C^\infty(\Omega)$ onto $\mathbb{R}$

$$F[\phi] : \mathcal{D}(\Omega) \to \mathbb{R} ,$$

where $\mathcal{D}(\Omega)$ is a vector space of C^∞ functions on Ω. Let the vector space of generalized functions on Ω be denoted as $\mathcal{D}'(\Omega)$ and notice that $\mathcal{D}(\Omega) \subset \mathcal{D}'(\Omega)$. The test functions in $\mathcal{D}(\Omega)$ are required to have compact support, but if we substitute compact support with the requirement that $\phi(\mathbf{x})$ must be rapidly decreasing[2] we get the so-called tempered distributions. Let $\mathcal{S}(\Omega)$ denote the vector space of test functions $\phi(\mathbf{x})$ which are rapidly decreasing. Linear functionals of elements from $\mathcal{S}(\Omega)$ are called tempered generalized functions and the set of tempered generalized functions is denoted as $\mathcal{S}'(\Omega)$. Notice that $\mathcal{D}(\Omega) \subset \mathcal{S}(\Omega)$ and $\mathcal{S}'(\Omega) \subset \mathcal{D}'(\Omega)$.

[1] To avoid confusion between probability distributions and Schwartz distributions I will use the term generalized function for Schwartz distributions.

[2] That is, all derivatives must decrease faster than polynomials.

I assume that naturally occurring images belong to the space of tempered generalized functions $\mathcal{S}'(\Omega)$, $\Omega \subset \mathbb{R}^2$. In the continuous setting I therefore seek a probability distribution on $\mathcal{S}'(\Omega)$.

If we consider locally L_1-functions[3] $L_1(\mathbb{R}^2)$ on the plane $f(\mathbf{x}) : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^2$, we get an important example of tempered generalized functions by

$$\langle f, \phi \rangle = F[\phi] = \int_\Omega f(\mathbf{x})\phi(\mathbf{x})\, d\mathbf{x}$$

where $\phi(\mathbf{x}) \in \mathcal{S}(\Omega)$. Locally L_1-functions are themselves examples of tempered generalized functions $f(\mathbf{x}) \in \mathcal{S}'(\Omega)$. That is, we have $L_1(\mathbb{R}^2) \subset \mathcal{S}'(\Omega) \subset \mathcal{D}'(\Omega)$.

Florack (1997) has pointed out the connection between linear scale-space theory (Iijima, 1962; Witkin, 1983; Koenderink, 1984) and viewing images as tempered distributions. In scale-space theory the test functions are the Gaussian aperture function and its derivatives, which all decrease rapidly and therefore belong to $\mathcal{S}(\Omega)$. The linear scale-space aperture function is

$$\psi(\mathbf{x}; s) \equiv \frac{1}{2\pi s^2} \exp\left(-\frac{(x^2 + y^2)}{2s^2}\right)$$

where $\mathbf{x} = (x, y)^T$ and $s \geq 0$ denotes the scale parameter, and its partial derivatives are written as $\partial_{x^n y^m}\psi(\mathbf{x}; s)$. The scale-space of an image $f(\mathbf{x})$, which is assumed to be a locally L_1-function, is a tempered generalized function given by

$$L(\mathbf{x}; s) = F[\psi(\mathbf{x}; s)] \equiv \int_\Omega f(\mathbf{x}')\psi(\mathbf{x} - \mathbf{x}'; s)\, d\mathbf{x}' = f(\mathbf{x}) * \psi(\mathbf{x}; s)$$

with $L(\mathbf{x}; s = 0) \equiv f(\mathbf{x})$. Instead of using the cumbersome notation $F[\psi(\mathbf{x}; s)]$ to denote the scale-space of $f(\mathbf{x})$ I use the, in the scale-space literature, more common notation $L(\mathbf{x}; s)$, making the dependency on $\psi(\mathbf{x}; s)$ implicit.

By introducing generalized functions the concept of differentiation of otherwise non-differentiable functions is made sensible (see e.g. Friedlander and Joshi (1998)). In the analysis of stochastic functions the concept of generalized functions has been proposed for solving the problem of differentiability of realizations of stochastic functions (see e.g. Mandelbrot and van Ness (1968)). By using the linear scale-space representation we can define derivatives of images $f(\mathbf{x})$ by applying the differential operator $\partial_{x^n y^m}$ to the test function $\psi(\mathbf{x}; s)$

$$L_{x^n y^m}(\mathbf{x}; s) = f(\mathbf{x}) * \partial_{x^n y^m}\psi(\mathbf{x}; s) \ .$$

[3] A function $f(\mathbf{x})$ is a locally L_1-function if $\int_K |f|\, d\mathbf{x} < \infty$ for all compact sets $K \subset \Omega$.

The scale-space partial derivative $L_{x^n y^m}(\mathbf{x}; s)$ is a tempered generalized function belonging to $\mathcal{S}'(\Omega)$.

Throughout this chapter I will discuss properties of the Gaussian and Brownian image models both in the discrete and continuous setting. For an in-depth exposé of Gaussian image models the interested reader is encouraged to look at Mumford (2001) and for Gaussian stochastic functions in general Lifshits (1995) provides insight.

5.2 The Gaussian Model

I start out by discussing the Gaussian image model in both the discrete and continuous setting.

The Gaussian distribution $p(\mathbf{I})$ on $\mathbb{R}^{NM}$ is defined such that all projections of vectors $\mathbf{I} \in \mathbb{R}^{NM}$ onto arbitrary vectors $\mathbf{y} \in \mathbb{R}^{NM}$ are 1-dimensional Gaussian stochastic variables, i.e. have a Gaussian distribution on $\mathbb{R}$. Furthermore, a Gaussian distribution on $\mathbb{R}^{NM}$ is fully parameterized by its mean vector $\mathbf{m}$ and covariance matrix C. The elements of the covariance matrix are given by

$$C_{\alpha\beta} = \int_{\mathbb{R}^{NM}} (I_\alpha - m_\alpha)(I_\beta - m_\beta)\, dp(\mathbf{I}) \tag{5.1}$$

and the mean vector is

$$\mathbf{m} = \int_{\mathbb{R}^{NM}} \mathbf{I}\, dp(\mathbf{I}) \ . \tag{5.2}$$

Let us start by defining the discrete zero mean Gaussian image model with independently identically distributed (i.i.d.) pixel intensity values, also known as *discrete white noise*. The zero mean Gaussian density function for pixel values I_α with variance σ^2 is

$$p(I_\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{I_\alpha^2}{2\sigma^2}\right) \ . \tag{5.3}$$

The Gaussian probability density for discrete white noise images can be written as the density of the stochastic vector $\mathbf{I} = (I_1, \ldots, I_{NM})^T \in \mathbb{R}^{NM}$

$$p(\mathbf{I}) = p(I_1, \ldots, I_{NM}) = \prod_{\alpha=1}^{NM} p(I_\alpha) = (2\pi\sigma^2)^{-\frac{NM}{2}} \exp\left(-\frac{\sum_\alpha I_\alpha^2}{2\sigma^2}\right) \ . \tag{5.4}$$

Discrete white noise is stationary which implies translational invariance. See Fig. 5.1 for a realization of the zero mean i.i.d. Gaussian image model.

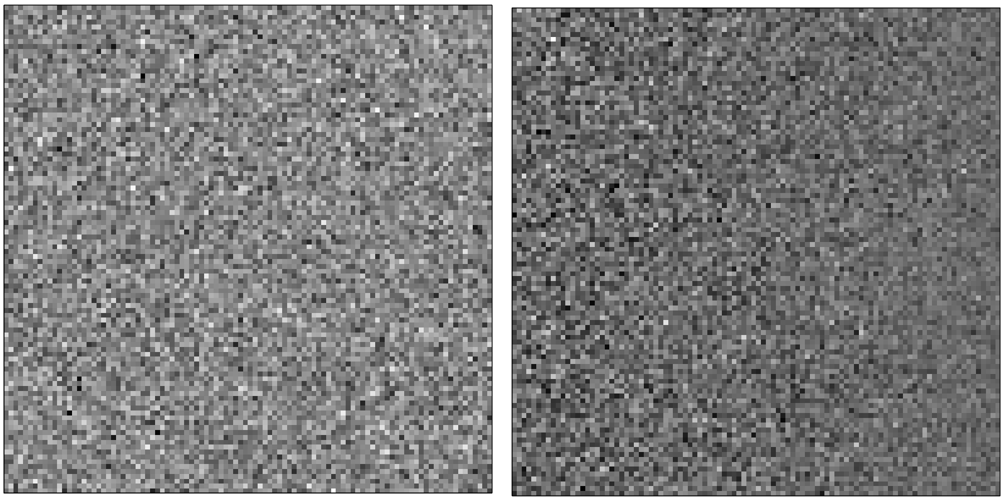


Figure 5.1: (Left) A 100×100 pixel sample from the zero mean i.i.d. Gaussian image model (Eq. (5.4)). (Right) Sample from the general Gaussian image model (Eq. (5.5)). The image on the right is a 100×100 pixel sample from the zero mean general Gaussian image model with a covariance matrix with eigenvalues (variances along the eigenvectors) decreasing linearly from 10 to 1. The eigenvectors are generated by rotating the columns of the identity matrix around 2500 randomly chosen axes by 45 degrees. Notice the introduced correlation between the pixels.

In general, the Gaussian distribution on the vector space $\mathbf{I} \in \mathbb{R}^{NM}$ with mean vector $\mathbf{m} \in \mathbb{R}^{NM}$ and positive definite covariance matrix $C \in \mathbb{R}^{NM \times NM}$ can be written as

$$p(\mathbf{I}) = (2\pi)^{-\frac{NM}{2}} (\det C)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{I}-\mathbf{m})^T C^{-1}(\mathbf{I}-\mathbf{m})\right) . \tag{5.5}$$

In Fig. 5.1 (right) a sample from the general Gaussian image model is shown.

In order to define the continuous Gaussian image model one has to define a Gaussian probability measure on an infinite dimensional vector space of functions. The space of functions I am interested in is the space of tempered generalized functions $\mathcal{S}'(\mathbb{R}^2)$. The Gaussian distribution on the function space $\mathcal{S}'(\mathbb{R}^2)$ is defined in a similar fashion as in the finite dimensional case. All the distributions of stochastic vectors $(f_1, \ldots, f_N)$ obtained by sampling a stochastic function $f(\mathbf{x}) \in \mathcal{S}'(\mathbb{R}^2)$ at different sets of $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^2$ are called the finite distributions of the stochastic function $f(\mathbf{x})$. A stochastic function $f(\mathbf{x}) \in \mathcal{S}'(\mathbb{R}^2)$ is said to be Gaussian if all its finite distributions are Gaussian distributed stochastic vectors.

In general all linear functionals of a Gaussian distributed stochastic function are also Gaussian distributed. Specifically, generalized functions $\langle f, \phi \rangle = \int f(x)\phi(x)\, dx$ (for any test functions $\phi(x) \in \mathcal{S}(\mathbb{R}^2)$) of a Gaussian distributed stochastic function $f(x)$ are also Gaussian distributed stochastic functions. Hence the scale-space and scale-space derivatives of a Gaussian stochastic function are Gaussian distributed.

The mean function $\bar{f} \in \mathcal{S}'(\mathbb{R}^2)$ of a probability measure p on $\mathcal{S}'(\mathbb{R}^2)$ is defined such that the expectation $E[\cdot]$ for all continuous linear functionals $h : \mathcal{S}'(\mathbb{R}^2) \to \mathbb{R}$ is

$$E[h] = h(\bar{f}) .$$

The function $C(\mathbf{x}, \mathbf{y})$ is the covariance function of the probability measure p on $\mathcal{S}'(\mathbb{R}^2)$, if for all continuous linear functionals h_1, h_2 given by

$$h_1(f) = \int f(\mathbf{x})\phi_1(\mathbf{x})\, d\mathbf{x} \quad \text{and} \quad h_2(f) = \int f(\mathbf{x})\phi_2(\mathbf{x})\, d\mathbf{x} ,$$

we have

$$E[(h_1 - h_1(\bar{f})) \cdot (h_2 - h_2(\bar{f}))] = \iint \phi_1(\mathbf{x})\phi_2(\mathbf{y}) C(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y} . \tag{5.6}$$

The continuous equivalent of the discrete i.i.d. Gaussian model is *Gaussian white noise*. Gaussian white noise on $\mathcal{S}'(\mathbb{R}^2)$ is defined as a Gaussian stochastic function $\nu \in \mathcal{S}'(\mathbb{R}^2)$ with zero mean and covariance

$$E[\langle \nu, \phi_1 \rangle \langle \nu, \phi_2 \rangle] = \sigma^2 \int \phi_1(\mathbf{x})\phi_2(\mathbf{x})\, d\mathbf{x} . \tag{5.7}$$

This is a generalization of the discrete i.i.d. Gaussian model, since if ϕ_1 and ϕ_2 have disjoint support, we have $E[\langle \nu, \phi_1 \rangle \langle \nu, \phi_2 \rangle] = 0$ and any disjoint open sets are independently distributed (Mumford, 2001). In linear scale-space the Gaussian white noise $W(\mathbf{x}; s) = \nu(\mathbf{x}) * \psi(\mathbf{x}; s)$ has the covariance

$$\begin{aligned}
E[W(\mathbf{x}; s)W(\mathbf{y}; s)] &= \sigma^2 \int_{\mathbb{R}^2} \psi(\mathbf{x}' - \mathbf{x}; s)\psi(\mathbf{x}' - \mathbf{y}; s)\, d\mathbf{x}' \\
&= \frac{\sigma^2}{4\pi^2 s^4} \int_{\mathbb{R}^2} \exp\left(-\frac{(\mathbf{x}' - \mathbf{x})^2}{2s^2}\right) \exp\left(-\frac{(\mathbf{x}' - \mathbf{y})^2}{2s^2}\right) d\mathbf{x}' \\
&= \frac{\sigma^2}{4\pi s^2} \exp\left(-\frac{(\mathbf{x} - \mathbf{y})^2}{4s^2}\right) .
\end{aligned}$$

Hence the covariance function of Gaussian white noise in scale-space $W(\mathbf{x}; s)$ may be written as

$$C_W(\mathbf{x}, \mathbf{y}; s) = \frac{\sigma^2}{4\pi s^2} \exp\left(-\frac{(\mathbf{x} - \mathbf{y})^2}{4s^2}\right) . \tag{5.8}$$

Thus scale-space smoothing changes the covariance of white noise and introduces spatial dependencies. Remember that by definition $W(\mathbf{x}; s = 0) = \nu(\mathbf{x})$ and for scale $s = 0$ the spatial covariance of $W(\mathbf{x}; s)$ reduces to $E[\nu(\mathbf{x})\nu(\mathbf{y})] = \sigma^2\delta(\mathbf{x} - \mathbf{y})$, where $\delta(\cdot)$ is the Dirac delta function[4].

The probability density of continuous Gaussian white noise is given by

$$p(\langle \nu, \phi \rangle) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma^2} \int \langle \nu(\mathbf{x}), \phi(\mathbf{x}) \rangle^2 \, d\mathbf{x}\right) , \tag{5.9}$$

where Z is a normalization factor that ensures $\int p(\langle \nu, \phi \rangle)\, d\mathbf{x} = 1$. We cannot write the probability density for the Gaussian white noise image model in scale-space, because the covariance function in Eq. (5.8) is not invertible. But it is possible to write the characteristic function[5] of the probability distribution of continuous white noise in scale-space $W(\mathbf{x}; s)$,

$$\begin{aligned}
E[e^{iW(\mathbf{x};s)}] &= \exp\left(-\frac{1}{2}\langle C_W W, W \rangle\right) \\
&= \exp\left(-\frac{1}{2} \iint W(\mathbf{x}; s) C_W(\mathbf{x}, \mathbf{y}; s) W(\mathbf{y}; s)\, d\mathbf{x}\, d\mathbf{y}\right) .
\end{aligned}$$

[4] The Dirac delta function $\delta(x)$ is a generalized function defined such that $\int \delta(x)\, dx = 1$ and $\langle \delta(x), \phi(x) \rangle = \phi(0)$.

[5] The characteristic function of a distribution is a complete description of the distribution.

A Gaussian stochastic function $f(\mathbf{x})$ is stationary if it has the same mean value $E[f(\mathbf{x})] = a$, $a \in \mathbb{R}$ for all $\mathbf{x} \in \mathbb{R}^2$, and its covariance $E[(f(\mathbf{x}) - E[f(\mathbf{x})])(f(\mathbf{y}) - E[f(\mathbf{y})])]$ depends on one parameter only $\mathbf{x}-\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, i.e. does not depend on position. The covariance function of Gaussian white noise $\nu(\mathbf{x})$ may be written as $C(\mathbf{x}-\mathbf{y}) = \sigma^2\delta(\mathbf{x}-\mathbf{y})$, hence white noise is stationary. The Gaussian white noise model in scale-space $W(\mathbf{x}; s)$ defined by the covariance function of Eq. (5.8) is also a stationary stochastic function.

Sample functions of Gaussian white noise $\nu(\mathbf{x})$ are not differentiable. The power spectrum[6] of a zero mean white noise image is flat, i.e. constant,

$$S(\xi) = \iint_{-\infty}^{\infty} C(\mathbf{x}) e^{-i\xi\cdot\mathbf{x}}\, d\mathbf{x} = \sigma^2 \iint_{-\infty}^{\infty} \delta(\mathbf{x}) e^{-i\xi\cdot\mathbf{x}}\, d\mathbf{x} = \sigma^2 \ .$$

This means that realizations of white noise have a uniform energy distribution across all frequencies. This implies that such functions have infinite frequencies, which in turn implies that realizations of the white noise image model are not differentiable. Nevertheless, by introducing the concept of generalized functions and linear scale-space we may define derivatives of realizations of the Gaussian white noise image model in a sensible way.

In the continuous case I define the derivatives of a realization of white noise by viewing the realization as a generalized function. By using the linear scale-space representation one can compute partial derivatives of $\nu(\mathbf{x})$ by convolution with partial derivatives of the scale-space aperture function $\phi_{x^n y^m}(\mathbf{x}; s)$. All linear functionals of Gaussian white noise $\nu(\mathbf{x})$ must be Gaussian, hence all partial scale-space derivatives $W_{x^n y^m}(\mathbf{x}; s)$ must be Gaussian stochastic functions. This follows immediately from the definition of the Gaussian distribution on the space of tempered generalized functions $\mathcal{S}'(\mathbb{R}^2)$. I have computed partial scale-space derivatives of sample images of Gaussian white noise and estimated the distribution of the derivatives (see Fig. 5.2). It can be seen that these scale-space derivatives are zero mean Gaussian distributed.

In the discrete case, image derivatives may be approximated in terms of differences among neighboring pixels I_α. The difference images $D_x I \approx \frac{\partial}{\partial x} I$ and $D_y I \approx \frac{\partial}{\partial y} I$ also have zero mean Gaussian distributed pixels (see Fig. 5.3). This can easily be seen by the following argument. Define $D_x I$ by forward differencing

$$(D_x I)_\alpha \equiv I_{i+1,j} - I_{i,j} \ , \tag{5.10}$$

[6]The power spectrum or power spectral density of a stationary stochastic function is usually defined as the Fourier transform of the autocorrelation function. In the case of zero mean stochastic functions the autocorrelation reduces to the spatial covariance.

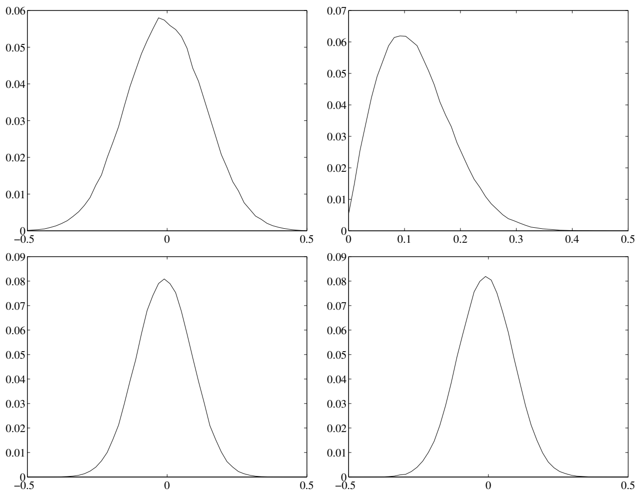


Figure 5.2: (Top left) Pixel histogram $p(W)$ of a white noise scale-space image at scale $s = 2$. (Top right) Histogram of the gradient magnitude $p(\|\nabla W\|)$, which is a close approximation to the Rayleigh distribution. (Bottom) Histograms $p(W_x)$ and $p(W_y)$ of partial derivatives in the x and y directions.

where α is the index of pixel (i, j). Since $(D_x I)_\alpha$ is a difference of Gaussian stochastic variables it is also a Gaussian stochastic variable. The density $p(Z)$ of the sum $Z = X + Y$ (difference $Z = X - Y$) of two independent stochastic variables X and Y is given by the convolution (correlation) between the densities of the two variables, $p(Z) = p(X) * p(Y)$. Hence we can derive the expression for the density of $(D_x I)_\alpha$ as follows

$$\begin{aligned} p((D_x I)_\alpha) &= \int_{-\infty}^{\infty} p(I_\alpha) p((D_x I)_\alpha + I_\alpha)\, dI_\alpha \\ &= \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{(D_x I)_\alpha^2}{4\sigma^2}\right) \end{aligned} \tag{5.11}$$

and similarly for $(D_y I)_\alpha$

$$p((D_y I)_\alpha) = \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{(D_y I)_\alpha^2}{4\sigma^2}\right) . \tag{5.12}$$

The Gaussian distributions of $(D_x I)_\alpha$ and $(D_y I)_\alpha$ are stationary. In general, the increments (differences) in all directions (i.e. not just along the x and y axes) of a Gaussian stochastic function are Gaussian distributed and stationary. This can be formalized in the continuous setting as follows.

Any directional derivative $\partial W(\mathbf{x}; s)/\partial \mathbf{m}\big|_{P_0}$ in the direction $\mathbf{m}$, ($\|\mathbf{m}\| = 1$), at the point P_0 is a Gaussian stochastic variable, since $\partial W(\mathbf{x}; s)/\partial \mathbf{m}\big|_{P_0}$

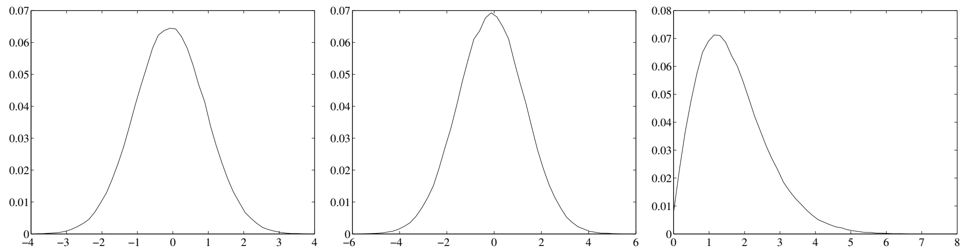


Figure 5.3: (Left) Pixel histogram $p(I_\alpha)$. (Center) Histogram $p((D_xI)_\alpha)$ of pixel differences in the x direction. (Right) Histogram of the gradient magnitude $p(\sqrt{(D_xI)_\alpha^2 + (D_yI)_\alpha^2})$, which is a close approximation to the Rayleigh distribution.

is linear in the two Gaussian variables $W_x(P_0; s)$ and $W_y(P_0; s)$

$$\left.\frac{\partial W(\mathbf{x}; s)}{\partial \mathbf{m}}\right|_{(x,y)=P_0} = W_x(P_0; s)\cos(\theta) + W_y(P_0; s)\sin(\theta) \tag{5.13}$$

where θ is the angle between the positive x-axis and the directional vector $\mathbf{m}$. The distribution of directional derivatives of a Gaussian function is Gaussian and stationary. From Eq. (5.13) one can also conclude that $W_x(P_0; s)$ and $W_y(P_0; s)$ are independently distributed, since when $\mathbf{m}$ is parallel to the x-axis we have $\theta = 0$ and $\partial W(\mathbf{x}; s)/\partial \mathbf{m}\big|_{P_0} = W_x(P_0; s)$, similarly for the y-axis case. Furthermore, $p(W_x)$ and $p(W_y)$ are identically Gaussian distributed. This implies that the joint distribution of (W_x, W_y), where $W_x = W_x(P_0; s)$ and $W_y = W_y(P_0; s)$, is an isotropic Gaussian distribution $p(W_x, W_y)$ with density

$$p(W_x, W_y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(W_x^2 + W_y^2) \right) . \tag{5.14}$$

The gradient magnitude $\|\nabla W\|$ of an image sampled from a Gaussian stochastic function must be Rayleigh distributed. To see this, let us compute the probability of the gradient magnitude $\|\nabla W\|$ having a value in the annulus in the (W_x, W_y)-plane given by the inner r_1 and outer r_2 radius. The joint distribution of (W_x, W_y) at the point P_0 is an isotropic Gaussian and we can make a change of variables to polar coordinates $(r = \|\nabla W\|, \theta)$ and write

$$p(W_x, W_y) = p(r, \theta) . \tag{5.15}$$

The distribution $p(W_x, W_y)$ is uniform along the θ axis since $p(W_x, W_y)$ is isotropic, and $p(r, \theta) = \frac{1}{2\pi\sigma^2} \exp\big(-\frac{r^2}{2\sigma^2}\big)$ is Gaussian. The probability of a value in the annulus is therefore

$$P(\|\nabla W\| \in [r_1; r_2]) = \int_{r_1}^{r_2} \int_0^{2\pi} p(r,\theta) r \, d\theta \, dr = \int_{r_1}^{r_2} \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) dr \ .$$

Thus the probability density of the gradient magnitude is

$$p(\|\nabla W\|) = \frac{\|\nabla W\|}{\sigma^2} \exp\left(-\frac{\|\nabla W\|^2}{2\sigma^2}\right) \ , \tag{5.16}$$

which is exactly the density function of the Rayleigh distribution. See Fig. 5.2 and Fig. 5.3 for experimental evidence.

From Eq. (5.10 – 5.12) it can be seen that differences introduce spatial dependencies. In general, differentiation of stochastic functions introduces spatial dependencies. For derivatives of continuous Gaussian white noise in scale-space we may write the spatial covariance function as

$$\begin{aligned} &E[W_{x^n y^m}(\mathbf{x}_1; s) W_{x^n y^m}(\mathbf{x}_2; s)] \\ &\qquad = \sigma^2 \int_{\mathbb{R}^2} \partial_{x^n y^m} \psi(\mathbf{x}_0; s)\Big|_{\mathbf{x}_0 = \mathbf{x}' - \mathbf{x}_1} \partial_{x^n y^m} \psi(\mathbf{x}_0; s)\Big|_{\mathbf{x}_0 = \mathbf{x}' - \mathbf{x}_2} d\mathbf{x}' \ . \end{aligned}$$

It is also possible to say something about the dependency and correlation between partial derivatives of different order. The distribution of Gaussian white noise in k-jet space is Gaussian, since scale-space derivatives of Gaussian white noise are also Gaussian distributed. The k-jet space is the space of partial derivatives up to order k, i.e. $(W(\mathbf{x}; s), \ldots, W_{x^n y^m}(\mathbf{x}; s)) \in \mathbb{R}^N$ where $n + m = k$ and $N = (2 + k)!/(2k!)$. The covariance matrix of Gaussian white noise in jet space is (see derivation in appendix Sec. 5.6)

$$\begin{aligned} &E[\langle W_{x^{n_1} y^{m_1}}(\mathbf{x}; s), W_{x^{n_2} y^{m_2}}(\mathbf{x}; s)\rangle] \\ &\qquad = (-1)^{\frac{(n+m)}{2} + n_2 + m_2} \frac{\sigma^2}{\pi s^{n+m+2}} \frac{n!m!}{2^{n+m+2}(n/2)!(m/2)!} \end{aligned} \tag{5.17}$$

whenever both $n = n_1 + n_2$ and $m = m_1 + m_2$ are even integers, otherwise

$$E[\langle W_{x^{n_1} y^{m_1}}(\mathbf{x}; s), W_{x^{n_2} y^{m_2}}(\mathbf{x}; s)\rangle] = 0 \ .$$

We see that the derivatives $W_{x^{n_1} y^{m_1}}(\mathbf{x}; s)$ and $W_{x^{n_2} y^{m_2}}(\mathbf{x}; s)$ are correlated whenever n and m are both even.

The Gaussian image model is not scale invariant. To see this, one can compute block averages of a Gaussian image and plot the resulting pixel histogram (see Fig. 5.4). One immediately sees that the variance of the empirically estimated pixel distribution has changed under this operation of

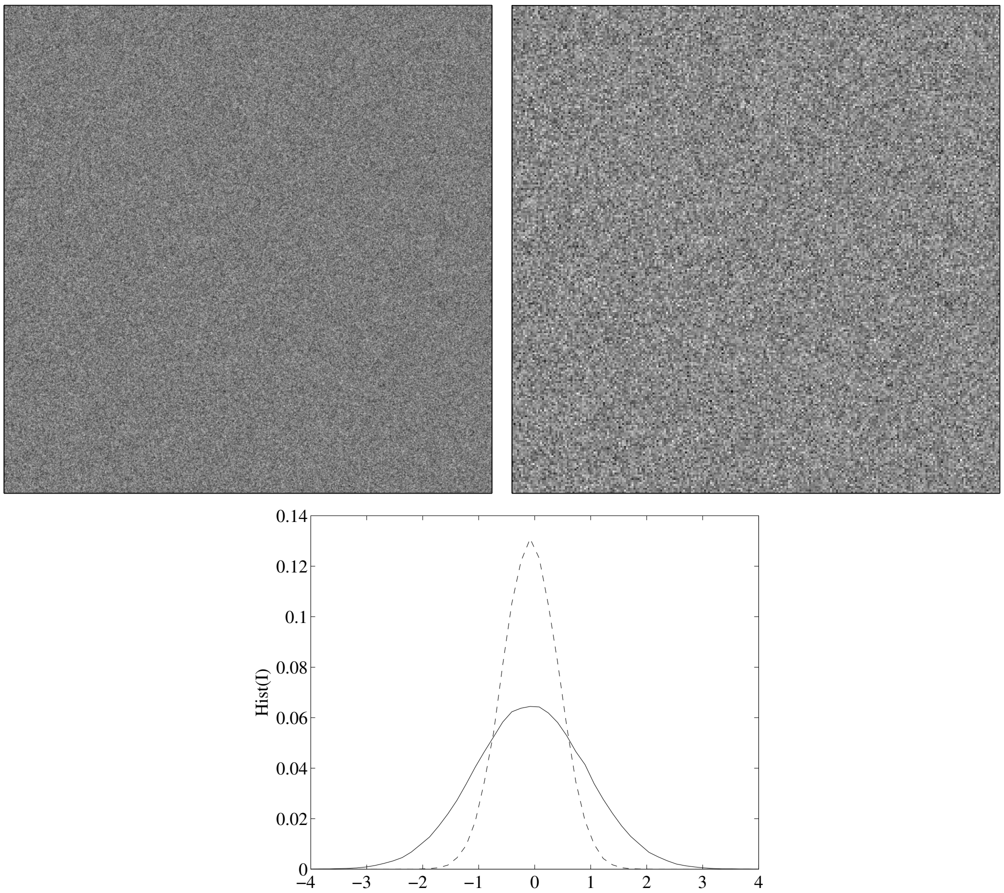


Figure 5.4: The Gaussian image model is not scale invariant. (Top row) A Gaussian image and the same image down-sampled 50 % by block averaging. (Bottom row) Pixel intensity histograms for the two images. The two histograms are Gaussian but clearly have different variances.

rescaling. We also see this lack of scale invariance from the fact that the spatial covariance function $C_W(\mathbf{x}, \mathbf{y}; s)$, Eq. (5.8), of white noise in scale-space $W(\mathbf{x}; s)$ is dependent on the scale s. Furthermore, in the covariance matrix for k-jet space in Eq. (5.17) there is a scale dependency, $1/s^{n+m+2}$, which we can not get rid of by standard scale normalization of derivatives, i.e. by multiplying each derivative $W_{x^n y^m}$ with s^{n+m}.

The Gaussian image model is a poor model of natural images, because several of its properties differ from the empirical findings for natural images. First of all pixel intensity values of natural images are not Gaussian (Field, 1987). Secondly, pixel differences and general filter responses are not Gaussian distributed and in fact have a distribution that may be approximated by either the generalized Laplacian distribution (Mallat, 1989; Huang and Mumford, 1999) or the so-called Bessel K forms proposed by Srivastava et al. (2002). But most importantly, natural images exhibit scale invariance of the covariance structure (Field, 1987; Ruderman and Bialek, 1994), which is not the case for the Gaussian image model as we have just seen. I will now discuss an extension of the Gaussian image model — the Brownian image model — which has the property of scale invariance.

5.3 The Brownian Model

Brownian motion was first described by the botanist R. Brown in 1828 as the random movement of pollen in water, i.e. a path in $\mathbb{R}^3$. Around 1905–1908 Einstein (see e.g. Einstein (1956)) and Bachelier developed a mathematical theory of Brownian motion. In 1923, Wiener proposed a model of the Brownian motion in the form of a stochastic function on $\mathbb{R}$. Here we are interested in the Brownian stochastic function on $\mathbb{R}^2$ or the *Brownian image*, which is sometimes called the Lévy Brownian function. The Brownian image (Mandelbrot and van Ness, 1968; Pentland, 1984; Mumford, 2001) is a stationary scale invariant Gaussian stochastic function. The *fractional Brownian image* (Mandelbrot and van Ness, 1968; Pentland, 1984) is a generalization of this kind of stochastic function, which generally is not scale invariant but exhibit the property of self similarity (see Appendix 5.7 for an explanation of this concept).

The Brownian image may be defined in several ways (see e.g. (Mandelbrot and van Ness, 1968; Pentland, 1984; Lifshits, 1995; Mumford, 2001)), but I will only give the definition, as proposed by Pentland (1984), through the more general fractional Brownian image. The image $\beta_H(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$ is a *fractional Brownian image* if for all $\mathbf{x}, \Delta\mathbf{x} \in \mathbb{R}^2$

$$P\left(\frac{\beta_H(\mathbf{x}+\Delta\mathbf{x}) - \beta_H(\mathbf{x})}{\|\Delta\mathbf{x}\|^H} < y\right) = F(y) \tag{5.18}$$

where $F(y)$ is the cumulative probability distribution function and the parameter $0 < H \leq 1$ controls the "roughness" of the image. The H parameter is related to the fractal dimension D via $D = 3 - H$ (see e.g. Ott (1993) for a definition of the fractal dimension).

If $F(y)$ is the cumulative distribution function of a zero mean Gaussian distribution and $H = 1/2$, then I call the image $\beta(\mathbf{x}) = \beta_{1/2}(\mathbf{x})$ a Gaussian Brownian image or simply a *Brownian image*. The Brownian image is a zero mean Gaussian stochastic function on $\mathbb{R}^2$.

The variance of the distribution of increments of a fractional Brownian image is

$$E[(\beta_H(\mathbf{x}) - \beta_H(\mathbf{y}))^2] = \|\mathbf{x} - \mathbf{y}\|^{2H} \; . \tag{5.19}$$

The Brownian image model has stationary Gaussian distributed increments or more general stationary Gaussian distributed directional derivatives. Obviously, from Eq. (5.19) we have that the variance of the increments of the Brownian image is

$$E[(\beta(\mathbf{x}) - \beta(\mathbf{y}))^2] = \|\mathbf{x} - \mathbf{y}\| \; . \tag{5.20}$$

Brownian motion on $\mathbb{R}$ has independent increments, but this does not generalize to higher dimensions (Lifshits, 1995).

The power spectrum of the fractional Brownian image is $S(\xi) = \sigma_0^2/\|\xi\|^\alpha$ where $\alpha = 2H + 1$. For Brownian images ($H = 1/2$) this reduces to $S(\xi) = \sigma_0^2/\|\xi\|^2$, which means that the Brownian image has equal energy at equal frequency octaves[7]. This property implies scale invariance of the covariance structure of the Brownian image model. This in turn implies that Brownian images have a scale invariant distribution since Brownian images are Gaussian, thence are completely parameterized by its mean and covariance. Field (1987) has shown that ensembles of natural images have the same power spectrum thus giving evidence of scale invariance of the second order statistics of natural images.

The Brownian image is a zero mean stochastic function and from Wiener-Khinchin relations[8] between the power spectrum and the autocorrelation

[7]An octave is an interval which is multiplied by some constant from the start to the end of the interval, $[\xi_0; n\xi_0]$.

[8]The Wiener-Khinchin relations between the power spectrum $S(\xi)$ and the autocorrelation function $E[f(\mathbf{x})f(\mathbf{x}+\mathbf{y})]$ of a stationary stochastic function $f(\mathbf{x})$ are $S(\xi) = \iint E[f(\mathbf{x})f(\mathbf{x}+\mathbf{y})]e^{-i\xi\cdot\mathbf{y}}\, d\mathbf{y}$ and $E[f(\mathbf{x})f(\mathbf{x}+\mathbf{y})] = \frac{1}{(2\pi)^2}\iint S(\xi)e^{i\xi\cdot\mathbf{y}}\, d\xi$. For zero mean stochastic functions the autocorrelation function reduces to the spatial covariance function.

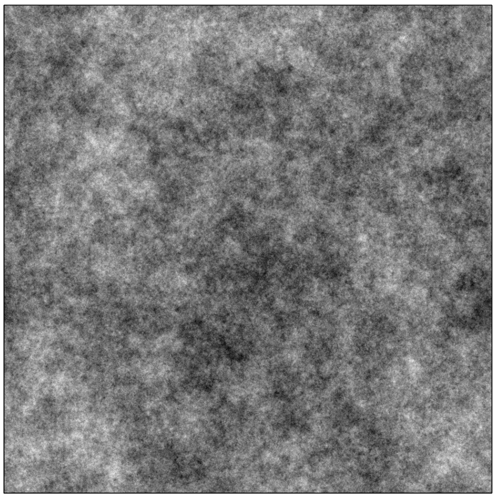


Figure 5.5: A sample from the Brownian image model ($H = 1/2$).

function we have that the power spectrum of a Brownian image is the Fourier transform of the spatial covariance function $C(\mathbf{x}, \mathbf{y})$. The inverse Fourier transform of the power spectrum $S(\xi) = \sigma_0^2/\|\xi\|^2$ yields the spatial covariance function $C(\mathbf{x}, \mathbf{y})$, which is obviously not proportional to the Dirac delta function, which leads to the conclusion that Brownian images have spatial dependencies, contrary to white noise. Furthermore, the covariance $C(\mathbf{x}, \mathbf{y})$ does not dependent on position, only on the separation $\mathbf{x} - \mathbf{y}$ and we can conclude that the Brownian image model is also stationary.

A sample of the periodic Brownian image can be constructed by sampling zero mean i.i.d. Gaussian distributed Fourier coefficients ξ in the frequency domain with variance $\sigma_0^2/\|\xi\|^2$ and enforcing conjugate symmetry. The constant σ_0^2 can be thought of as a global variance offset. The Fourier transformation of this image yields a periodic Brownian image. See Fig. 5.5 for a sample of a Brownian image.

As for all Gaussian stochastic functions the Brownian image is fully described by specifying its mean and covariance functions. By definition the Brownian image has mean zero. The covariance function of the Brownian image is defined in several ways in the literature (Lifshits, 1995; Mumford, 2001). Mumford (2001) argues that under the assumptions of stationarity and scale invariance the spatial covariance function of the Brownian image $\beta(\mathbf{x})$ must be

$$C(\mathbf{x}, \mathbf{y}) = \sigma_0^2 \log\left(\frac{1}{\|\mathbf{x} - \mathbf{y}\|}\right) . \tag{5.21}$$

For Brownian images in scale-space, $B(\mathbf{x}; s) = \beta(\mathbf{x}) * \psi(\mathbf{x}; s)$, we have the spatial covariance

$$\begin{aligned} C_B(\mathbf{x}, \mathbf{y}; s) &= E[B(\mathbf{x}; s) B(\mathbf{y}; s)] \\ &= \int_{\mathbb{R}^2} \psi(\mathbf{x}' - \mathbf{x}; s) \psi(\mathbf{x}' - \mathbf{y}; s) C(\mathbf{x}, \mathbf{y}) \, d\mathbf{x}' \\ &= -\sigma_0^2 \int_{\mathbb{R}^2} \psi(\mathbf{x}' - \mathbf{x}; s) \psi(\mathbf{x}' - \mathbf{y}; s) \log\left(\|\mathbf{x} - \mathbf{y}\|\right) \, d\mathbf{x}' \, . \end{aligned}$$

We can write the probability density function for the Brownian image $\beta(\mathbf{x})$ explicitly as

$$p(\langle \beta, \phi \rangle) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma_0^2} \int_\Omega \| \langle \beta, \nabla \phi \rangle \|^2 \, d\mathbf{x} \right) \, . \tag{5.22}$$

It is easier to get an understanding of this probability density function in the discrete case. For discrete images one could write the gradient magnitude in terms of forward differences among neighboring pixels (4-connected neighborhood). The discrete version of Eq. (5.22) therefore is

$$p(\mathbf{I}) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma_0^2} \sum_{i=1}^{N} \sum_{j=1}^{M} \left((I_{i+1,j} - I_{i,j})^2 + (I_{i,j+1} - I_{i,j})^2 \right) \right) \, . \tag{5.23}$$

Now we see the connection between the probability density and the definition of Brownian images in Eq. (5.18). That is, the increments of Brownian images are Gaussian distributed.

The Brownian image inherits the properties of the zero mean Gaussian stochastic function described in the previous section. The pixel statistics of the discrete Brownian image is Gaussian with density given by Eq. (5.3) and the same holds true for derivatives or pixel differences (see Eq. (5.11) and Eq. (5.12)). Following a similar argument as for the Gaussian white noise model we get that the gradient magnitude of the Brownian image is Rayleigh distributed and given by Eq. (5.16). Empirical evidence of these facts are shown in Fig. 5.6.

All partial derivatives of a Brownian image are Gaussian distributed, hence the distribution of image points mapped into k-jet space is a zero mean Gaussian. The covariance matrix for the zero mean Brownian image in k-jet space can be calculated analytically (see appendix in Sec. 5.6 for the derivation)

$$\begin{aligned} &E\left[\langle B_{x^{n_1} y^{m_1}}(\mathbf{x}; s), B_{x^{n_2} y^{m_2}}(\mathbf{x}; s) \rangle \right] \\ &\qquad = (-1)^{\frac{n+m}{2} + n_2 + m_2} \frac{\sigma_0^2 n! m!}{2^{n+m+1} \pi s^{n+m} (n+m)(n/2)!(m/2)!} \end{aligned} \tag{5.24}$$

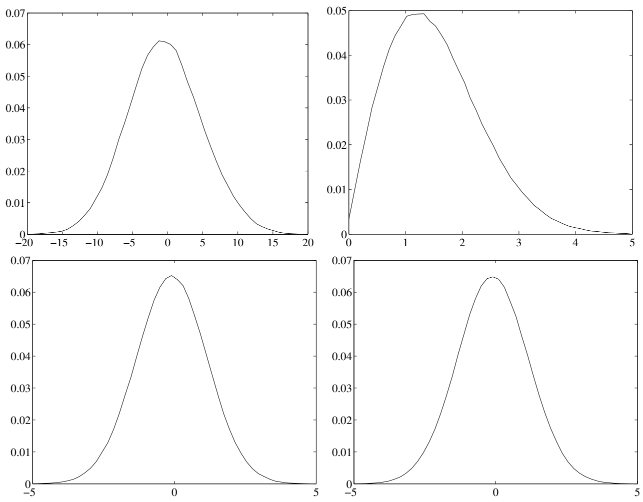


Figure 5.6: (Top left) Pixel histogram $p(B)$ of a Brownian scale-space image at scale $s = 2$. (Top right) Histogram of the gradient magnitude $p(\|\nabla B\|)$, which is a close approximation to the Rayleigh distribution. (Bottom) Histograms $p(B_x)$ and $p(B_y)$ of partial derivatives in the x and y directions.

whenever both $n = n_1 + n_2$ and $m = m_1 + m_2$ are even integers, otherwise

$$E\left[\langle B_{x^{n_1}y^{m_1}}(\mathbf{x};s), B_{x^{n_2}y^{m_2}}(\mathbf{x};s)\rangle\right] = 0 \ .$$

As was the case for Gaussian white noise (Eq. (5.17)), we see that partial derivatives of Brownian images are correlated whenever n and m are even integers. If we use scale normalized derivatives based on dimensional analysis (Florack et al., 1994) we get rid of the scale dependency, $1/s^{n+m}$.

Interestingly, it turns out that the covariance structure of Brownian images in k-jet space, within some scale interval, is a good model of the covariance structure of natural images mapped into k-jet space. To be more precise, the eigenvalues of the covariance matrix of natural images and Brownian images in k-jet space are equivalent up to a multiplicative constant within a scale interval. Fig. 5.7 shows graphs of eigenvalues of estimated covariance matrices in 3-jet space for four classes of images; Brownian images, white noise, natural images[9], and logarithm of power spectrum of natural images. The jet space used here is $\tilde{\mathcal{J}}^3(\mathbb{R}^2 \mapsto \mathbb{R})$ which was discussed in Chapter 4 and defined by the map $\tilde{j}^3 B(\mathbf{x};s)$ of Eq. (4.9), which basically means that the zeroth order term of the corresponding Taylor expansion has

[9]All images from the van Hateren natural stimuli collection (van Hateren and van der Schaaf, 1998).

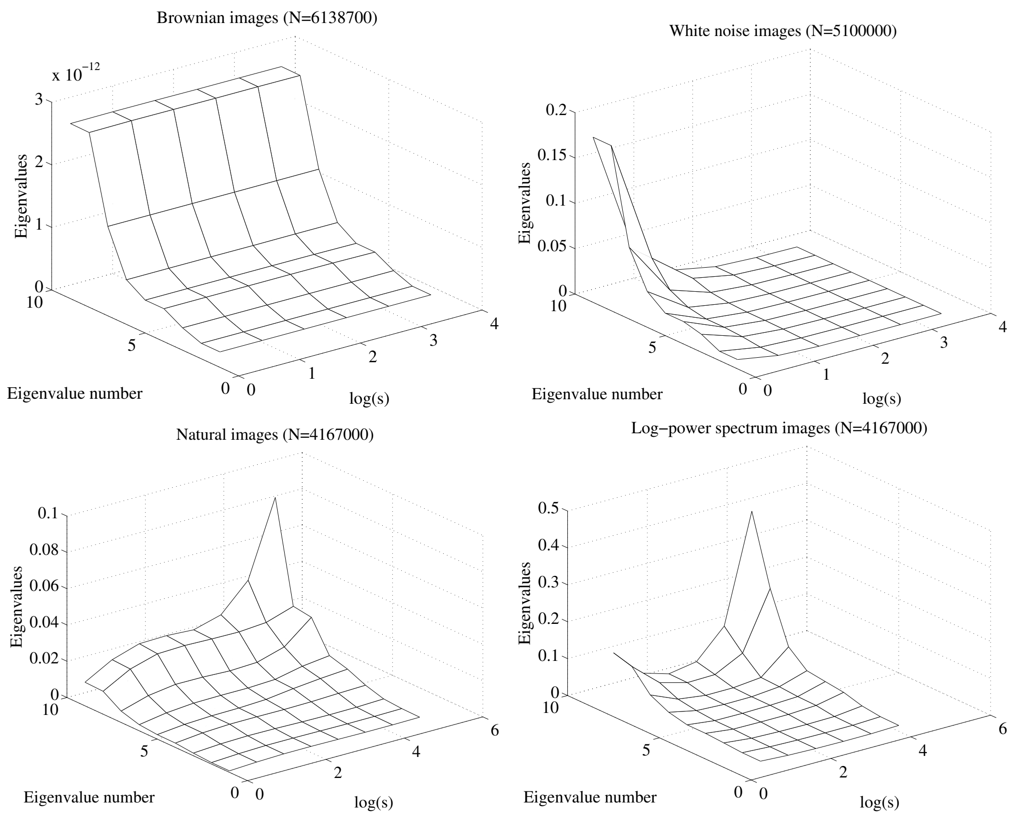


Figure 5.7: Eigenvalues across scale s of an empirically estimated covariance matrix $E\left[\langle (s^{n_1+m_1} B_{x^{n_1}y^{m_1}}), (s^{n_2+m_2} B_{x^{n_2}y^{m_2}})\rangle\right]$ of images mapped into 3-jet space ($n_1 + m_1 \leq 3$, $n_2 + m_2 \leq 3$ and omitting the zeroth order term) by scale-normalized scale-space derivatives. The top row show eigenvalues for Brownian and white noise images in 3-jet space. The bottom row shows eigenvalues for natural images and images of log-power spectrum of natural images. The size of each data set is given by the number N in parenthesis.

been thrown away. Data sets were produced by applying the scale normalized map $\tilde{j}^3 B(\mathbf{x}; s)$ to images at different scales s. Each data set has zero mean and the covariance matrices were estimated from these data sets. It can be seen from Fig. 5.7 that not surprisingly Brownian images have constant eigenvalues across scale. White noise has decreasing eigenvalues across scale due to this image models lack of scale invariance. Natural images have a plateau of constant eigenvalues in the scale interval between approximately $s = 4$ and $s = 32$.

As a control experiment I have included images of logarithm of power

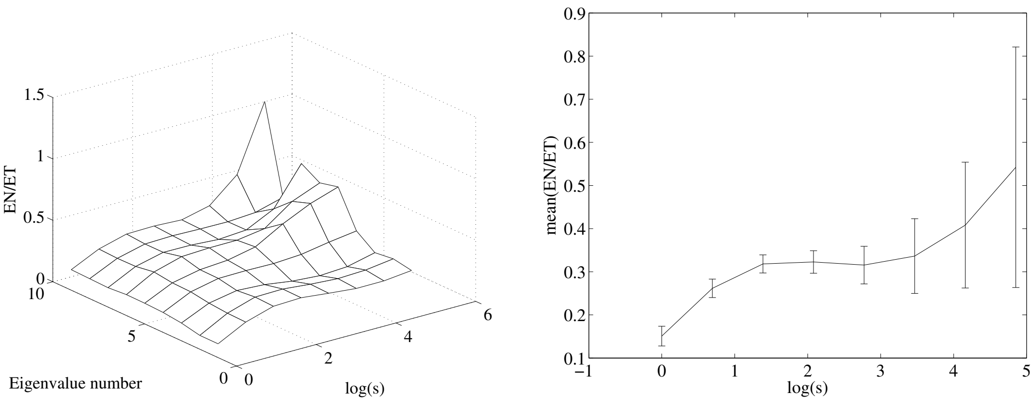


Figure 5.8: (Left) Eigenvalues of natural images $EN(s)$ at different scales s divided by the theoretical eigenvalues of Brownian images ET, EN/ET. (Right) mean value, mean(EN/ET), and standard deviation (error bars) of the 9 eigenvalue factors at each scale s.

spectra of natural images[10]. These images have a statistics that is very different from both natural and Gaussian images and may therefore be used to evaluate whether the results of constant eigenvalues for Brownian and natural images are an artifact of the method. It can be seen from Fig. 5.7 that log-power spectrum images do not have constant eigenvalues across scale.

Besides having constant eigenvalues in a scale interval, natural images seem to have the same relative interrelation among the eigenvalues as Brownian images. This can be seen from Fig. 5.8 (Left), which shows the graph of the natural image eigenvalues divided by the corresponding theoretical eigenvalues for Brownian images. Fig. 5.8 (Right) shows the mean value and standard deviation of the constant factor between eigenvalues across scales. We see that in the scale interval $s = 4$ to $s = 32$ the factor is nearly constant across eigenvalues and scale.

There might be several reasons why we do not have true scale invariance, i.e. invariance across all scales, for the eigenvalues of the covariance matrix of natural images. An image obviously has an inner and outer scale, bounded by the physical constraints of the camera. The lower bound is given by the pixel size, but the actual inner scale might be larger than this lower bound since the inner scale is equivalent to the size of the smallest discernable objects in the image. Similarly, the outer scale is often smaller than the physical size of the image and is given by the scale of the largest discernable objects.

[10] Again I use the van Hateren natural stimuli collection (van Hateren and van der Schaaf, 1998).

The actual inner and outer scales reflect the tendency of the photographer to take pictures of discernable objects, hence introducing a bias in represented object scales. In both Fig. 5.7 (Bottom left) and Fig. 5.8 we see that for small scales ($1/2 \leq s < 4$) the eigenvalues increase with scale. This might be due to a lack of sufficient small scale structure in the van Hateren image database. That is, $s < 4$ might be below the average inner scale of van Hateren images. Similarly, at large scales ($s > 32$) the eigenvalues increase again, which might be due to the scale being larger than the average outer scale of van Hateren images.

The eigenvectors of the covariance matrix in k-jet space are the same for all types of images and do not depend on the scale, since the covariance matrix is build by inner products of partial derivatives of the Gaussian scale-space aperture (see Fig. 5.9, Eq. (5.17), and Eq. (5.24)).

The connection between the covariance matrices for Brownian and natural images in jet space comes from the scale invariance property of the second order statistics of both types of images. The Brownian image model captures the second order statistics of natural images. That is, we can use the Brownian image model as an analytical model of the covariance structure of natural images.

5.4 The Edge Manifold and Brownian Images

Since the Brownian image is a scale invariant Gaussian model of images, which produces a Gaussian distribution in jet space, we would expect that if we do the experiments (with whitening) outlined in Chapter 4 and in Pedersen and Lee (2002) on Brownian images, we would see a uniform (and scale invariant) distribution on the 8-sphere and around the manifold of edges. And indeed, this is what we see in the results. The results included here are based on Brownian data whitened by the theoretical covariance matrix described in Appendix 5.6.

Fig. 5.10 shows a log-log plot of the density, Eq. (4.26), of Brownian images in jet space as a function of distance to the manifold of edges. Fig. 5.11 (bottom) shows the cumulative volume versus the cumulative number of data points. From these graphs it is clear that the Brownian image model is uniformly distributed on the surface of the 8-sphere.

The density shown in Fig. 5.10 is nearly flat, but at approximately 60 degrees the density estimate suddenly goes up towards infinity. This behavior is due to both the choice of histogram bin size and lack of sufficient statistics in both the volume estimate and the dataset. Fig. 5.12 shows both the normalized histogram of the dataset (top) and the bin volume estimate

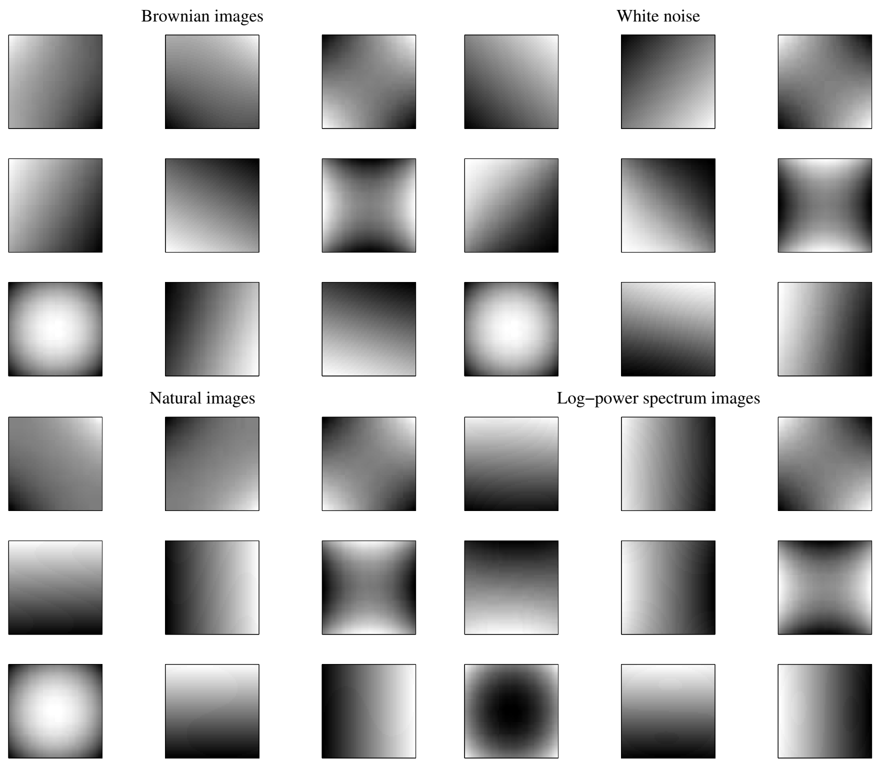


Figure 5.9: Eigen-images of the empirical covariance matrices in 3-jet space of Fig. 5.7. Each eigen-image is generated by drawing the Taylor series corresponding to the 3-jet eigenvector within one scale unit.

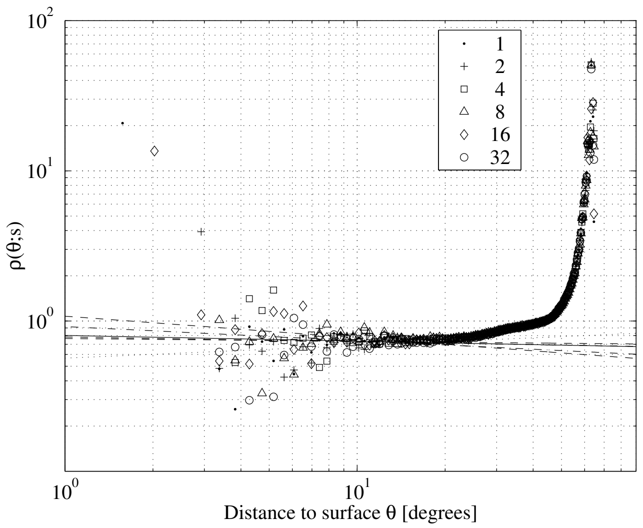


Figure 5.10: Log-log plot of the density as defined by equation Eq. (4.26) in Chapter 4. Whitening based on theoretical covariance matrix.

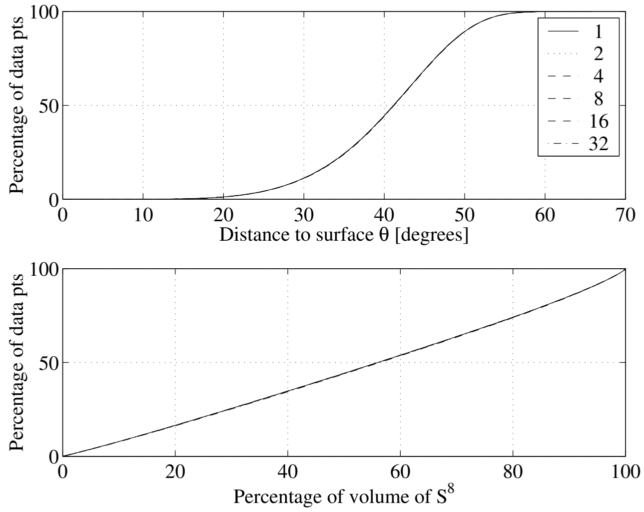
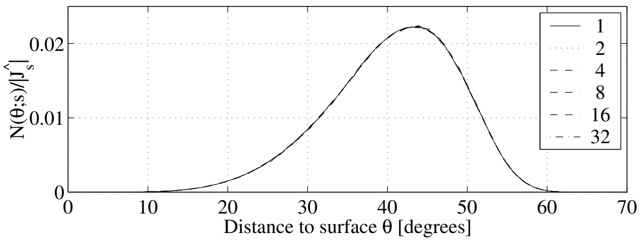


Figure 5.11: (Top) Cumulative sum $\sum_{\beta \leq \theta} N(\beta; s)/|\hat{J}_s|$ (in percent) of the number of data points in $\hat{J}_s$ as a function of the distance to the manifold of edges. (Bottom) Cumulative volume versus cumulative number of data points for data sets $\hat{J}_s$.



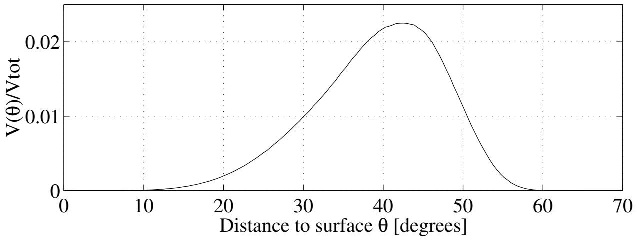


Figure 5.12: (Top) Histogram of the Brownian surface data set ($6.1 \cdot 10^6$ data points) with respect to distance to the edge manifold. (Bottom) Histogram of the Monte Carlo estimate of the surface volume of the 8-sphere.

(bottom) used in the density estimate. It is clear that the volume estimate decreases rapidly towards zero at 60 degrees, which implies that the bins are small and a lot more samples are therefore needed to estimate the bin volume by a Monte Carlo integration. I therefore suspect that the bin volume is underestimated for bins after 60 degrees. This underestimated volume would of course lead to a false increase in the density estimate. Evidence of this problem is also seen in the experiments on natural images in Chapter 4 especially in Fig. 4.3.

5.5 Summary

In this chapter I have discussed which statistical properties of natural images may be modeled by Gaussian image models, including both the Gaussian white noise image model and the Gaussian Brownian image model.

The conclusion is that the Gaussian white noise model is a poor model of the statistics of natural images. On the other hand the Brownian image model is a model which captures the scale invariant covariance structure of natural images. I have shown that the covariance structure of natural images in k-jet space corresponds to that of Brownian images within a scale interval. Therefore the Brownian image model could be thought of as a least committed scale invariant model of the second order statistics of natural images, in the sense that the Gaussian distribution is the maximum entropy solution given only constraints on the mean and covariance structure (see Jaynes (1957; 1968) and Nielsen (1995)).

Since the Brownian image model gives a uniform distribution around the edge manifold of Chapter 4 and Pedersen and Lee (2002), contrary to what we see for natural images, one might conclude that natural images have more complex dependencies among pixel values than Brownian images. That is, the edge manifold model captures a variation in natural images which is special for this class of images. This also highlights one of the weaknesses of the Brownian image model with respect to being a model of natural images. The Brownian image model does not capture the higher order statistics of natural images.

5.6 Appendix I: Analytical Expression for Covariance Matrix in Jet Space

Preliminaries

We will derive the expectation value $E[\cdot]$ of the inner product of two linear functionals of a stationary stochastic function f, $\phi_1 * f$ and $\phi_2 * f$, that is the correlation between these two functionals. Let the inner product be defined as

$$\langle(\phi_1 * f), (\phi_2 * f)\rangle \equiv \iint (\phi_1 * f) \cdot (\phi_2 * f)\, dx\, dy$$

and its expectation value as

$$E[\langle(\phi_1 * f), (\phi_2 * f)\rangle] = E\left[\iint (\phi_1 * f) \cdot (\phi_2 * f)\, dx\, dy\right] .$$

We can simplify this expectation value by going into Fourier space using Plancherel's theorem, $\int_{\mathbb{R}^d} f(x)\overline{g(x)}dx = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi$,

$$\begin{aligned} E\left[\iint (\phi_1 * f) \cdot (\phi_2 * f)\, dx\, dy\right] &= E\left[\frac{1}{(2\pi)^2}\iint \hat{\phi}_1\overline{\hat{\phi}_2}|\hat{f}(\xi,\eta)|^2\, d\xi\, d\eta\right] \\ &= \frac{1}{(2\pi)^2}\iint \hat{\phi}_1\overline{\hat{\phi}_2} E\left[|\hat{f}(\xi,\eta)|^2\right]\, d\xi\, d\eta\, . \end{aligned} \tag{5.25}$$

We will now look at the covariance of the stochastic function $f(x,y)$ in jet scale-space. Due to Eq. (5.25) the covariance matrix of a zero mean stationary stochastic function $f(x,y)$ mapped into jet space can be written as

$$\begin{aligned} &E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * f), (\partial_{x^{n_2}y^{m_2}}\psi * f)\rangle\right] \\ &\quad = E\left[\iint (\partial_{x^{n_1}y^{m_1}}\psi * f) \cdot (\partial_{x^{n_2}y^{m_2}}\psi * f)\, dx\, dy\right] \\ &= \frac{1}{(2\pi)^2}\iint (i\xi)^{n_1}(i\eta)^{m_1}(-i\xi)^{n_2}(-i\eta)^{m_2} e^{-s^2(\xi^2+\eta^2)} E\left[|\hat{f}(\xi,\eta)|^2\right]\, d\xi\, d\eta \\ &= \frac{1}{(2\pi)^2}\iint (-1)^{n_2+m_2} i^{n+m}\xi^n\eta^m e^{-s^2(\xi^2+\eta^2)} E\left[|\hat{f}(\xi,\eta)|^2\right]\, d\xi\, d\eta \end{aligned} \tag{5.26}$$

where $n = n_1 + n_2$ and $m = m_1 + m_2$.

Covariance Matrix of White Noise in Jet Space

Let the stochastic function $\nu(x,y)$ be a white noise image with variance σ^2 and power spectrum $E[|\hat{\nu}(\xi,\eta)|^2] = \sigma^2$. The expectation value of Eq. (5.26) is therefore,

$$
\begin{aligned}
&E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * \nu),(\partial_{x^{n_2}y^{m_2}}\psi * \nu)\rangle\right] \\
&\qquad = \frac{\sigma^2}{(2\pi)^2}\iint (-1)^{n_2+m_2} i^{n+m}\xi^n\eta^m e^{-s^2(\xi^2+\eta^2)}\,d\xi\,d\eta \\
&\qquad = (-1)^{n_2+m_2} i^{n+m}\frac{\sigma^2}{(2\pi)^2}\int_{-\infty}^{\infty}\xi^n e^{-s^2\xi^2}\,d\xi\int_{-\infty}^{\infty}\eta^m e^{-s^2\eta^2}\,d\eta\ .
\end{aligned}
$$

If we now use that

$$
\int_{-\infty}^{\infty}\xi^n e^{-s^2\xi^2}\,d\xi = \begin{cases} \frac{1}{s^{1+n}}\Gamma\left(\frac{1+n}{2}\right) & \text{if } n \text{ even} \\ 0 & \text{otherwise} \end{cases}
$$

and

$$
\Gamma\left(\frac{1+n}{2}\right) = \frac{n!}{2^n(n/2)!}\sqrt{\pi}
$$

we get that

$$
\begin{aligned}
&E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * \nu),(\partial_{x^{n_2}y^{m_2}}\psi * \nu)\rangle\right] \\
&\qquad\qquad = (-1)^{\frac{n+m}{2}+n_2+m_2}\frac{\sigma^2}{\pi s^{2+n+m}}\frac{n!m!}{2^{n+m+2}(n/2)!(m/2)!}
\end{aligned}
$$

whenever both n and m are even integers, otherwise

$$
E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * \nu),(\partial_{x^{n_2}y^{m_2}}\psi * \nu)\rangle\right] = 0\ .
$$

Covariance Matrix of Brownian Images in Jet Space[11]

Let the stochastic function $\beta(x,y)$ be a Brownian image, i.e. the power spectrum of $\beta(x,y)$ is $E[|\hat{\beta}(\xi,\eta)|^2] = \sigma_0^2/\|\xi\|^2$ where σ_0^2 is a constant which can be interpreted as a global variance offset. Hence the expectation value of Eq. (5.26) can be written as

$$
\begin{aligned}
&E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * \beta),(\partial_{x^{n_2}y^{m_2}}\psi * \beta)\rangle\right] \\
&\qquad = \frac{1}{(2\pi)^2}\iint (-1)^{n_2+m_2} i^{n+m}\xi^n\eta^m e^{-s^2(\xi^2+\eta^2)}\frac{\sigma_0^2}{\xi^2+\eta^2}\,d\xi\,d\eta\ .
\end{aligned}
$$

[11]This derivation was developed in discussion with D. Mumford.

If we introduce polar coordinates $(\|\xi\|, \theta)$ instead of (ξ, η) we get

$$\frac{\sigma_0^2}{(2\pi)^2} \iint \frac{(-1)^{n_2+m_2} i^{n+m} (\|\xi\| \cos\theta)^n (\|\xi\| \sin\theta)^m}{\|\xi\|^2} e^{-s^2\|\xi\|^2} \|\xi\| \, d\|\xi\| \, d\theta$$
$$= (-1)^{n_2+m_2} i^{n+m} \frac{\sigma_0^2}{(2\pi)^2} \int_0^\infty \|\xi\|^{n+m-1} e^{-s^2\|\xi\|^2} \, d\|\xi\|$$
$$\cdot \int_0^{2\pi} (\cos\theta)^n (\sin\theta)^m \, d\theta.$$

Notice that

$$\int_0^{2\pi} (\cos\theta)^n (\sin\theta)^m \, d\theta = \begin{cases} 0, \text{if } n \text{ or } m \text{ is odd} \\ \frac{[(n-1)(n-3)\cdots 1]\cdot[(m-1)(m-3)\cdots 1]}{(n+m)(n+m-2)\cdots 2} 2\pi \end{cases}$$

and

$$\begin{aligned} \int_0^\infty r^{n+m-1} e^{-s^2 r^2} \, dr &= \int_0^\infty \frac{1}{s} \left(\frac{t}{s}\right)^{n+m-1} e^{-t^2} \, dt \\ &= \frac{1}{s^{n+m}} \int_0^\infty t^{n+m-1} e^{-t^2} \, dt \\ &= \frac{1}{s^{n+m}} \int_0^\infty h^{\frac{n+m-1}{2}} e^{-h} \frac{1}{2h^{1/2}} \, dh \\ &= \frac{1}{2s^{n+m}} \int_0^\infty h^{\frac{n+m}{2}-1} e^{-h} \, dh \\ &= \frac{1}{2s^{n+m}} \Gamma\left(\frac{n+m}{2}\right) \end{aligned}$$

by substitution with $t = sr$ followed by $h = t^2$. Using $\Gamma(n+1) = n!$ when n is an integer, this can be written as

$$\begin{aligned} \frac{1}{2s^{n+m}} \Gamma\left(\frac{n+m}{2}\right) &= \frac{1}{2s^{n+m}} \left(\frac{n+m}{2} - 1\right) \left(\frac{n+m}{2} - 2\right) \cdots 1 \\ &= \frac{1}{2^{\frac{n+m}{2}} s^{n+m}} (n+m-2)(n+m-4) \cdots 2. \end{aligned}$$

We can now write the expectation value of interest as

$$E\left[\langle(\partial_{x^{n_1}y^{m_1}} \psi * \beta), (\partial_{x^{n_2}y^{m_2}} \psi * \beta)\rangle\right] =$$
$$(-1)^{n_2+m_2} i^{n+m} \frac{\sigma_0^2}{2\pi s^{n+m}} \frac{[(n-1)(n-3)\cdots 1]\cdot[(m-1)(m-3)\cdots 1]}{2^{\frac{n+m}{2}}(n+m)}$$
$$= (-1)^{\frac{n+m}{2}+n_2+m_2} \frac{\sigma_0^2}{2\pi s^{n+m}} \frac{n!m!}{2^{n+m}(n+m)(n/2)!(m/2)!}$$

whenever both n and m are even integers, otherwise

$$E\left[\langle(\partial_{x^{n_1}y^{m_1}}\psi * \beta), (\partial_{x^{n_2}y^{m_2}}\psi * \beta)\rangle\right] = 0 \ .$$

If we use dimensionless scale-normalized derivatives (Florack et al., 1994) we will get rid of the scale factor $1/s^{n+m}$.

As an example, let us write the covariance matrix of Brownian images mapped into 3-jet space ($n_1 + m_1 \leq 3$, $n_2 + m_2 \leq 3$ and omitting the zeroth order term) by scale-normalized scale-space derivatives,

$$E\left[\langle(s^{n_1+m_1}\partial_{x^{n_1}y^{m_1}}\psi * \beta), (s^{n_2+m_2}\partial_{x^{n_2}y^{m_2}}\psi * \beta)\rangle\right] =$$
$$\frac{\sigma_0^2}{2\pi}\cdot\begin{pmatrix} 1/4 & 0 & 0 & 0 & 0 & -3/16 & 0 & -1/16 & 0 \\ 0 & 1/4 & 0 & 0 & 0 & 0 & -1/16 & 0 & -3/16 \\ 0 & 0 & 3/16 & 0 & 1/16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/16 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/16 & 0 & 3/16 & 0 & 0 & 0 & 0 \\ -3/16 & 0 & 0 & 0 & 0 & 5/16 & 0 & 1/16 & 0 \\ 0 & -1/16 & 0 & 0 & 0 & 0 & 1/16 & 0 & 1/16 \\ -1/16 & 0 & 0 & 0 & 0 & 1/16 & 0 & 1/16 & 0 \\ 0 & -3/16 & 0 & 0 & 0 & 0 & 1/16 & 0 & 5/16 \end{pmatrix} .$$

5.7 Appendix II: Scale Invariance and Self Similarity

Scale invariance is a special case of self similarity. A function is self similar if an affine transformation exists, such that the function is invariant with respect to that transformation. Let us state this more formally by writing a definition of self similarity (Mallat, 1998):

Definition 5.1 (Self similarity of functions)
Let f be a continuous function with a compact support $S \subset \mathbb{R}$. We say that f is self similar if disjoint subsets $S_1, \ldots, S_k$ exist such that the graph of f restricted to each S_i is an affine transformation of f. This means that a scale λ_i, a translation r_i, a weight p_i and a constant c_i exist such that

$$\forall x \in S_i \ , \ f(x) = c_i + p_i f(\lambda_i(x - r_i)) \ . \tag{5.27}$$

A function $f(x)$ is scale invariant if a rescaling with scale parameter λ has no effect, that is $f(x) = f(\lambda x)$.

5.7.1 Statistical Self Similarity

We can also define a statistical version of self similarity, where one requires that the moments of a stochastic function $f(x)$ or the probability distribution of this function is self similar. The definition of statistical self similarity is a weaker formulation than deterministic self similarity of definition 5.1. Nevertheless, statistical self similarity is a much more useful concept in the context of natural image statistics.

Definition 5.2 (Equality in law)
The notation $\{f(t)\} \stackrel{\triangle}{=} \{g(t)\}$ *means that the two stochastic functions* $f(t)$ *and* $g(t)$ *have the same probability distribution function.*

Equality in law is a weak definition of stationarity of the stochastic function. I assume stationarity of the stochastic function and only define statistical self similarity in terms of scaling transformations.

Definition 5.3 (Statistical self similarity)
The stochastic function $f(t)$ *is said to be self similar (SS) with scaling exponent* h *if for any* $\tau > 0$ *and* t

$$\{f(t)\} \stackrel{\triangle}{=} \{\tau^{-h} f(\tau t)\} \; .$$

Part II

Motion Analysis

Chapter 6

Optic Flow by Scale-Space Integration of Normal Flow

6.1 Introduction

Motion analysis is a large topic within computer vision and image analysis, because knowledge of motion, as perceivable in image sequences, is necessary for various tasks such as object tracking, time-to-contact problems, and structure from motion. Motion analysis is conducted by associating a vector field of velocities to the image sequence, which describes the rate and direction of change of the intensity values. We define the *optic flow* as the velocity field, which describes the temporal changes of the intensity values of the sequence. Optic flow is not equivalent to a projection of the true physical motion onto the image plane. There are several reasons for this, one being the changing intensity values caused by reflections and varying lighting conditions, but the main reason is the so-called *aperture problem*. That is, we can only deduce the motion that results in a change in the intensity patterns of the captured image sequence. The problem of determining the motion along image isophotes (i.e. iso-intensity curves) is inherently ambiguous. By introducing assumptions about the structure of the flow we can obtain various types of realizations of optic flow. Usual assumptions used in other methods are local or global rigidity of objects and affine motion. The only unbiased local motion we can obtain is the motion orthogonal to the isophotes of the image. This type of velocity field is usually called the *normal flow*.

A large variety of methods exist for estimation of optic flow (Alvarez et al., 2000; Arnspang, 1988; Florack et al., 1998; Horn and Schunck, 1981; Lindeberg, 1995; Otte and Nagel, 1994) (which is far from an exhaustive list). See also Barron et al. (Barron et al., 1994) for a discussion and evaluation of various methods.

We wish to develop an algorithm, which can be used for doing motion analysis in experimental fluid dynamics (see Pedersen (2000)). For that reason it is important that it is least committed, in order to reduce possible biases in the approximated flow. Since normal flow is the only unbiased or least committed realization of optic flow, we will use it as a basis for obtaining an estimate of the full flow. We believe that a least committed estimate can be obtained by locally modeling the full optic flow on top of the normal flow. Ideally we seek to estimate the optic flow field which at each point has the corresponding normal flow vector as its normal component[1]. This constraint only fix one degree of freedom (d.o.f.), which means that an infinite number of solutions exist. In order to circumvent this problem, we introduce a polynomial model of the local optic flow. We use a linear Gaussian scale-space to formalize the concept of locality or the integration scale of our model. This integration scale-space regularizes the problem and lets us fix the missing tangential d.o.f. by estimating, in a least squares sense, the parameters of our local model constrained by the normal flow found at the integration scale or put in another way, in the region of model validity. The solution to this constrained minimization problem can be stated in closed form and it is expressed in terms of a Taylor expansion of the normalized structure tensor. In order to take full advantage of this multi-scale approach we use an automatic local scale selection mechanism, based on a method by Niessen and Maas (1997), to select the scale of model validity.

In this paper we compute the normal flow by using the method proposed by Florack et al. (1998). It is an incorporation of the so-called *Optic Flow Constraint Equation* (OFCE), originally proposed by Horn and Schunck (1981), into the linear Gaussian scale-space formalism. In this method the normal flow is in general modeled by an M'th order polynomial, but we choose the, at first glance restrictive, zeroth order model. The zeroth order model produce flow vectors which are normals to the isophotes, giving a normal flow field in the usual sense. Higher order models give flow vectors which are not necessarily normals to the isophotes. Other choices of normal flow methods are possible in the setting of the method we describe in this paper.

We believe that scale-space integration of normal flow and modeling locally the optic flow in this scale-space is different from other methods. The conscious inclusion of measurement scale lets us select the appropriate neighborhood in which our local model is valid. Other authors (Arnspang,

[1]By normal component we mean the projection of the sought optic flow onto the direction orthogonal to the local isophote.

1988; Florack et al., 1998; Lindeberg, 1995; Otte and Nagel, 1994) have introduced polynomial models of the flow structure directly into the OFCE, contrary to modeling the optic flow on top of the normal flow as we propose. The use of an integration neighborhood for the approximation of optic flow and in the related stereo matching problem has been used by several authors (Alvarez et al., 2000; Florack et al., 1998; Kanade and Okutomi, 1994; Lindeberg, 1995; Otte and Nagel, 1994).

We evaluate the performance of our algorithm by using the methods proposed by Barron et al. (1994) in their survey of performance of various optic flow methods. Our performance results will be compared with the results for other optic flow techniques applied among other on a synthetic fire sequence. In order to show a possible application of our method, we also apply it on a real sequence of smoke circulation in a pigsty.

The organization of this paper is the following: In Sec. 6.2 we describe our scale-space representation as well as some notation. Sec. 6.3 is a brief introduction to the theory of normal flow. The proposed method for computation of optic flow is presented in Sec. 6.4 and finally, the performance of the presented algorithm is discussed in Sec. 6.5.

6.2 Spatiotemporal Gaussian Scale-Space

Both Koenderink (1988) and Lindeberg (1997) have proposed causal temporal scale-space representations. In this paper we choose simply to disregard temporal causality, since we are not interested in real-time applications. It is our opinion that a generalization of our method to one of the above mentioned causal temporal representations is feasible. We choose instead to use the linear Gaussian scale-space representation, introduced among others by Koenderink (1984).

The spatiotemporal scale-space representation $L(\mathbf{x};\sigma,\tau) : \mathbb{R}^{N+1} \times \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}$ of the spatiotemporal image $f(\mathbf{x}) : \mathbb{R}^{N+1} \mapsto \mathbb{R}$ is defined by the convolution of the image with the *scale-space aperture function* $G(\mathbf{x};\sigma,\tau) : \mathbb{R}^{N+1} \times \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}$,

$$L(\mathbf{x};\sigma,\tau) = f(\mathbf{x}) * G(\mathbf{x};\sigma,\tau)\,. \tag{6.1}$$

The aperture function is the spatiotemporal Gaussian

$$G(\mathbf{x};\sigma,\tau) = \frac{1}{\sqrt{2\pi\tau^2}(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{x^i x^i}{2\sigma^2} - \frac{(x^t)^2}{2\tau^2}\right) \tag{6.2}$$

where σ and τ are called the *spatiotemporal scale parameters.* Throughout this paper, we will use the index notation, where x^i denotes the i'th spatial component of the vector $\mathbf{x}$, and x^t denotes the temporal component. When we are talking about both spatial and temporal components we write the component index with a Greek letter, e.g. x^μ. We will also sometimes make use of Einstein's summation convention, i.e. repeated lower and upper indices indicates summation $x^i x^i \equiv \sum_i x^i x^i$.

6.3 Theory of Normal Flow

We assume that we are given a method for computing the normal flow of an image sequence. The actual choice of method is arbitrary, but in this paper we use the scale-space OFCE proposed by Florack et al. (1998) as a method for computing normal flow. We will briefly outline this method in order to establish some necessary notation.

Horn and Schunck (1981) propose an optic flow method based on the local assumption that image intensities are preserved at points moving along the flow. Furthermore, it is assumed that the temporal velocity component is constant, $v^t(\mathbf{x}) \equiv 1$. This assumption expresses that the flow is everywhere non-vanishing and temporal causal. These assumptions lead to the well-known optic flow constraint equation (OFCE). It is interesting to notice that the later assumption breaks down generically in a countable set of points (Nielsen and Olsen, 1998). However in this paper we assume that this assumption is valid at every point in the image.

Florack et al. (1998) incorporate the ideas of Horn and Schunck into the scale-space paradigm, by defining the optic flow constraint equation under the spatiotemporal scale-space aperture. For a 2-dimensional image sequence $f(\mathbf{x})$, the OFCE can be written as the Lie derivative[2] under the scale-space aperture $G(\mathbf{x};\sigma,\tau)$ of the intensity function $f(\mathbf{x})$ along the flow $\mathbf{v}(\mathbf{x})$. We have

$$\mathcal{L}_{\mathbf{v}} L(\mathbf{x};\sigma,\tau) = \int_{\mathbf{x}\in\mathbb{R}^3} (f_t + v^x f_x + v^y f_y) G \, d\mathbf{x} \equiv 0 \tag{6.3}$$

where $f_\mu = \partial f / \partial x^\mu$.

Normal flow is defined by the so-called *normal flow constraint.* Florack et al. define the normal flow constraint under the scale-space aperture function as

$$\int_{\mathbf{x}\in\mathbb{R}^3} (v^x f_y - v^y f_x) G \, d\mathbf{x} \equiv 0 \,. \tag{6.4}$$

Furthermore, Florack et al. introduce a polynomial model of the flow $\mathbf{v}$ into Eq. (6.3) and Eq. (6.4), which lets them state the problem of normal flow

[2] The Lie derivative of a scalar function $f(\mathbf{x})$ is the derivative of $f(\mathbf{x})$ along the direction of a specified vector field $\mathbf{v}$, $\mathcal{L}_{\mathbf{v}} f = f_\mu v^\mu$.

estimation in terms of finding the solution to a system of linear equations. Each term of these equations can be expressed as a linear combination of scale-space image derivatives, due to the definition of Hermite polynomials in terms of Gaussian derivatives. As mentioned, we will only use the zeroth order model for approximation of normal flow, but we will use the M'th order polynomial model as our local model of the optic flow. We define the M'th order polynomial model $\mathbf{v}_M(\mathbf{x})$ of the vector field $\mathbf{v}(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$ as

$$v_M^\mu(\mathbf{x}) = \sum_{l=0}^{M} \frac{1}{l!} \tilde{v}^\mu_{\rho_1 \cdots \rho_l} x^{\rho_1} \cdots x^{\rho_l} \,. \tag{6.5}$$

Here x^ρ, v^μ denote the components of the spatiotemporal vectors $\mathbf{x}, \mathbf{v} \in \mathbb{R}^{N+1}$ and $\tilde{v}^\mu_{\rho_1 \cdots \rho_l}$ are the model coefficients approximating the partial derivatives of $\mathbf{v}(\mathbf{x})$ at $\mathbf{x} = \mathbf{0}$.

6.4 Least Committed Optic Flow

Ideally we want to obtain an optic flow field that has a normal component which is equal to the normal flow. This is an ill-posed problem, since an infinite number of solutions exist, due to our lack of knowledge about the tangential component. We intend to regularize this problem by locally modeling the tangential component of the optic flow field by a polynomial model, under the constraint that the normal component of the model should be close to the normal flow. We use a spatiotemporal integration scale-space to define the scale of validity of our local model.

This constrained optimization problem can be formalized by a functional $E(\mathbf{v}(\mathbf{x};\varpi))$ of the optic flow field $\mathbf{v}(\mathbf{x};\varpi)$ at integration scale ϖ. This functional describes the degree of discrepancy between the normal flow field $\mathbf{u}(\mathbf{x})$ and the normal component of the sought optic flow field $\mathbf{v}(\mathbf{x};\varpi)$ under the Gaussian integration aperture[3] $G(\mathbf{x};\varpi)$. We define the discrepancy as the least squares difference, thus we can write the functional as

$$E(\mathbf{v}(\mathbf{x};\varpi)) = \int_{\mathbf{x}\in\mathbb{R}^{N+1}} w(\mathbf{x}) \left\| \eta(\mathbf{v}\cdot\eta) - \mathbf{u} \right\|^2 G(\mathbf{x};\varpi)\, d\mathbf{x} \,, \tag{6.6}$$

where $\eta(\mathbf{x}) = \frac{\mathbf{u}}{\|\mathbf{u}\|}$ is the normalized direction of the normal flow $\mathbf{u}(\mathbf{x})$. The function $w(\mathbf{x})$ is a measure of the uncertainty of the estimates of the underlying normal flow and acts as a weight, which penalizes poorly estimated

[3] For the sake of simplicity we choose to use a scale-space representation with only one integration scale parameter ϖ. The spatiotemporal scale-space aperture function of Eq. (6.2) can readily be interchanged with the aperture function $G(\mathbf{x};\varpi)$.

normal flow vectors. In Sec. 6.5 we choose to use the numerical uncertainty of the normal flow as the penalty function $w(\mathbf{x})$, but other choices are of course possible.

We introduce the M'th order polynomial model $\mathbf{v}_M$, defined in Eq. (6.5), into the functional $E(\mathbf{v}(\mathbf{x};\varpi))$. The model parameters $\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}$ can be obtained by minimizing the functional $E(\mathbf{v}_M(\mathbf{x};\varpi))$, that is

$$\hat{v}^{\mu}_{\rho_1\ldots\rho_l} = \arg\min_{\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}} E(\mathbf{v}_M)\ ,\ \ l = 0,\ldots,M\,. \tag{6.7}$$

The integrand of the functional $E(\mathbf{v}_M)$ is quadratic, which trivially means that a global minimum exists and that this minimum $\hat{v}^{\mu}_{\rho_1\ldots\rho_l}$ is the solution to the system of differential equations given by $\partial E/\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l} = 0$. We therefore arrive at:

Result 6.1 (*M*'th Order Optic Flow)
The M'th order optic flow approximation $\mathbf{v}_M(\mathbf{x};\varpi)$ is given by Eq. (6.5), where the model parameters $\tilde{v}^{\beta}_{\nu_1\ldots\nu_i}$ are obtained by solving the minimization problem of Eq. (6.7). The solution is given by the following set of linear equations

$$\frac{\partial E}{\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}} = \sum_{i=0}^{M}\frac{1}{i!}\tilde{v}^{\beta}_{\nu_1\ldots\nu_i}\int_{\mathbf{x}\in\mathbb{R}^{N+1}} w(\mathbf{x})\eta^{\mu}\eta^{\beta}\mathcal{M}_{\nu_1\ldots\nu_i\rho_1\ldots\rho_l}(\mathbf{x};\varpi)\,d\mathbf{x}$$
$$-\int_{\mathbf{x}\in\mathbb{R}^{N+1}} w(\mathbf{x})u^{\mu}\mathcal{M}_{\rho_1\ldots\rho_l}(\mathbf{x};\varpi)\,d\mathbf{x} = 0\,, \tag{6.8}$$

where $\mathcal{M}_{\rho_1\ldots\rho_l}(\mathbf{x};\varpi) = x^{\rho_1}\cdots x^{\rho_l}G(\mathbf{x},\varpi)$ is the l'th order mixed Gaussian monomials at integration scale ϖ.

Proof: We seek the solution of

$$\frac{\partial E}{\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}} = 0\,. \tag{6.9}$$

By using the chain rule $\partial E/\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l} = \int \partial q/\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}\partial F/\partial q\,d\mathbf{x}$, where $F(\mathbf{v}_M)$ is the integrand of $E(\mathbf{v}_M)$ and $q = \mathbf{v}_M\cdot\eta$, we get

$$\frac{\partial E}{\partial\tilde{v}^{\mu}_{\rho_1\ldots\rho_l}} = \int_{\mathbf{x}\in\mathbb{R}^{N+1}}\frac{2w(\mathbf{x})}{l!}\cdot\left(\eta^{(\alpha)^2}\eta^{\mu}\eta^{\beta}\left(\sum_{i=0}^{M}\frac{1}{i!}\tilde{v}^{\beta}_{\nu_1\ldots\nu_i}x^{\nu_1}\cdots x^{\nu_i}\right)\right.$$
$$\left. -\,\eta^{\mu}(\eta\cdot\mathbf{u})\right)\cdot x^{\rho_1}\cdots x^{\rho_l}G(\mathbf{x};\varpi)\,d\mathbf{x} = 0\,. \tag{6.10}$$

Notice that $\eta^{(\alpha)^2} = 1$, $\eta^\mu(\eta \cdot \mathbf{u}) = u^\mu$, and that the constant $2/l!$ can be removed. If we introduce the notation $\mathcal{M}_{\rho_1\ldots\rho_l}(\mathbf{x};\varpi) = x^{\rho_1}\cdots x^{\rho_l}G(\mathbf{x},\varpi)$, this system of linear equations can be written as

$$\frac{\partial E}{\partial \tilde{v}^\mu_{\rho_1\ldots\rho_l}} = \sum_{i=0}^{M} \frac{1}{i!}\tilde{v}^\beta_{\nu_1\ldots\nu_i} \int_{\mathbf{x}\in\mathbb{R}^{N+1}} w(\mathbf{x})\Big(\eta^\mu\eta^\beta\mathcal{M}_{\nu_1\ldots\nu_i\rho_1\ldots\rho_l}(\mathbf{x};\varpi) - u^\mu\mathcal{M}_{\rho_1\ldots\rho_l}(\mathbf{x};\varpi)\Big)\,d\mathbf{x} = 0\,. \quad (6.11)$$

□

We developed our method at the origin of the vector field $\mathbf{v}(\mathbf{x}=\mathbf{0})$, which means that in general a translation to the point of interest should be introduced, which in turn transforms the integrals of Result 6.1 into convolution integrals.

The l'th order mixed Gaussian monomials $\mathcal{M}_{\rho_1\ldots\rho_l}(\mathbf{x};\varpi)$ can be expressed in terms of linear combinations of the partial derivatives of the Gaussian by using the definition of Hermite polynomials and the separability of the Gaussian (see e.g. Florack et al. (1998) or Pedersen (2000)). This lets us interpret the two integrals of Result 6.1 as a set of scale-space derivatives of the normal flow $\mathbf{u}$ and the matrix $\eta^\mu\eta^\beta$.

Since the normal flow is defined to be parallel to the gradient direction, the matrix $(w\eta^\mu\eta^\beta) * G(\mathbf{x};\varpi)$ of Result 6.1 can be interpreted as the normalized spatiotemporal structure tensor (see e.g. Weickert (1998)). This gives another possible interpretation of Result 6.1, namely that we are seeking the flow field which, by the product with the normalized structure tensor, is equal to the normal flow. The introduction of our flow model corresponds to a Taylor expansion of the normalized structure tensor. We should mention that other authors have used the structure tensor as the basis of methods for doing motion analysis (Granlund and Knutsson, 1995).

The structure of image sequences exists on different scales and consequently the optimal measurement scale will vary between different regions in an image sequence. In conjunction with optic flow and the related stereo vision problem different approaches to the problem of scale selection has been taken; Kanade and Okutomi (1994), Weber and Malik (1995), Gårding and Lindeberg (1996), Niessen and Maas (1997), and Nielsen et al. (1999).

In this paper an automatic local scale selection method will be used similar to the one proposed by Niessen and Maas (1997). They propose to use the numerical stability as a criterion for scale selection. As a measure of numerical stability they use the Frobenius norm $\|\mathbf{A}^{-1}\|_F^2$ of the system of linear equations $\mathbf{A}\mathbf{v} = \mathbf{b}$ used in the method by Florack et al. (1998).

The Frobenius norm of a matrix is the sum of the singular values of that matrix. Since we use the Florack et al. multi-scale normal flow method, the outcome of Result 6.1 is a spatiotemporal integration scale-space of the vector field $\mathbf{v}_M(\mathbf{x};\sigma,\tau,\varpi)$. An automatic scale selection mechanism can therefore be stated as the selection of the spatiotemporal integration scale triplet (σ,τ,ϖ) at each point in space-time, for which $\|\boldsymbol{\Psi}^{-1}\|_F^2$ is minimal. Here $\Psi^{\mu\beta}_{\nu_1\ldots\nu_i\rho_1\ldots\rho_l}(\mathbf{x};\sigma,\tau,\varpi) = (w(\mathbf{x})\eta^\mu\eta^\beta) * \mathcal{M}_{\nu_1\ldots\nu_i\rho_1\ldots\rho_l}$ represent the matrix part of the system of linear equations described in Result 6.1.

In Result 6.1, we assume that the Gaussian derivatives are scale normalized and furthermore we use natural coordinates, $\frac{x}{\sigma}$. We prefer the standard method of scale normalization which is based on dimensional analysis (Florack et al., 1994) and can be stated as

$$\left[\frac{\partial^n L}{\partial x^{\mu^n}}\right]_{\text{norm}} = \sigma^n \frac{\partial^n L}{\partial x^{\mu^n}}\,. \tag{6.12}$$

In the context of scale selection for feature detection Lindeberg (1998a) has proposed a method in which the structure of the raw image controls the normalization factor. Pedersen and Nielsen (2000) has augmented this method by introducing the fractal dimension of the underlying image structure as a control parameter of the normalization factor. In this paper, we have chosen to use the standard normalization method instead of the more advanced normalization procedures, but for future work, it could be interesting to examine the use of image structure as a control parameter for normalization of derivatives in the context of multi-scale optic flow estimation.

The algorithm proposed in this paper is clearly a sequential process of the computation of normal flow followed by the computation of the optic flow. The computations of the different scale-space derivatives in the normal flow and optic flow steps can readily be parallelized. The structure of the process lets us implement the proposed algorithm in a highly modular fashion.

6.5 Discussion of Performance

We compute the zeroth and first order optic flow for different benchmark image sequences and compare the results with the findings of other authors. We use the angular error measure used by Barron et al. (1994) as well as three of their synthetic benchmark image sequences with known ground truth: The translating trees (TTS), diverging trees (DTS), and Yosemite sequences. We also do a comparison on a synthetic fire particle sequence[4]

[4] Available through anonymous FTP at ftp://ftp.diku.dk/diku/users/kimstp/fire.tar.gz.

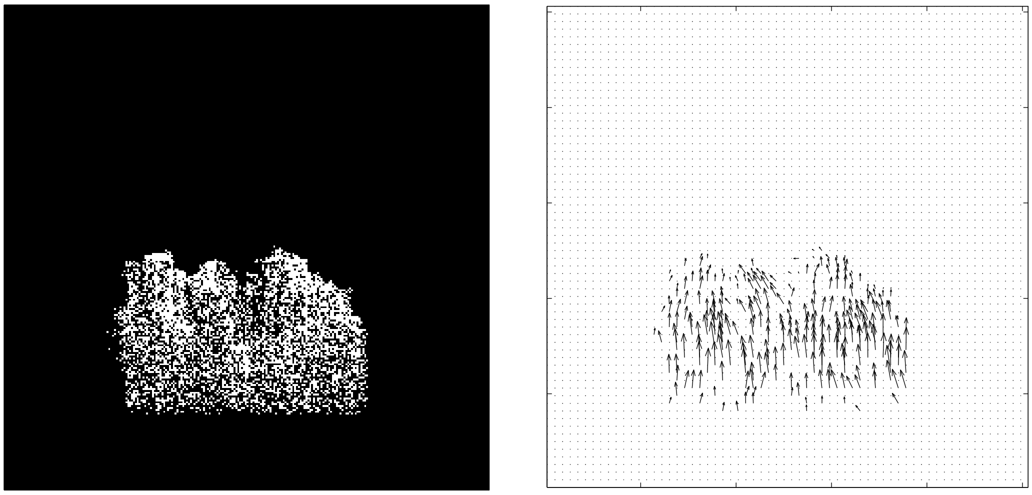


Figure 6.1: The middle image from the synthetic fire particle sequence (left) and the corresponding correct flow field (right). The sequence consists of 32 images of 256×256 gray value pixels with velocities between 0 and 3.79 pixel per frame. We show every fourth vector for the middle image scaled by a factor of 3.

(see Fig. 6.1). In order to show that our method can handle complex non-rigid motion, we compute the optic flow of a real sequence of smoke circulation in a pigsty (see Fig. 6.2). The angular error $\epsilon = \arccos(\hat{\mathbf{v}}_c \cdot \hat{\mathbf{v}}_e)$ is defined as the angle between the estimated vector $\mathbf{v}_e$ and the correct vector $\mathbf{v}_c$, where $\mathbf{v} = (1, v^x, v^y)$ and $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$.

In the following we use the abbreviations

$$\begin{aligned} \Psi^{ij}_{kl}(\mathbf{x};\sigma,\tau,\varpi) &= (w(\mathbf{x})\eta^i(\mathbf{x};\sigma,\tau)\eta^j(\mathbf{x};\sigma,\tau)) * \mathcal{M}_{kl}(\mathbf{x};\varpi) \\ \Phi^{i}_{kl}(\mathbf{x};\sigma,\tau,\varpi) &= (w(\mathbf{x})u^i(\mathbf{x};\sigma,\tau)) * \mathcal{M}_{kl}(\mathbf{x};\varpi)\,. \end{aligned} \tag{6.13}$$

In order to keep the notation simple we assume that each partial derivative of the Gaussians in $\mathcal{M}_{kl}(\mathbf{x};\varpi)$ is scale normalized by Eq. (6.12). When using scale normalization of the derivatives it is important to remember that this effectively corresponds to changing the measurement units, hence the computed flow vectors are expressed in units of the scale. As an uncertainty measure of the normal flow, we use $w(\mathbf{x}) = 1/\|\mathbf{A}^{-1}\|_F^2$, where the matrix $\mathbf{A}$ is given by the linear equations $\mathbf{A}\mathbf{u} = \mathbf{b}$ defining the normal flow (Eq. (6.3) and Eq. (6.4)).

We use the zeroth and first order spatial model for the optic flow even though our method in general lets us use spatiotemporal models. This means, that we purely base our approximation of optic flow on a spatial analysis of the underlying spatiotemporal normal flow field. According to Result 6.1 the optic flow approximated by the zeroth order spatial model $v_0^i = \tilde{v}^i$ is given by the solution to

$$\begin{aligned} \Psi^{xx}\tilde{v}^x + \Psi^{xy}\tilde{v}^y &= \Phi^x \\ \Psi^{xy}\tilde{v}^x + \Psi^{yy}\tilde{v}^y &= \Phi^y . \end{aligned} \tag{6.14}$$

The optic flow modeled by the first order spatial model $v_1^i = \tilde{v}^i + \tilde{v}_x^i x + \tilde{v}_y^i y$ is given by the solution to the system of linear equations given by the partial derivatives of the functional $E(\mathbf{v}_1)$ with respect to the six model parameters $\tilde{v}_j^i = (\tilde{v}^x, \tilde{v}^y, \tilde{v}_x^x, \tilde{v}_x^y, \tilde{v}_y^x, \tilde{v}_y^y)^T$,

$$\begin{aligned} \frac{\partial E}{\partial \tilde{v}_j^x} &= \Psi_j^{xx}\tilde{v}^x + \Psi_j^{xy}\tilde{v}^y + \Psi_{xj}^{xx}\tilde{v}_x^x + \Psi_{xj}^{xy}\tilde{v}_x^y + \Psi_{yj}^{xx}\tilde{v}_y^x + \Psi_{yj}^{xy}\tilde{v}_y^y - \Phi_j^x &= 0 \\ \frac{\partial E}{\partial \tilde{v}_j^y} &= \Psi_j^{xy}\tilde{v}^x + \Psi_j^{yy}\tilde{v}^y + \Psi_{xj}^{xy}\tilde{v}_x^x + \Psi_{xj}^{yy}\tilde{v}_x^y + \Psi_{yj}^{xy}\tilde{v}_y^x + \Psi_{yj}^{yy}\tilde{v}_y^y - \Phi_j^y &= 0 . \end{aligned} \tag{6.15}$$

In order to solve these linear equations we have used the pseudo inverse of the matrix Ψ_{kl}^{ij} because it can be close to singular.

We have computed the optic flow for the four synthetic benchmark sequences using fixed scales and automatic scale selection (Table 6.1). We see, that for some types of sequences the automatic scale selection improves the results. This is not true for the fire sequence and zeroth order results for the DTS sequence. This indicates that scale selection based on numerical stability might not be the best solution in all situations. We believe that a way to improve this would be to incorporate information of the structure of the normal flow into the scale selection mechanism. Furthermore, for certain fixed fine scales we find that the first order model does not produce better results than the zeroth order model. The reason for this is that at fine scales the accuracy of the higher order partial derivatives needed in the first order model reduces. Note as well that the zeroth order optic flow model does not handle the sequences consisting of non-translational motion well; this concerns the DTS, Yosemite, and fire sequences. This is not surprising considering that these sequences consist of a type of motion which is poorly modeled by this type of model.

In Table 6.2 we show some of the results from Table 6.1 together with the best results of other optic flow techniques. Unfortunately we could only get results for the Yosemite sequence for the Alvarez et al. (2000) method. We see that both the zeroth and first order scale selected optic flow models perform as well as, and in some cases better than, other methods for most benchmark sequences. For the diverging trees and Yosemite sequences the results of the zeroth order optic flow are mediocre. The reason for this is, as mentioned above, that the zeroth order model is a poor model of this type of flow. We expect that the first order result for the Yosemite sequence is mediocre, because of the apparent limitations of the scale selection method

Parameters	TTS		DTS	
	Mean	St. dev.	Mean	St. dev.
M=0, $(1, 2, 2)$	0.99	4.50	2.43	3.42
M=1, $(1, 2, 2)$	0.87	1.76	1.17	1.74
M=0, $(2, 2, 4)$	0.42	0.87	3.41	2.64
M=1, $(2, 2, 4)$	0.52	1.23	1.49	1.88
M=0, $(4, 2, 8)$	0.31	0.19	5.19	3.10
M=1, $(4, 2, 8)$	0.25	0.25	1.49	0.88
M=0, multi-scale	0.34	0.23	5.02	3.86
M=1, multi-scale	0.15	0.11	0.80	0.47

Parameters	Yosemite		Fire	
	Mean	St. dev.	Mean	St. dev.
M=0, $(1, 2, 2)$	20.15	17.37	5.92	14.61
M=1, $(1, 2, 2)$	21.83	19.00	5.78	14.79
M=0, $(2, 2, 4)$	17.16	13.89	7.88	14.60
M=1, $(2, 2, 4)$	16.94	13.45	7.42	14.78
M=0, $(4, 2, 8)$	18.42	11.95	12.88	16.62
M=1, $(4, 2, 8)$	17.30	10.69	11.71	17.75
M=0, multi-scale	11.50	15.66	7.85	15.57
M=1, multi-scale	8.51	12.57	7.55	15.65

Table 6.1: Mean angular errors and corresponding standard deviations of approximated zeroth ($M = 0$) and first ($M = 1$) order optic flow based on zeroth order normal flow. The results have been computed at different sets of spatiotemporal and integration scales (σ, τ, ϖ). The last two rows show results produced with the automatic scale selection method discussed in Sec. 6.4. The scales were selected from the sets $\sigma \in \{1.0, 1.414, 2.0, 2.828, 4.0, 5.656, 8.0\}$, $\tau \in \{1.0, 2.0, 3.0\}$, $\varpi \in \{1.0, 2.0, 4.0, 8.0\}$.

Techniques	TTS		DTS	
	Mean	St. dev.	Mean	St. dev.
Horn & Schunck	2.02	2.27	2.55	3.67
Uras et al.	0.62	0.52	4.64	3.48
Nagel	2.44	3.06	2.94	3.23
Anandan	4.54	3.10	7.64	4.96
Singh (step 2)	1.25	3.29	8.60	5.60
Florack et al.	0.49	1.92	1.15	3.32
M=1, $(1, 2, 2)$	0.87	1.76	1.17	1.74
M=0, multi-scale	0.34	0.23	5.02	3.86
M=1, multi-scale	0.15	0.11	0.80	0.47

Techniques	Yosemite		Fire	
	Mean	St. dev.	Mean	St. dev.
Horn & Schunck	9.78	16.19	9.08	18.97
Uras et al.	8.94	15.61	14.68	25.64
Nagel	10.22	16.51	10.84	21.75
Anandan	13.36	15.64	14.38	22.77
Singh (step 2)	10.44	13.94	9.94	19.71
Alvarez et al.	5.53	7.40	–	–
M=1, $(1, 2, 2)$	21.83	19.00	5.78	14.79
M=0, multi-scale	11.50	15.66	7.85	15.57
M=1, multi-scale	8.51	12.57	7.55	15.65

Table 6.2: Mean angular errors and standard deviations for different optic flow techniques obtained from Barron et al. (1994) (the best results for different methods), Alvarez et al. (2000), and Florack et al. (1998) (first order multi-scale results). The last three rows are results obtained with our method. The fire results for other methods were computed by using the implementations by Barron et al. (1994).

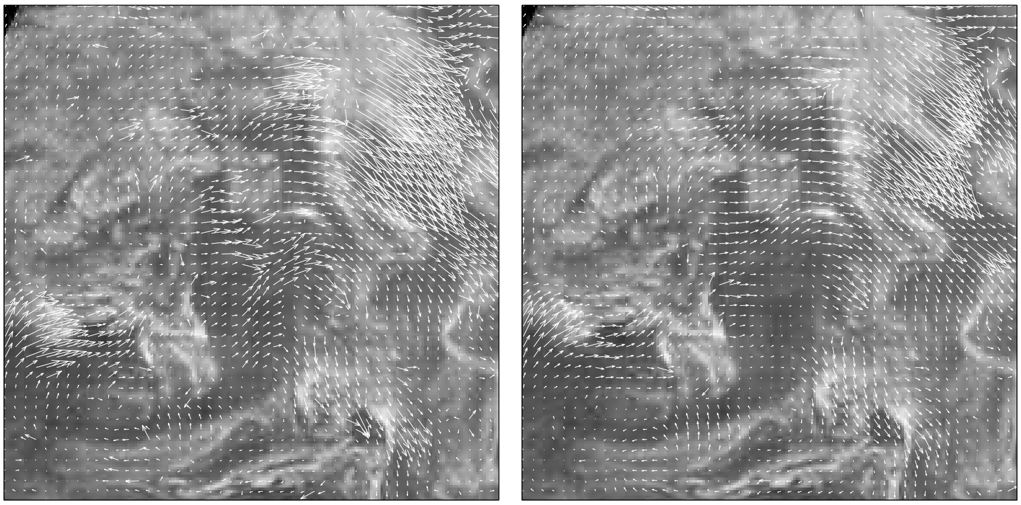


Figure 6.2: Smoke circulation in a pigsty. The sequence consists of 16 images of 256×256 gray value pixels. We only show a part of the middle image of the sequence with the corresponding zeroth (left) and first (right) order scale selected optic flow. We plot every fourth vector scaled by a factor of 5. In the top of the image there is a circular motion from left to right and in the bottom there is a slow motion from right to left.

and our limiting choice of spatial models of the optic flow. Our method delivers good results for the fire sequence, which shows that it works well on sequences consisting of complex non-rigid motion.

In Fig. 6.2 we show the zeroth and first order optic flow for a real complex sequence of smoke circulation in a pigsty[5]. Notice that our method captures the circular motion in the sequence and that the produced flow fields follow the structure of the smoke. It is clear that the zeroth order model breaks down at several points contrary to the first order model which seems to do a good job at almost all image points. We would expect that higher order models would improve the results to some extend, but it is also well-known that too complex models would lead to over-fitting to the data and it is therefore important to choose the model order carefully.

6.6 Conclusions

In this paper we have presented an algorithm for estimation of optic flow based on a polynomial model of the local flow and regularized by a Gaussian integration scale-space. The method fits the normal component of the

[5]See http://www.diku.dk/users/kimstp/demos/ for more details.

model to the underlying normal flow, which we presume is given, and the tangential component is extracted by integration of the local structure of the normal flow. In order to take full advantage of the multi-scale property of the method, we have suggested the use of an automatic local scale selection mechanism proposed by Niessen and Maas (1997). We have compared the performance of the proposed method based on zeroth and first order models with that of other optic flow methods, (Alvarez et al., 2000; Barron et al., 1994; Florack et al., 1998). We thereby show that the method with these models performs as well as, and in some cases outperforms, other optic flow methods.

The optic flow method proposed in this paper is least committed in the sense that we model the local variation of the optic flow by a polynomial model for which the range of validity is determined by the choice of local integration scale. This makes the proposed algorithm a useful tool in e.g. experimental fluid dynamics (Pedersen, 2000), which we illustrated with an analysis of a sequence of smoke circulation in a pigsty. We measured the performance of our and other methods on a complex synthetic fire particle sequence consisting of non-rigid motion, thereby showing that our method is able to handle this type of motion better than other methods.

When using this, as well as other, methods it is important to choose the model order carefully, because the actual number of d.o.f. will vary across the image. In regions with non or little local flow structure we would expect that a zeroth order flow would give us an accurate model, because of the low number of d.o.f. Regions with a large amount of local flow structure lead to a large number of d.o.f., hence higher order models are necessary. We therefore believe that a local model order selection mechanism, like the minimum description length principle, would be a valuable tool combined with the optic flow method proposed in this paper.

Acknowledgments

We thank Theo Engell-Nielsen for producing the synthetic fire particle image sequence.

Part III

Discussion

Chapter 7

Discussion of Results and Open Problems

This chapter is a summary of the results presented in the thesis together with a discussion of open problems and future research directions.

7.1 Natural Image Statistics

In Chapter 2 I developed a scale normalization procedure based on the assumption that natural images can be modeled by the fractal Brownian image model (called the fractal Brownian motion in Chapter 2). This scale normalization procedure is an extension of the automatic scale selection method introduced by Lindeberg (1998b). The choice of the fractal Brownian image as a model of natural images is justified by the fact that the power spectrum of this model, i.e. the scaling of the second order statistics, is similar to what we find for different classes of natural images. Natural images have a power spectrum proportional to $1/|\xi|^{\alpha}$ where the value of α varies around 2. This variation can be modeled by the fractal Brownian image model. In Chapter 2 I have focused on the fractal nature of the fractal Brownian image, and only discussed its statistical properties briefly.

The Brownian and Gaussian image models were discussed in Chapter 5 and various statistical properties and results were presented. The Brownian image model is a scale invariant Gaussian stochastic field which seems to be a good model of the scaling of the second order statistics of ensembles of natural images. I showed empirically that the covariance structure of natural images mapped into the so-called jet space can be approximated by the Brownian image model. The fractal Brownian image model is a generalization of the Brownian image model, which in general is not scale invariant but self similar.

To sum up, the fractal Brownian image model captures the scaling of the second order statistics of single natural images, and the Brownian image model captures the scaling of the second order statistics of ensembles of natural images.

Chapters 3 and 4 are two studies of the statistical variation of local image geometry with a focus on edge-like structures. In Chapter 3 I studied high-contrast 3×3 pixel patches of both optical and 3D range images of natural scenes. Both types of image modalities captures similar geometric information about the scene. I investigated how the data of high-contrast patches distribute in state space with respect to a model of edges. A blurred step edge was used as the model, which trace out a continuous two dimensional manifold in state space. The results show that optical patches concentrate around the edge manifold in state space. In fact, the density with respect to distance θ to the manifold is a power law of the form $\theta^{-2.5}$ with infinite density at the manifold. 3D range patches have a completely different distribution in state space. They concentrate around what can be characterized as binary patches, again with a density following a power law. It is not surprising that we get different results from different image modalities, but the interesting thing about these results is that the differences between the sub-resolution properties of optical and range images apparently manifest themselves in the form of how data clusters in state space. I believe that clusters of data correspond to interesting image primitives or features and it is noteworthy that these primitives and the averaging effect of optical cameras lead to continuous manifolds in state space. Similarly, the effect of the 3D range finder can be seen as a morphological filter (such as a median or mean filter) which maps primitives into compact clusters in state space.

In Chapter 4 I extended the results on optical images to filter responses. I used linear scale-space theory and computed image derivatives in scale-space and mapped the images into the so-called 3-jet space. An analytical model of the scale-space of step edges was introduced, which generates a two dimensional C^∞ manifold in the 3-jet state space. The results confirm the previous findings on 3×3 high-contrast optical patches. Optical images mapped into 3-jet space seem to cluster around the edge manifold in this space. The density of data with respect to distance to the manifold follows a power law proportional to $\theta^{-0.7}$ with infinite probability density at the manifold. Besides showing that the 3×3 patch results generalize to filter responses, the results also showed that the distribution is approximately scale invariant. I believe that in general image primitives generates manifolds in state space with a dimension that reflects the complexity of the primitives, e.g. dimension 2 for edges and 3 for bars. Furthermore, these manifolds are related to each other in a hierarchical form, e.g. the edge manifold is part of the manifold of bars.

7.1.1 Open Problems

The results of Chapters 3 and 4 opens several new questions. My work can be seen as pointing towards a way of discovering what are sensible image primitives or features. These primitives could be defined as the local geometric structures in images which we are most likely to see. These structures will generate clusters or manifolds in state space, hence we can learn features by learning models of the corresponding manifolds. That is, we can learn the primitive features of Marr's primal sketch. Unfortunately this is a difficult learning problem.

Another approach, opposite to learning, is to specify models of features of interest as was done in Chapters 3 and 4. A future study could be to work out the distribution of both natural and Brownian images in jet space on and around manifolds of interesting differential invariants, e.g. the manifold of so-called toppoints (Johansen et al., 1986). The Brownian image model provides us with a void model, which can be used to get an understanding of the complexity of the distribution of natural images. Knowing the statistics on or around the manifolds of features might lead us to an information theoretic measure of how much information is captured by these features. We can use this as a quality measure of proposed features. This is probably the most feasible approach to building a probabilistic primal sketch or at least a primal sketch based on empirical findings.

In the analyses in Chapters 3 and 4 the spatial information is lost, which means that we can not examine the dependency between the geometry at different spatial locations. The local structure of images is often part of large structures such as objects, hence there must be a high correlation of the geometry between different points belonging to the same large scale structure. In the analysis of Chapters 3 and 4, I have not touched upon the problem of describing the long range dependencies in natural images. The obvious question is therefore how do we extend the prior on natural images to capture long range dependencies?

As an example of possible applications of the results presented in this thesis, I would like to mention a novel image reconstruction method by Nielsen and Lillholm (2001) and Lillholm et al. (2002). They have proposed a variational approach to image reconstruction based on simple probabilistic models of the statistics of natural images. It would be interesting to incorporate the results of this thesis into a prior distribution of natural images which could be used in this image reconstruction method. The results on edge statistics should be combined with the second order statistics of the

Brownian image model. In order to construct the prior from this information we might use the ideas of Jaynes (1968) as proposed in the introduction (Chapter 1). That is, the empirical results should be stated as constraints which the prior distribution must fulfill while having maximum entropy.

Although I believe the work on natural image statistics presented in this thesis answers several important questions, it is clear from this discussion that much more work is needed. First of all we are far from having a full understanding of the statistical variation of images of natural scenes. In the work presented in this thesis I have focused my investigations on the statistical variation of local image geometry inspired by Marr's (Marr, 1982) idea of a primal sketch consisting of a set of primitive building blocks called features. The next important step is to take advantage of this knowledge in the solution of image analysis and computer vision problems. In Chapters 1 and 2 I discussed some examples of possible applications.

7.2 Motion Analysis

In Chapter 6 a multi-scale algorithm for the computation of optic flow was presented. This method is based on fitting a local polynomial model of the flow to the analyzed image sequence. The normal component of the model flow is fitted to the normal flow of the sequence, and the tangential component is extracted by integrating the flow structure in a neighborhood. The involved local scale parameters are selected automatically based on a measure of numerical stability. The performance of the algorithm was evaluated and it was found that its performance compares with that of other existing methods. The method was originally developed for use in experimental fluid dynamics and the performance evaluation shows promising results on image sequences of non-rigid motion as found in smoke circulation.

7.2.1 Open Problems

There are at least two open problems with the optic flow method of Chapter 6. The first problem is that the proposed automatic scale selection algorithm leaves room for improvements. It would for instance be interesting to see if Lindeberg's (1998b) automatic scale selection scheme for image features can be extended to image sequences. That is, one could make a similar analysis of the scaling of optic flow as was done in Chapter 2 for differential features in images. In my opinion a proper analysis would require empirical knowledge of the scaling of image sequences of natural scenes. It is interesting to note that Nielsen et al. (2002) have introduced Brownian warps as a least committed model of deformations of images, which leads to the question: Does the Brownian warp provide a good model of motion in natural image sequences and can we derive a scale normalization procedure that can be used in conjunction with automatic scale selection for optic flow?

The answer to this question could bridge the two research tracks of this thesis, namely the research on natural image statistics and on motion analysis. It seems necessary to open an investigation into the statistics of image sequences of natural scenes in order to improve scale selection in the proposed optic flow method.

The second problem with the proposed optic flow method is that of model order selection. The structure of motion varies across the image sequence in both space and time, leading to the conclusion that different orders of local polynomial models are needed across the sequence. The question is how to select proper model orders locally?

There are several approaches to model order selection for optic flow. One might state the problem of model selection via the minimum description length (MDL) principle (Rissanen, 1989), but there are other possibilities. Nielsen and Olsen (1998; 2002) have shown how to divide the local optic flow into flow equivalence classes. A flow equivalence class consists of local flow fields that are equivalent with respect to some transformation or warp of the vector field. It has been proposed (Nielsen and Olsen, 2002) that the Brownian warp introduced by Nielsen et al. (2002) can be used as an equivalence transformation. The Brownian warp may give us a least committed metric for dividing the local optic flow into equivalence classes of different complexity. We might assign a model order to local flow regions based on which equivalence class it belongs to.

I am convinced that a better approach to scale selection and a solution to the model selection problem would greatly improve the optic flow method presented in Chapter 6. If a further study of this method should be done, these are the two most important problems to be handled.

Bibliography

Alvarez, L., Y. Gousseau, and J. Morel: 1999a, 'Scales in Natural Images and a Consequence on their Bounded Variation Norm'. In: M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert (eds.): *Proc. of Scale-Space'99*. Corfu, Greece, pp. 247–258, Springer.

Alvarez, L., Y. Gousseau, and J. Morel: 1999b, 'The size of objects in natural and artificial images'. In: P. W. Hawkes (ed.): *Advances in Imaging and Electron Physics*. Academic Press.

Alvarez, L., Y. Gousseau, and J. Morel: 1999c, 'The size of Objects in Natural Images'. Technical Report CMLA9921, CMLA.

Alvarez, L., J. Weickert, and J. Sanchez: 2000, 'Reliable Estimation of Dense Optical Flow Fields with Large Displacements'. *International Journal of Computer Vision* **39**(1), 41 – 56.

Arfken, G.: 1985, *Mathematical Methods for Physicists*, Computational Imaging and Vision. London: Academic Press.

Arnspang, J.: 1988, 'Optic Acceleration'. In: *Second International Conference on Computer Vision*. pp. 364–373.

Barlow, H. B.: 1961, 'Possible principles underlying the transformations of sensory messages'. In: W. Rosenblith (ed.): *Sensory Communication*. Cambridge, MA, USA: MIT Press, pp. 217–234.

Barron, J. L., D. J. Fleet, and S. S. Beauchemin: 1994, 'Performance of Optical Flow Techniques'. *International Journal of Computer Vision* **12**(1), 43–77.

Bell, A. J. and T. J. Sejnowski: 1997, 'The "independent components" of natural scenes are edge filters'. *Vision Research* **37**, 3327–3338.

Bronshtein, I. N. and K. A. Semendyayev: 1998, *Handbook of Mathematics*. Springer-Verlag, 3rd edition.

Buccigrossi, R. W. and E. P. Simoncelli: 1999, 'Image Compression via Joint Statistical Characterization in the Wavelet Domain'. *IEEE Transaction on Image Processing* **8**(12), 1688–1701.

Caselles, V., B. Coll, and J. Morel: 2002, 'Geometry and Color in Natural Images'. *Journal of Mathematical Imaging and Vision* **16**(2), 89–105.

Conway, J. H. and N. J. A. Sloane: 1988, *Sphere Packings, Lattices and Groups*, No. 290 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag.

Cover, T. M. and J. A. Thomas: 1991, *Elements of Information Theory.* New York: John Wiley & Sons.

Deriugin, N. G.: 1956, 'The power spectrum and the correlation function of the television signal'. *Telecommunications* **1**, 1–12.

Desolneux, A., L. Moisan, and J. Morel: 2001, 'Partial gestalts'. Technical Report CMLA 2001-22, CMLA, ENS Cachan, France.

Einstein, A.: 1956, *Investigations on the theory of the Brownian movement.* Dover. Republication of a collection of papers first published in 1926.

Elder, J. H. and S. W. Zucker: 1996, 'Local Scale Control for Edge Detection and Blur Estimation'. In: B. Baxton and R. Cipolla (eds.): *Proc. of European Conference on Computer Vision.* pp. 57–69, Springer.

Elliott, J. P. and P. G. Dawber: 1979, *Symmetry in Physics*, Vol. 1. New York: Oxford University Press.

Fazzalari, N. L. and I. H. Parkinson: 1996, 'Fractal dimension and architecture of trabecular bone'. *Journal of Pathology* **82**(1), 100–105.

Field, D. J.: 1987, 'Relations between the statistics of natural images and the response properties of cortical cells'. *Journal of Optical Society of America* **4**(12), 2379–2394.

Florack, L.: 1997, *Image Structure*, Computational Imaging and Vision. Dordrecht: Kluwer Academic Publishers.

Florack, L., W. Niessen, and M. Nielsen: 1998, 'The Intrinsic Structure of Optic Flow Incorporating Measurement Duality'. *International Journal of Computer Vision* **27**(3), 263–286.

Florack, L., B. ter Haar Romeny, M. Viergever, and J. Koenderink: 1996, 'The Gaussian Scale-Space Paradigm and the Multiscale Local Jet'. *International Journal of Computer Vision* **18**, 61–75.

Florack, L. M., B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever: 1992, 'Scale and the differential structure of images'. *Image and Vision Computing* **10**(6), 376–388.

Florack, L. M., B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever: 1994, 'Linear Scale-Space'. *Journal of Mathematical Imaging and Vision* **4**(4), 325–351.

Friedlander, G. and M. Joshi: 1998, *Introduction to The Theory of Distributions*. Cambridge University Press, 2nd edition.

Friedman, J. H.: 1987, 'Exploratory Projection Pursuit'. *Journal of the American Statistical Association* **82**(397), 249–266.

Gårding, J.: 1987, 'A note on the application of fractals in image analysis'. Technical Report TRITA-NA-P8716 CVAP 49, Computer Vision and Associative Pattern Processing Laboratory, Royal Institute of Technology.

Gårding, J. and T. Lindeberg: 1996, 'Direct Computation of Shape Cues Using Scale-Adapted Spatial Derivative Operators'. *International Journal of Computer Vision* **17**(2), 163 – 191.

Geman, D. and A. Koloydenko: 1999, 'Invariant Statistics and Coding of Natural Microimages'. In: *Proc. of the IEEE Workshop on Statistical and Computational Theories of Vision, http://pascal.stat.ucla.edu/~sczhu/Workshops/sctv99.html*. Published on the Web.

Gousseau, Y. and J. Morel: 2001, 'Are natural images of bounded variation?'. *SIAM Journal of Mathematical Analysis* **33**(3), 634–648.

Granlund, G. H. and H. Knutsson: 1995, *Signal Processing for Computer Vision*. Kluwer, Dordrecht.

Grenander, U. and A. Srivastava: 2001, 'Probability Models for Clutter in Natural Images'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **23**(4), 424–429.

Griffin, L. D., M. Lillholm, and M. Nielsen: 2002, 'Natural image profiles are most likely to be step edges'. *Vision Research*. Submitted.

Heyden, A., G. Sparr, M. Nielsen, and P. Johansen (eds.): 2002, 'Proceedings of 7th European Conference on Computer Vision', LNCS 2350–2353. Copenhagen, Denmark: Springer Verlag.

Horn, B. K. P. and B. G. Schunck: 1981, 'Determining Optical Flow'. *Artificial Intelligence* **17**, 185–203.

Huang, J., A. B. Lee, and D. Mumford: 2000, 'Statistics of Range Images'. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1. Hilton Head Island, SC, pp. 324–331.

Huang, J. and D. Mumford: 1999, 'Statistics of Natural Images and Models'. In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*.

Huang, Z. and J. Yan: 2000, *Introduction to Infinite Dimensional Stochastic Analysis*. Kluwer Academic Publishers.

Huber, P. J.: 1985, 'Projection Pursuit'. *The Annals of Statistics* **13**(2), 435–475.

Hyvärinen, A.: 1999, 'Survey on Independent Component Analysis'. *Neural Computing Surveys* **2**, 94–128.

Iijima, T.: 1962, 'Basic Theory on Normalization of a Pattern'. *Bulletin of Electrical Laboratory* **26**, 368–388. In Japanese.

Jaynes, E. T.: 1957, 'Information Theory and Statistical Mechanics'. *Physical review* **106**(4), 620–630.

Jaynes, E. T.: 1968, 'Prior Probabilities'. *IEEE Transaction on Systems Science and Cybernetics* **SSC-4**, 227–241.

Jaynes, E. T.: 1996, *Probability Theory: The Logic of Science*. Fragmentary Edition of March 1996.

Johansen, P.: 1997, 'Local Analysis of Image Scale Space'. In (Sporring et al., 1997), Chapt. 10, pp. 139–146.

Johansen, P., S. Skelboe, K. Grue, and J. Andersen: 1986, 'Representing signals by their toppoints in scale space'. In: *Proceedings of the International Conference on Image Analysis and Pattern Recognition*. pp. 215–217, IEEE Computer Society Press.

Kanade, T. and M. Okutomi: 1994, 'A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment'. *IEEE Transaction of Pattern Analysis and Machine Intelligence* **16**(9), 920–932.

Kerckhove, M. (ed.): 2001, 'Scale-Space and Morphology in Computer Vision: Third international Conference, Scale-Space 2001', LNCS 2106. Vancouver, Canada: Springer.

Knill, D. C., D. Field, and D. Kersten: 1990, 'Human discrimination of fractal images'. *Journal of Optical Society of America* **7**(6), 1113–1123.

Koenderink, J. J.: 1984, 'The Structure of Images'. *Biological Cybernetics* **50**, 363–370.

Koenderink, J. J.: 1988, 'Scale-time'. *Biological Cybernetics* **58**, 159–162.

Koenderink, J. J. and A. J. van Doorn: 1987, 'Representation of Local Geometry in the Visual System'. *Biological Cybernetics* **55**, 367–375.

Koenderink, J. J. and A. J. van Doorn: 1996, 'Metamerism in Complete Sets of Image Operators'. In: K. Bowyer and N. Ahuja (eds.): *Advances in Image Understanding.* IEEE Computer Society Press, pp. 113–129.

Kretzmer, E. R.: 1952, 'Statistics of television signals'. *Bell Syst. Tech. Journal* **31**, 751–763.

Lee, A. B.: 2001, 'Statistics, Models and Learning in BCM Theory of a Natural Visual Environment'. Ph.D. thesis, Department of Physics at Brown University, USA.

Lee, A. B., K. S. Pedersen, and D. Mumford: 2001, 'The Complex Statistics of High-Contrast Patches in Natural Images'. In: *WWW Proceedings of Second International IEEE Workshop on Statistical and Computational Theories of Vision, http://pascal.stat.ucla.edu/~szhu/Workshop/SCTV2001.html.* Vancouver, Canada.

Lee, A. B., K. S. Pedersen, and D. Mumford: 2002, 'The Nonlinear Statistics of High-Contrast Patches in Natural Images'. *International Journal of Computer Vision.* Accepted for publication.

Levy Vehel, J.: 1996, 'Introduction to the multifractal analysis of images'. In: Y. Fisher (ed.): *Fractal Image Encoding and Analysis.* Springer Verlag.

Lifshits, M. A.: 1995, *Gaussian Random Functions.* Kluwer Academic Publishers.

Lillholm, M., M. Nielsen, and L. D. Griffin: 2002, 'Feature-based Image Analysis'. *International Journal of Computer Vision.* Accepted.

Lindeberg, T.: 1994a, 'Scale Selection for Differential Operators'. Technical Report ISRN KTH/NA/P–9403-SE, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology.

Lindeberg, T.: 1994b, *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Computer Science. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lindeberg, T.: 1995, 'Direct Estimation of Affine Image Deformations Using Visual Front-End Operations With Automatic Scale Selection'. In: *Proc. of International Conference on Computer Vision*. pp. 134–141, Springer.

Lindeberg, T.: 1996, 'Scale-space: A framework for handling image structures at multiple scales'. Technical Report CVAP-TN15, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology.

Lindeberg, T.: 1997, 'Linear Spatio-Temporal Scale-Space'. In: B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever (eds.): *Scale-Space Theory in Computer Vision: First International Conference, Scale-Space'97, Proceedings*. pp. 113–127, Springer Verlag.

Lindeberg, T.: 1998a, 'Edge detection and ridge detection with automatic scale selection'. *International Journal of Computer Vision* **30**(2), 117–154.

Lindeberg, T.: 1998b, 'Feature detection with automatic scale selection'. *International Journal of Computer Vision* **30**(2), 79–116.

Malik, J., S. Belongie, T. Leung, and J. Shi: 2001, 'Contour and Texture Analysis for Image Segmentation'. *International Journal of Computer Vision* **43**(1), 7–27.

Mallat, S.: 1989, 'A theory for multiresolution signal decomposition: The wavelet representation'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **11**(7), 674–693.

Mallat, S.: 1998, *A wavelet tour of signal processing*. Academic Press, 1st edition.

Mandelbrot, B. B.: 1982, *The fractal geometry of nature*. San Francisco: W. H. Freeman and company.

Mandelbrot, B. B. and J. W. van Ness: 1968, 'Fractional Brownian motions, fractional noises and applications'. *SIAM Review* **10**(4), 422–437.

Marr, D.: 1982, *Vision*. New York: W. H. Freeman.

Mumford, D.: 1996, 'The Statistical Description of Visual Signals'. In: O. K.Kirchgassner and R.Mennicken (eds.): *ICIAM'95*. Akademie Verlag.

Mumford, D.: 2001, 'Gaussian Models for Images'. Unpublished book chapter.

Mumford, D. and B. Gidas: 2001, 'Stochastic Models for Generic Images'. *Quarterly of Applied Mathematics* **59**(11), 85–111.

Nielsen, M.: 1995, 'From Paradigm to Algorithms in Computer Vision'. Ph.d. dissertation, Datalogisk Institut, Københavns Universitet (DIKU), Universitetsparken 1, DK-2100 Copenhagen N, Denmark. DIKU Tech. report no. 95/8.

Nielsen, M., L. Florack, and R. Deriche: 1997, 'Regularization, Scale-Space, and Edge Detection Filters'. *Journal of Mathematical Imaging and Vision* **7**(4), 291–307.

Nielsen, M., P. Johansen, A. D. Jackson, and B. Lautrup: 2002, 'Brownian Warps: A Least Committed Prior for Non-rigid Registration'. In: T. Dohi and R. Kikinis (eds.): *MICCAI*. pp. 557–564, Springer Verlag. LNCS 2489.

Nielsen, M. and M. Lillholm: 2001, 'What Do Features Tell about Images?'. In (Kerckhove, 2001), pp. 39–50, Springer.

Nielsen, M., R. Maas, W. J. Niessen, L. L. Florack, and B. M. ter Haar Romeny: 1999, 'Binocular Stereo from Grey-Scale Images'. *Journal of Mathematical Imaging and Vision* **10**(2), 103–122.

Nielsen, M. and O. F. Olsen: 1998, 'The Structure of the Optic Flow Field'. In: *Proceedings of European Conference on Computer Vision*. pp. 271–287, Springer Verlag.

Nielsen, M. and O. F. Olsen: 2002, 'Personal communication'.

Niessen, W. and R. Maas: 1997, 'Optic Flow and Stereo'. In (Sporring et al., 1997), Chapt. 3, pp. 31–42.

O'Hagan, A.: 1994, *Kendall's Advanced Theory of Statistics - Volume 2B: Bayesian Inference*. Edward Arnold.

Olshausen, B. A. and D. J. Field: 1996, 'Natural image statistics and efficient coding'. *Network: Computation in Neural Systems* **7**(2), 333–339.

Ott, E.: 1993, *Chaos in dynamical systems*. Cambridge University Press.

Otte, M. and H. Nagel: 1994, 'Optical flow estimation: Advances and comparisons'. In: *Proc. of European Conference of Computer Vision*. Springer.

Pece, A. E. C.: 2002, 'The Problem of Sparse Image Coding'. *Journal of Mathematical Imaging and Vision* **17**(2), 89–108.

Pece, A. E. C. and A. D. Worrall: 2002, 'Tracking with the EM Contour Algorithm'. In (Heyden et al., 2002), pp. 3–17, Springer Verlag.

Pedersen, K. S.: 2000, 'Turbulence in Optical Flow Fields'. Technical Report DIKU-00-3, Department of Computer Science, University of Copenhagen (DIKU), Universitetsparken 1, 2100 København Ø, Denmark.

Pedersen, K. S. and A. B. Lee: 2002, 'Toward a Full Probability Model of Edges in Natural Images'. In (Heyden et al., 2002), pp. Vol–I 328–342, Springer Verlag.

Pedersen, K. S. and M. Nielsen: 2000, 'The Hausdorff Dimension and Scale-Space Normalisation of Natural Images'. *Journal of Visual Communication and Image Representation* **11**(2), 266 – 277.

Pedersen, K. S. and M. Nielsen: 2001, 'Computing Optic Flow by Scale-Space Integration of Normal Flow'. In (Kerckhove, 2001), pp. 14–25, Springer.

Pentland, A. P.: 1984, 'Fractal-based description of natural scenes'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**(6), 661–674.

Reinagel, P. and A. M. Zador: 1999, 'Natural Scene Statistics at the Centre of Gaze'. *Network: Computation in Neural Systems* **10**(4), 341–350.

Ren, X. and J. Malik: 2002, 'A Probabilistic Multi-scale Model for Contour Completion Based on Image Statistics'. In (Heyden et al., 2002), pp. Vol–I 312–327, Springer Verlag.

Rissanen, J.: 1989, *Stochastic Complexity in Statistical Inquiry*. World Scientific.

Ruderman, D. L. and W. Bialek: 1994, 'Statistics of Natural Images: Scaling in the Woods'. *Physical Review Letters* **73**(6), 814–817.

Sidenbladh, H. and M. J. Black: 2001, 'Learning Image Statistics for Bayesian Tracking'. In: *Proc. of International Conference on Computer Vision*, Vol. II. Vancouver, Canada, pp. 709–716.

Simoncelli, E. P.: 1999a, 'Bayesian Denoising of Visual Images in the Wavelet Domain'. In: P. Müller and B. Vidakovic (eds.): *Bayesian Inference in Wavelet Based Models*. New York: Springer-Verlag, pp. 291–308.

Simoncelli, E. P.: 1999b, 'Modeling the Joint Statistics of Images in the Wavelet Domain'. In: *Proc. of SPIE, 44th Annual Meeting*, Vol. 3813. Denver, CO, pp. 188–195.

Simoncelli, E. P. and B. A. Olshausen: 2001, 'Natural image statistics and neural representation'. *Annual Review of Neuroscience* **24**, 1193–1216.

Sporring, J., M. Nielsen, L. Florack, and P. Johansen (eds.): 1997, *Gaussian Scale-Space Theory*. Kluwer Academic Publishers.

Srivastava, A., A. B. Lee, E. P. Simoncelli, and S. Zhu: 2003, 'On Advances in Statistical Modeling of Natural Images'. *Journal of Mathematical Imaging and Vision* **18**(1). To appear.

Srivastava, A., X. Liu, and U. Grenander: 2002, 'Universal Analytical Forms for Modeling Image Probabilities'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **24**(9), 1200–1214.

Sullivan, J., A. Blake, M. Isard, and J. MacCormick: 1999, 'Object Localization by Bayesian Correlation'. In: *Proc. of International Conference on Computer Vision*. Corfu, Greece, pp. 1068–1075.

Torr, P. H. S.: 2002, 'Bayesian Model Estimation and Selection for Epipolar Geometry and Generic Manifold Fitting'. *International Journal of Computer Vision* **50**(1), 35–61.

Tu, Z. W., S. C. Zhu, and H. Y. Shum: 2001, 'Image Segmentation by Data Driven Markov Chain Monte Carlo'. In: *Proc. of International Conference on Computer Vision*. Vancouver, Canada.

van Hateren, J. H. and A. van der Schaaf: 1998, 'Independent Component Filters of Natural Images Compared with Simple Cells in Primary Visual Cortex'. *Proc. of Royal Society London Series B* **265**, 359 – 366.

Veenland, J. F., J. L. Grashius, F. van Meer, A. L. Beckers, and E. S. Gelsema: 1996, 'Estimation of fractal dimension in radiographs'. *Med. Phys.* **82**(4), 585–594.

Voss, R. F.: 1985, 'Random Fractal Forgeries'. In: R. A. Earnshaw (ed.): *Fundamental Algorithms for Computer Graphics*, Vol. 17. Berlin: Springer-Verlag, pp. 805–835.

Wainwright, M. J. and E. P. Simoncelli: 2000, 'Scale Mixtures of Gaussians and the Statistics of Natural Images'. In: T. K. L. S. A. Solla and K.-R. Müller (eds.): *Advances in Neural Information Processing Systems 12*. MIT Press, pp. 855–861.

Weber, J. and J. Malik: 1995, 'Robust Computation of Optical Flow in a Multi-Scale Differential Framework'. *International Journal of Computer Vision* **14**, 67–81.

Wegmann, B. and C. Zetzsche: 1990, 'Statistical Dependence Between Orientation Filter Outputs Used in an Human Vision Based Image Code'. In: *Proc. of SPIE Visual Communication and Image Processing*, Vol. 1360. Lausanne, Switzerland, pp. 909–922.

Weickert, J.: 1998, *Anisotropic Diffusion in Image Processing*. Teubner.

Weickert, J., S. Ishikawa, and A. Imiya: 1997, 'On the History of Gaussian Scale-Space Axiomatics'. In (Sporring et al., 1997), Chapt. 4, pp. 45–59.

Winkler, G.: 1995, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical introduction*. Springer.

Witkin, A. P.: 1983, 'Scale Space Filtering'. In: *Proc. of the Eighth International Joint Conference on Artificial Intelligence*, Vol. 2. Karlsruhe, Germany, pp. 1019–1023.

Wu, Y. N., S. C. Zhu, and C. Guo: 2002, 'Statistical Modeling of Texture Sketch'. In (Heyden et al., 2002), pp. Vol–III 240–254, Springer Verlag.

Yuille, A. L. and J. M. Coughlan: 2000, 'Fundamental Limits of Bayesian Inference: Order Parameters and Phase Transitions for Road Tracking'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **22**(2), 160–173.

Zetzsche, C., B. Wegmann, and E. Barth: 1993, 'Nonlinear aspects of primary vision: Entropy reduction beyond decorrelation'. In: *International Symposium, Society for Information Display*, Vol. 24. pp. 933–936.

Zhu, S. C. and D. Mumford: 1997, 'Prior Learning and Gibbs Reaction-Diffusion'. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **19**(11), 1236–1250.

Zhu, S. C. and D. Mumford: 1998, 'GRADE: Gibbs Reaction And Diffusion Equations. — a framework for pattern synthesis, image denoising, and clutter removal'. In: *Proc. of International Conference on Computer Vision*.

Zhu, S. C., Y. N. Wu, and D. Mumford: 1997, 'Minimax Entropy Principle and Its Application to Texture Modelling'. *Neural Computation* **9**(8), 1627–1660.

Sammenfatning

Denne Ph.D. afhandling består af to dele, som hver især beskriver de to linjer af forskning jeg har udført. Afhandlingens titel *Statistics of Natural Image Geometry* afspejler min forskning angående statistikken af geometri i naturlige billeder, som udgør hovedvægten af denne afhandling (Part I). Den anden del af afhandlingen (Part II) omhandler analyse af videosekvenser med henblik på detektion og beregning af bevægelse.

Statistik af naturlige billeder

Der er blevet argumenteret for, af denne og andre forfattere (Mumford, 1996; Srivastava et al., 2003), at kendskab til statistikken af naturlige billeder kan være nyttig ved udvikling af datamatsyn og billedanalysealgoritmer. Betragter man datamatsyn og billedanalyseproblemer som statistisk inferensproblemer kan man med fordel benytte den såkaldte Bayesianske inferens. Her er det nødvendigt at tildele en sandsynlighedsfordeling til a priori viden eller modelparametre. Der er i denne afhandling blevet argumenteret for at sådanne sandsynlighedsfordelinger kan konstrueres ud fra empirisk viden om statistikken af naturlige billeder. Fokus har specielt været på den statistiske variation af lokal geometri i billeder ud fra betragtningen, at lokal geometri udgør lav-niveau primitiver/features i mange datamatsynsopgaver.

I afhandlingen er den statistiske variation i høj-kontrast 3×3 mikrobilleder af naturlige billeder blevet studeret. Der blev anvendt billeder med to forskellige modaliteter — billeder optaget med et konventionelt optisk kamera, samt billeder optaget med en 3D-laserskanner. Bemærk at begge billedtyper indeholder information om den betragtede scenes geometri. Det blev analyseret hvorledes disse mikrobilleder fordeler sig i tilstandsrummet med hensyn til en model af kanter. Den benyttede model er en udglattet trinkant, som danner en to-dimensional manifold i tilstandsrummet. Empiriske studier viser at naturlige optiske mikrobilleder fordeler sig tæt omkring kantmanifolden i tilstandsrummet. Fordelingen af optiske mikrobilleder med hensyn til afstand θ til kantmanifolden er proportionel med $\theta^{-2.5}$. 3D-laser mikrobilleder har en markant anderledes fordeling. Denne type mikrobilleder ligger tæt fordelt omkring punkter i tilstandsrummet, der kan betegnes som binære mikrobilleder. Nogle af disse binære mikrobilleder ligger på kantmanifolden, mens andre ligger spredt i tilstandsrummet. Fordelingen af 3D-laser mikrobilleder omkring binære mikrobilleder følger, ligesom optiske mikrobilleder, en potenslov. Forskellen i de to billedtypers fordeling af lokal geometri skyldes forskellene i de to modaliteters subpixel-egenskaber: Det optiske kamera kan sammenlignes med et udglattende filter og 3D-laserskanneren med et morfologisk filter som gør den lokale struktur skarpere.

Studiet af høj-kontrast 3×3 pixel mikrobilleder er blevet udvidet med et studium af den statistiske variation af naturlige optiske billeder afbildet ind i det såkaldte jet-rum. Akserne i jet-rummet består af de partielle billedafledte, og i dette studium benyttedes afledte op til tredje orden, hvilket definere 3-jettet. Billedafledte defineredes i dette studium ved hjælp af lineær skalarumsteori. I denne teori introduceres en skala hvorpå billedafledte måles. Et punkt i jet-rummet beskriver den lokale geometri i et punkt i billedet. Igen fokuseres der på analyse af kanter og en analytisk kantmodel introduceres. Modellen består af skalarummet af en trinkant og den udspænder en to-dimensional C^∞ manifold i jet-rummet. Dette studium viser at den lokale geometri af naturlige billeder er tæt fordelt omkring kantmanifolden. Fordelingen af lokal geometri med hensyn til afstand θ til kantmanifolden er proportionel med $\theta^{-0.7}$, hvilket er i overenstemmelse med forsøgene med høj-kontrast 3×3 pixel optiske mikrobilleder. Ydermere er denne fordeling skalainvariant.

Disse to studier antyder at de såkaldte features, dvs. kanter, blobs, hjørner etc., danner klynger eller manifolder i tilstandsrummet uafhængigt af billedrepræsentation, dvs. pixel-intensitetsværdier eller filtersvar. Dimensionen af en feature-manifold afhænger ikke af billedrepræsentationen, men bestemmes udelukkende af kompleksiteten af featuren eller antallet af frihedsgrader den pågældende feature besidder, f.eks. er dimensionen af kant manifolden 2 og dimensionen af manifolden for en isotrop blob 1. Det er rimeligt at antage at feature-manifolder er indlejret i hinanden og er del af et hierarki af manifolder, f.eks. er det oplagt at kantmanifolden er en del af manifolden for hjørner. Man kan forestille sig muligheden for at lære features ved at lære modeller for de tilhørende manifolder i tilstandsrummet.

Udover disse studier af lokal geometri, er der blevet arbejdet med en matematisk model for billeder. Denne model kaldes *den Brownske billedmodel* og er en stokastisk funktion på planet. Den Brownske model har en række attraktive egenskaber, blandt andet er den skalainvariant og har en Gaussisk sandsynlighedsfordeling. Vigtigst er dog at denne model fanger anden ordens statistikken af naturlige billeder. Empiriske studier viser nemlig at naturlige billeder har en skalainvariant kovariansstruktur. Det er i denne afhandling blevet vist empirisk at den Brownske billedmodel modellerer kovariansstrukturen af naturlige billeder afbildet ind i jet-rummet. Ydermere er den Brownske billedmodel og dens generalisering — *den fraktale Brownske billedmodel* — blevet anvendt som grundlag for en udledning af en metode til skalanormalisering af billedafledte i skalarum. Denne skalanormalisering er en udvidelse af en metode foreslået af Lindeberg (1998b) i forbindelse med en metode til automatisk skalaselektion i feature-detektorer.

Bevægelsesanalyse

I afhandlingens anden del udvikles en algoritme til analyse af bevægelse i videosekvenser. I en bevægelsesanalyse ønsker man at beregne et vektorfelt som beskriver størrelse og retning af ændringen af intensitetsværdier i den analyserede videosekvens. Dette vektorfelt kaldes for *det optiske flowfelt*. Bevægelsesanalyse er vigtigt for løsningen af en række problematikker inden for datamatsyn. Fokus har i denne afhandling været på udviklingen af en metode til at beregne hastighedsvektorfelter med henblik på analyse af væskebevægelse. Her er det vigtigt ikke at lave forsimplende antagelser såsom at den analyserede scene kun indeholder stive legemer og at bevægelsen kan beskrives ved en affin transformation. Disse antagelser bryder generisk sammen ved analyse af væskebevægelse.

I den udviklede metode antages det, at det såkaldte normal-flow er givet på forhånd. Det optiske flow modelleres lokalt med en polynomisk model. Denne models parametre estimeres ved at tilpasse modellens projektion ind på den lokale iso-intensitetskurvenormal til det underliggende normal-flow. Modellens tangentielle komponent udledes ved en integration over et naboområde. I metoden benyttes lineær skalarumsteori til at beregne afledte af billeder og normal-flow, samt at definere integrationsområdet. Modelparametrene estimeres lokalt ved at finde løsningen til et lineært ligningssystem. I metoden indgår også en teknik til automatisk selektion af de involverede skalaparametre.

I afhandlingen sammenlignedes den her beskrevne algoritme til beregning af optisk flow med andre kendte metoder. Denne sammenligning bestod af en kvalitativ vurdering af de involverede metoders evne til at analysere billedsekvenser med et kendt bevægelsesfelt. Sammenligningen viser at den her foreslåede metode producerer hastighedsfelter som er lige så gode, og i visse tilfælde bedre, end de bedste af de andre metoder. Det blev også sandsynliggjort at den her foreslåede metode kan beregne bevægelsesfelter af virkelige videosekvenser af væskebevægelse, som er en god approksimation af det faktiske bevægelsesfelt.