Scale-Space Theory in Computer Vision
Scale-Space Theory in Computer Vision

Tony Lindeberg

Royal Institute of Technology
Stockholm, Sweden
Foreword

The problem of scale pervades both the natural sciences and the visual arts. The earliest scientific discussions concentrate on visual perception (much like today!) and occur in Euclid's (c. 300 B.C.) Optics and Lucretius' (c. 100–55 B.C.) On the Nature of the Universe. A very clear account in the spirit of modern "scale-space theory" is presented by Boscovitz (in 1758), with wide ranging applications to mathematics, physics and geography. Early applications occur in the cartographic problem of "generalization", the central idea being that a map in order to be useful has to be a "generalized" (coarse grained) representation of the actual terrain (Miller and Voskuil 1964). Broadening the scope asks for progressive summarizing. Very much the same problem occurs in the (realistic) artistic rendering of scenes. Artistic generalization has been analyzed in surprising detail by John Ruskin (in his Modern Painters), who even describes some of the more intricate generic "scale-space singularities" in detail: Where the ancients considered only the merging of blobs under blurring, Ruskin discusses the case where a blob splits off another one when the resolution is decreased, a case that has given rise to confusion even in the modern literature.

It is indeed clear that any physical observation of some extended quantity such as mass density or surface irradiance presupposes a scale-space setting due to the inherent graininess of nature on the small scale and its capricious articulation on the large scale. What is the "right scale" does indeed depend on the problem, i.e., whether one needs to see the forest, the trees or the leaves. (Of course this list could be extended indefinitely towards the microscopic as well as the the mesoscopic domains, as has been done in the popular film Powers of Ten (Morrison and Morrison, 1984)). The physicist almost invariably manages to pick the right scale for the problem at hand intuitively. However, in many modern applications the "right scale" need not be obvious at all, and one really needs a principled mathematical analysis of the scale problem.

In applications such as vision the front end system has to process the radiance function blindly (since no meaning resides in the photons as such) and the problem of finding the right scale becomes especially acute. This is true for biological and artificial vision systems alike. Here a principled theory is mandatory and can a priori be expected to yield important insights and lead to mechanistic models. The modern scale-space theory has indeed led to an increased understanding of the low level
operations and novel handles on ways to design algorithms for problems in machine vision.

In this book the author presents a commendably lucid outline of the theory of scale-space, the structure of low level operations in a scale-space setting and algorithmic schemes to use these structures such as to solve important problems in computer vision. The subjects range from a mathematical underpinning, over issues in implementation (discrete scale-space structures) to more open ended algorithmic methods for computer vision problems. The latter methods seem to me to point a way to a range of potentially very important applications. This approach will certainly turn out to be part of the foundations of the theory and practice of machine vision.

It was about time for somebody to write a monograph on the subject of scale-space structure and scale-space based methods, and the author has no doubt performed an excellent service to many in the field of both artificial and biological vision.

_Utrecht, October 4th, 1993_  
*Jan Koenderink*
Preface

We perceive objects in the world as having structures both at coarse and fine scales. A tree, for instance, may appear as having a roughly round or cylindrical shape when seen from a distance, even though it is built up from a large number of branches. At a closer look, individual leaves become visible, and we can observe that the leaves in turn have texture at an even finer scale.

The fact that objects in the world appear in different ways depending upon the scale of observation has important implications when analysing measured data, such as images, with automatic methods. A straightforward way of exemplifying this is to note that every operation on image data must be carried out on a window, whose size can range from a single point to the whole image. The type of information we can get from such an operation is largely determined by the relation between structures in the image and the size of the window. Hence, without prior knowledge about what we are looking for, there is no reason to favour any particular scale. We should therefore try them all and operate at all window sizes.

These insights are not completely new in computer vision. Multi-scale representations of images in terms of pyramids were developed already around 1970. A main motivation then was to achieve computational efficiency by coarse-to-fine strategies. This approach was also supported by findings in neurophysiology about the primate visual system. However, it was soon discovered that relating structures from different levels in the multi-scale representation was far from trivial. Structures at coarse levels could sometimes not be assigned any direct interpretation, since they were hard to trace to finer scales. Despite considerable efforts to develop techniques for matching between scales, a theoretical foundation was missing.

In 1983, Witkin proposed that scale could be considered as a continuous parameter, thereby generalizing the existing notion of Gaussian pyramids. He noted the relation to the diffusion equation and hence found a well-founded way of relating image structures between different scales. Koenderink soon furthered the approach, which has been developed into what we now know as scale-space theory.

Since that work, we have seen the theory develop in many ways, and also realized that it provides a framework for early visual computations of a more general nature. The aim of this book is to provide a coherent overview of this recently developed theory, and to make material, which
has earlier existed only in terms of research papers, available to a larger audience. The presentation provides an introduction into the general foundations of the theory and shows how it applies to essential problems in computer vision such as computation of image features and cues to surface shape. The subjects range from the mathematical foundation to practical computational techniques. The power of the methodology is illustrated by a rich set of examples.

I hope that this work can serve as a useful introduction, reference, and inspiration for fellow researchers in computer vision and related fields such as image processing, signal processing in general, photogrammetry, and medical image analysis. Whereas the book is mainly written in the form of a research monograph, the level of presentation has been adapted so that it can be used as a basis for advanced courses in these fields.

The presentation is organized in a logical bottom-up way, following the ordering of the processing modules in an imagined vision system. It is, however, not necessary to read the book in such a sequential manner. Several of the chapters are relatively self-contained, and it should be possible to read them independently. A guide to the reader describing the mutual dependencies is given in section 1.7 (page 22). I wish the reader a pleasant tour into this highly stimulating and challenging subject.

Stockholm, September 1993,  

Tony Lindeberg
Abstract

The presentation starts with a philosophical discussion about computer vision in general. The aim is to put the scope of the book into its wider context, and to emphasize why the notion of scale is crucial when dealing with measured signals, such as image data. An overview of different approaches to multi-scale representation is presented, and a number of special properties of scale-space are pointed out.

Then, it is shown how a mathematical theory can be formulated for describing image structures at different scales. By starting from a set of axioms imposed on the first stages of processing, it is possible to derive a set of canonical operators, which turn out to be derivatives of Gaussian kernels at different scales.

The problem of applying this theory computationally is extensively treated. A scale-space theory is formulated for discrete signals, and it demonstrated how this representation can be used as a basis for expressing a large number of visual operations. Examples are smoothed derivatives in general, as well as different types of detectors for image features, such as edges, blobs, and junctions. In fact, the resulting scheme for feature detection induced by the presented theory is very simple, both conceptually and in terms of practical implementations.

Typically, an object contains structures at many different scales, but locally it is not unusual that some of these “stand out” and seem to be more significant than others. A problem that we give special attention to concerns how to find such locally stable scales, or rather how to generate hypotheses about interesting structures for further processing. It is shown how the scale-space theory, based on a representation called the scale-space primal sketch, allows us to extract regions of interest from an image without prior information about what the image can be expected to contain. Such regions, combined with knowledge about the scales at which they occur constitute qualitative information, which can be used for guiding and simplifying other low-level processes.

Experiments on different types of real and synthetic images demonstrate how the suggested approach can be used for different visual tasks, such as image segmentation, edge detection, junction detection, and focus-of-attention. This work is complemented by a mathematical treatment showing how the behaviour of different types of image structures in scale-space can be analysed theoretically.
It is also demonstrated how the suggested scale-space framework can be used for computing direct cues to three-dimensional surface structure, using in principle only the same types of visual front-end operations that underlie the computation of image features.

Although the treatment is concerned with the analysis of visual data, the notion of scale-space representation is of much wider generality and arises in several contexts where measured data are to be analyzed and interpreted automatically.
Acknowledgments

This book is based on the author’s thesis Discrete Scale-Space Theory and the Scale-Space Primal Sketch, presented at KTH (Royal Institute of Technology) in Stockholm in May 1991. The material has been updated and extended with respect to research conducted since then.

It is a pleasure to take this opportunity to express my deep gratitude to the following people for their important contributions in various ways:

- Jan-Olof Eriksson for introducing me into this inspiring field, for excellent supervision during my period as a PhD student, and for providing the stimulating and pleasant research environment to work in known as the CVAP group at KTH.

- Jan Koenderink for his outstanding contributions to the field and for serving as the opponent at my defense.

- Kjell Brummström for the enjoyable collaboration with the work on junction classification.

- Jonas Gårding for the fruitful and inspiring collaboration on shape-from-texture and our many discussions.

- Harald Winroth with whom I shared office during a number of years. Our many talks and discussions have contributed substantially to the presentation of this work.

- Fredrik Bergholm for many discussions about scale-space as well as valuable comments on an endless number of manuscript drafts.

- Demetre Betsis for always being a reliable source of good advice.

- Stefan Carlsson for useful discussions, which served as a large source of inspiration to the discrete scale-space theory.

- Ambjörn Næve and Lars Svensson for always sharing their intuition and insights into many aspects of mathematics.

- Thure Valdsoo for providing the interesting aerosol problem.

- Matti Rendahl for maintaining an excellent systems environment at our laboratory, and for always providing wizardly help when needed.
I would also like to thank my other colleagues at the Computational Vision and Active Perception Laboratory, CVAP, for their friendship and for the help they have given me in so many ways: Magnus Andersson, Ann Bengtson, Kiyo Chinzei, Antonio Francisco, Akihiro Horii, Meng-Xiang Li, Anders Lundquist, Oliver Ludwig, Atsulo Maki, Peter Norlund, Niklas Nordström, Göran Olofsson, Lars Olsson, Kourosh Pahlavan, Eva Setterlund, Tomas Uhlén, and Wei Zhang.

I would like to express my gratitude to the other colleagues at KTH with whom I’ve had valuable interactions. In particular, I would like to mention Michael Benedict, Anders Björner, Germund Dahlquist, Lars Holst, Torbjörn Kolsrud, Heinz-Otto Kreiss, Anders Liljeborg, Bengt Lindberg, Jesper Oppelstrup, and Johan Philip.

This work has also benefitted from discussions with other colleagues in the field, in particular the partners in our national, European, and transatlantic collaborations. Especially, I would like to thank Luc Florack, Peter Johansen, Bart ter Haar Romeny, Stephen Pizer, and Richard Weiss with all of whom I have had several highly interesting discussions.

In preparing this manuscript I would like to thank Jan Michael Rynning for valuable advice on \LaTeX{} and typesetting, Birgitta Krasse for redrawing many of the figures, and Mike Casey at Kluwer Academic Publishers for his enthusiasm and patience with this manuscript.

Moreover, I would like to thank my parents Paul and Inga-Lill for always being there when I needed help in any way.

The research presented in this book has been made possible by a graduate position, “excellenstjänst,” and a postgraduate position, “forskarassistenttjänst,” provided by KTH together with financial support from the Swedish National Board for Industrial and Technical Development, NUTEK. Parts of the work have been performed under the ESPRIT-BRA projects InSight and VAP. This support is gratefully acknowledged.

All implementations have been made within the Candela and CanApp programming environment for image analysis developed at the Computational Vision and Active Perception Laboratory.
Contents

1 Introduction and overview .......................... 1
   1.1 Theory of a visual front-end ................. 4
   1.2 Goal ....................................... 4
   1.3 The nature of the problem ...................... 5
   1.4 Scale-space representation .................... 10
   1.5 Philosophies and ideas behind the approach ... 16
   1.6 Relations to traditional applied mathematics ... 19
   1.7 Organization of this book ..................... 22

Part I: Basic scale-space theory

2 Linear scale-space and related multi-scale representations 31
   2.1 Early multi-scale representations ............. 32
   2.2 Quad-tree ...................................... 32
   2.3 Pyramid representations ........................ 33
   2.4 Scale-space representation and scale-space properties ... 39
   2.5 Uniqueness of scale-space representation .......... 47
   2.6 Summary and retrospective ...................... 53
   2.7 Wavelets ....................................... 55
   2.8 Regularization .................................. 56
   2.9 Relations between different multi-scale representations ... 58

3 Scale-space for 1-D discrete signals .................. 61
   3.1 Scale-space axioms in one dimension .......... 62
   3.2 Properties of scale-space kernels .............. 65
   3.3 Kernel classification ............................ 76
   3.4 Axiomatic construction of discrete scale-space .... 82
   3.5 Axiomatic construction of continuous scale-space ... 88
   3.6 Numerical approximations of continuous scale-space ... 91
   3.7 Summary and discussion ........................ 98
   3.8 Conclusion: Scale-space for 1-D discrete signals .... 100

4 Scale-space for N-D discrete signals ................ 101
   4.1 Scale-space axioms in higher dimensions ....... 103
   4.2 Axiomatic discrete scale-space formulation ...... 106
   4.3 Parameter determination ....................... 112
   4.4 Summary and discussion ....................... 118
   4.5 Possible extensions ............................ 119

5 Discrete derivative approximations with scale-space properties 123
   5.1 Numerical approximation of derivatives .......... 123
### Part II:
The scale-space primal sketch: Theory

#### 7 The scale-space primal sketch

- 7.1 Grey-level blob ............................................. 166
- 7.2 Grey-level blob tree ........................................ 170
- 7.3 Motivation for introducing a multi-scale hierarchy .... 171
- 7.4 Scale-space blob ............................................. 172
- 7.5 Scale-space blob tree ....................................... 173
- 7.6 Grey-level blob extraction: Experimental results .... 174
- 7.7 Measuring blob significance ................................. 177
- 7.8 Resulting representation ................................... 185

#### 8 Behaviour of image structures in scale-space: Deep structure

- 8.1 Trajectories of critical points in scale-space .......... 189
- 8.2 Scale-space blobs ............................................. 192
- 8.3 Bifurcation events for critical points: Classification ... 194
- 8.4 Bifurcation events for grey-level blobs and scale-space blobs 201
- 8.5 Behaviour near singularities: Examples ....................... 204
- 8.6 Relating differential singularities across scales .......... 214
- 8.7 Density of local extrema as function of scale .............. 217
- 8.8 Summary ..................................................... 226

#### 9 Algorithm for computing the scale-space primal sketch

- 9.1 Grey-level blob detection .................................... 227
- 9.2 Linking grey-level blobs into scale-space blobs .......... 232
- 9.3 Basic blob linking algorithm ................................ 240
- 9.4 Computing scale-space blob volumes ......................... 242
- 9.5 Potential improvements of the algorithm .................... 243
- 9.6 Data structure ................................................. 244

### Part III:
The scale-space primal sketch: Applications

#### 10 Detecting salient blob-like image structures and their scales

- 10.1 Motivations for the assumptions ............................ 250
- 10.2 Basic method for extracting blob structures .............. 252
- 10.3 Experimental results ........................................ 252
10.4 Further treatment of generated blob hypotheses  
10.5 Properties of the scale selection method  
10.6 Additional experiments  

11 Guiding early visual processing with qualitative scale and region information  
11.1 Guiding edge detection with blob information  
11.2 Automatic peak detection in histograms  
11.3 Junction classification: Focus-of-attention  
11.4 Example: Analysis of aerosol images  
11.5 Other potential applications  

12 Summary and discussion  
12.1 Scale-space experiences  
12.2 Relations to previous work  
12.3 Grey-level blobs vs. Laplacian sign blobs  
12.4 Invariance properties  
12.5 Alternative approaches and further work  
12.6 Conclusions  

Part IV: 
Scale selection and shape computation in a visual front-end  

13 Scale selection for differential operators  
13.1 Basic idea for scale selection  
13.2 Proposed method for scale selection  
13.3 Blob detection  
13.4 Junction detection  
13.5 Edge detection  
13.6 Discrete implementation of normalized derivatives  
13.7 Interpretation in terms of self-similar Fourier spectrum  
13.8 Summary and discussion  

14 Direct computation of shape cues by scale-space operations  
14.1 Shape-from-texture: Review  
14.2 Definition of an image texture descriptor  
14.3 Deriving shape cues from the second moment matrix  
14.4 Scale problems in texture analysis  
14.5 Computational methodology and experiments  
14.6 Spatial selection and blob detection  
14.7 Estimating surface orientation  
14.8 Experiments  
14.9 Summary and discussion  

15 Non-uniform smoothing  
15.1 Non-linear diffusion: Review  
15.2 Linear shape-adapted smoothing  
15.3 Affine scale-space  
15.4 Image texture descriptor based on affine scale-space  
15.5 Outlook  

Contents  xi
Appendix

A Technical details 395
  A.1 Implementing scale-space smoothing . . . . . . . . . . . . . . . . . . . . 395
  A.2 Polynomials satisfying the diffusion equation . . . . . . . . . . . . . . . 398

Bibliography 399

Index 415
1

Introduction and overview

In our daily life we use vision as one of our main sources of information about the outside world. Compared to a sense like hearing, the visual sense gives a richer description of the world. Compared to a sense like touch, it allows us to gather information about objects at greater distance and without affecting the objects themselves physically. Considering the apparent ease with which we obtain information about the world from the light that enters our eyes, an intellectual effort is required to appreciate that this is a non-trivial task.

Computer vision addresses this problem computationally; it deals with the problem of deriving meaningful and useful information from visual data. What should be meant by “meaningful and useful information” is, of course, dependent on the goal of the analysis, that is, the underlying purpose why we want to make use of visual information and process it with automatic methods. One reason may be that of machine vision—the desire to provide machines and robots with visual abilities. Typical tasks to be solved are object recognition, object manipulation, and visually guided navigation. The type of information that needs to be computed to address a problem depends strongly on the task. For example, the problem of recognizing objects from complex scenes is generally regarded as one of the more complicated problems in the field, while under certain conditions descriptors like time-to-collision can be computed with comparably simpler low-level operations. Other common applications of techniques from computer vision can be found in image processing, where one can mention image enhancement, visualization and analysis of medical data, as well as industrial inspection, remote sensing, automated cartography, data compression, and the design of visual aids, etc.

A more theoretical reason why computer vision is studied is the desire of understanding mathematical and physical principles underlying the inference of scene characteristics from brightness data. If insights into such basic principles can be gained, then they may help us with the tremendously inspiring challenge of understanding the workings of the biological visual systems, which accomplish their tasks in a way that is essential for the survival of most living creatures.
The problem of understanding vision has interested and puzzled researchers through the centuries. Still, some of the most basic questions that remain to be answered concern what type of image information is relevant for accomplishing different tasks, how this information should be extracted from the sensory data, and how such features can be related to properties of environment? An indication of the complexity of the vision problem can be obtained from the fact that the term “vision” has been very hard to define. Then, what definitions have been stated? To the question “What does it mean to see?” Marr (1982) answered

... vision is the process of discovering from images what is present in the world and where it is.

emphasizing that vision is an information-processing task. He also stressed that the issue of internal representation of information is of utmost importance. Only by representation can information be captured and made available to decision processes. The purpose of a representation is to make certain aspects of the information content explicit, that is, immediately accessible without any need for additional processing.

While Marr’s definition captures several important aspects, the active and goal-oriented nature of vision is only implicit in this formulation. Clearly, the vision problem is undefined unless related to a task. The existence proofs of vision provided by nature, the biological vision systems, are usually not passively registering images of the world. Instead, biological vision is strongly tied to action, since the visual agent has to attend to and respond to dynamic changes in the outside world. It is also well-known in perception psychology that perception of pictures differs from perception of the three-dimensional world.

These are some of the main arguments behind the active vision methodology (Bajcsy 1988; Aloimonos et al. 1988; Ballard 1991; Pahlavan et al. 1993), which has received increasing attention during recent years. In this paradigm, the ability of the vision system to selectively control the image acquisition process is emphasized. Moreover, the desired behaviour of the visual agent is put into focus. If the visual system is allowed to acquire more information in difficult situations, then several problems occurring in the analysis of given pre-recorded images can be avoided.

A simple example is the problem of too low a resolution. It can be circumvented by foveating interesting structures, or if necessary, moving closer to the interesting object. The active approach makes it possible to acquire additional information about three-dimensional structure from cues like accommodation distance, vergence angles, etc. An active moving observer also has the potential of avoiding unfortunate situations like accidentally aligned structures. It is sometimes argued that accidentally coinciding structures are very singular cases that never occur in practice, but in reality such situations turn out to show up rather frequently, when
taking overview images of moderately complex scenes using cameras with normal resolution and opening angles.

There have been, and still are, different opinions in the computer vision community about how a visual system should be constructed. A long debate concerned the choice between bottom-up and top-down based reasoning. It has been argued by many authors that a visual system should be constructed in a modular way with different levels of processing. At the simplest level of abstraction three layers can be distinguished, usually denoted low-level, intermediate level, and high-level.

Although a natural implication of the active vision paradigm is that it may not be as easy to clearly separate out different processing levels as would be needed for a dogmatic interpretation of such a simple three-layer description, and although extreme stand points have been taken, such as “direct pick-up” (Gibson 1979), “labyrinthic design” (Aloimonos 1990), or “intelligence without representation” (Brooks 1991), one should be careful of not interpreting the active vision approach as excluding the need for competence theories, like concerning the computation of early retinotopic representations such as intrinsic images (Barrow and Tenenbaum 1978). The need for some kind of early low-level processing and representation for providing a sparse but rich set of primitives for other processing modules still remains highly motivated.

This book deals with a basic aspect of early image representation—the notion of scale. More specifically, the work deals with a certain type of approach, the use of scale-space representation, for analysing image data at the very lowest levels in the chain of information processing of a visual system. The aim is to operate directly on the raw pixel values without any type of pre-processing. The suggested methodology is intended as a first confrontation between the reasoning process and the raw image data. This part of the visual system is usually termed the visual front-end. No specific assumptions will be made about how higher-level processes are to operate on the output. Therefore, the approach is applicable to a variety of reasoning strategies.

Computer vision is a cross-disciplinary field with research methodologies from several scientific disciplines such as computer science, mathematics, neurophysiology, physics, and psychology. The approach taken here will be computational.1 A theory and a framework will be proposed for how certain aspects of image information can be represented and analysed at the earliest processing stages of a machine vision system.

1Although there are neurophysiological and psychophysical evidence for the existence of processing at multiple scales in biological vision (Campbell and Robson 1977; Wilson 1983; Young 1985, 1987; Jones and Palmer 1987), no claims will be made that the methodology proposed here describes how processing is done in human perception. The treatment is concerned with what visual information can be extracted by a computer. When biological vision is discussed, it is mainly as a source of inspiration.
1.1. Theory of a visual front-end

If we are to construct a machine vision system, the problem can be addressed in several ways. If the visual task is sufficiently domain specific, then it may be sufficient to come up with any set of algorithms that perform the given task up to some prescribed tolerance. On the other hand, if the aim is to construct a flexible system able to solve a large number of problems using visual information, then it may be advantageous to aim at a certain degree of generality in the design, so that similar low-level modules can be shared between several algorithms or processes for solving different visual tasks. If such modules also are to be constructed without built-in limitations that would restrict their applicability, then a natural requirement is that the first stages of processing should make as few irreversible decisions and be as uncommitted as possible.

This presentation follows the latter strategy. If the vision problem is approached without strong presumptions about what specific tasks are to be solved, then a fundamental question concerns what information should be extracted at the earliest stages, and what kinds of operations are natural to perform on the data that reach the visual sensor. Is any type of operation feasible? An axiomatic approach that has been adopted in order to restrict the space of possibilities is to assume that the very first stages of processing should be able to function without any direct knowledge about what can be expected to be in the scene. For an uncommitted vision system, the scale-space theory states that under certain conditions, there is a natural choice of first stage operations to perform in a visual front-end (this notion will be made more precise later). The output from these operations can then, in turn, be used as input to a large number of other visual modules. An attractive property of this type of approach is that it gives a uniform structure on the first stages of computation.

1.2. Goal

The main subject of this book is to give a mathematical description of such early visual operations. The goal the work aims at is a methodology, in which significant structures can be extracted from an image in a solely bottom-up way, and scale levels can be selected for handling those structures without any prior information. A short summary in terms of key words can be expressed as follows:

A ranking of events in order of significance will be suggested based on volumes of certain four-dimensional objects in a scale-space representation of the signal. In this scale-space, the scale dimension is treated as equally important as the spatial and grey-level coordinates. The associated extraction method is based on a systematic parameter variation principle, where locally stable states are detected and abstractions are determined from those.
It will be exemplified how qualitative scale and region information extracted in this way can be used for guiding the focus-of-attention and tuning other early visual processes so as to simplify their tasks. The general principle is to adapt the low-level processing to the local image structure. The main theme of the book is to construct a theoretical framework in which these operations can be formalized.

1.3. The nature of the problem

When given an image as obtained from a standard camera device, say a digitized video signal or a scanned photograph, all information is encoded in the pixel values represented as a matrix of numerical data. If this information is presented to a human observer with the pixel values coded as grey-level intensities, then the human will usually have no problems in perceiving and interpreting what the image represents.

However, if the same pattern of grey-level values is coded as decimal digits, or as a three-dimensional diagram with the grey-level values drawn as a function of the image coordinates, then the problem is no longer as easy for biological vision. A person not familiar with the field often underestimates the difficulties in designing algorithms for interpreting data on this numerical form. The problem with the matrix representation of the image is that the information is only implicit in the data.

1.3.1. Ill-posedness

A major subtask of a visual processing system is to extract meaningful information about the outside world from such a set of pixel values, which is the result of light measurements from a physical scene. The image data may either be given beforehand, like in image processing, or have been acquired by an active system, which has directed its attention towards some interesting structure. What is meant by meaningful is in turn given by the task the vision system has to solve.

In principle, this problem of deriving three-dimensional shape information about the scene is impossible to solve if stated as a pure mathematical problem. Assume first that a set of grey-level data is given. Then, there will always be an infinite number of scenes that could have given rise to the same result. To realize that this is the case, consider for example a photograph on a paper, or a slide projected onto a screen. We easily interpret such brightness distributions on flat surfaces as corresponding to three-dimensional objects with perceived depth variations.

In an active vision system additional cues may be available, like accommodation depth, vergence, etc. Nevertheless, it is always possible to present two cameras with (possibly time varying) brightness patterns that would give the system a completely false impression of the world. There are two basic reasons to this. The first is that we are not measuring di-
rect properties of the world, but light emitted from it. The second is a dimensionality problem; we are trying to analyse a three-dimensional world using two-dimensional image data.

From this viewpoint the vision problem is ill-posed\textsuperscript{2} in the sense of Hadamard, since it does not have any unique solution. A rigorous person without plenty of unspoiled optimism would probably take this as a very good motivation to study some other field of science, where the prerequisites could be more clearly stated and better suited for formal analysis. Nevertheless, despite this indeterminacy, the human visual system as well as other biological vision systems are capable of coping with the ill-posedness. Moreover, since vision is generally regarded as the highest developed of our senses, one can speculate that there must be some inherent properties in the image data reaching the retina that make the visual perception\textsuperscript{3} possible.

1.3.2. Grouping

A main purpose of the low-level processing modules is to provide a reasonable set of primitives that can be used for further processing or reasoning modules. A fundamental problem in this context concerns what points in the image can be regarded as related to each other and correspond to objects in the scene, i.e., which pixels in the image can be assumed to belong together and form meaningful entities. This is the problem of primitive grouping or perceptual organization. Before any such grouping operations have been performed, the matrix of grey-level values is, from the viewpoint of interpretation, in principle only a set of numerical values laid out on a given discrete grid.

The grouping problem has been extensively studied in psychology, especially by the Gestaltists (Koffka 1935), as well as in computer vision (Lowe 1985; Ahuja and Tuceryan 1989), and it seems to be generally agreed upon that the existence of active grouping processes in human perception can be regarded as established. Witkin and Tenenbaum (1983) discuss this property:

People are able to perceive structures in images, apart from the perception of three-dimensionality, and apart from the recognition of familiar objects. We impose organization on

\textsuperscript{2}For a mathematical problem to be regarded as well-posed, Hadamard stated three criteria: (i) a solution should exist, (ii) the solution should be unique, and (iii) the solution should depend continuously on the input data. A well-posed problem is not necessarily well-conditioned.

\textsuperscript{3}Of course, experiences and expectations are generally believed to play an important role in the perception process. However, also that information must be related to the incoming image data in some way. Moreover, the experiences must have been acquired (learned) in some way, at least partially based on visual data.
1.3. The nature of the problem

data ... even when we have no idea what it is we are organizing. What is remarkable is the degree to which such naively perceived structure survives more or less intact once a semantic context is established: the naive observer often sees essentially the same thing as an expert does. ... It is almost as if the visual system has some basis for guessing what is important without knowing why.

Although the gestalt school of psychology formulated rules as those of proximity, similarity, closure, continuation, symmetry, and familiarity, we still have no satisfactory understanding of how these mechanisms operate from a quantitative point of view.

1.3.3. Operator size

To be able to compute any type of representation from image data, it is necessary to extract information from it, and hence interact with the data using some operators. Some of the most fundamental problems in low-level vision and image analysis concern what operators to use, where to apply them, and how large they should be. If these problems are not appropriately addressed, then the task of interpreting the output results can be very difficult.

Figure 1.1. Illustration of the basic scale problem when computing gradients as a basis for edge detection. Assume that the dots represent (noisy) grey-level values along an imagined cross-section of an object boundary, and that the task is to find the boundary of the object. The lines show the effect of computing derivative approximations using a central difference operator with varying step size. Clearly, only a certain interval of step sizes is appropriate for extracting the major slope of the signal corresponding to the object boundary. Of course, this slope may also be interpreted as due to noise (or some other phenomena that should be neglected) if it is a part superimposed on some coarser-scale structure (not visible here).
To illustrate this problem, consider the task of detecting edges. It is generally argued that this type of feature represents important information, since under reasonably general assumptions, edges in an image can be assumed to correspond to discontinuities in depth, surface orientation, reflectance properties, or illumination phenomena in the scene. A standard way of extracting edges from an image is by gradient computation followed by some type of post-processing step, where “high values” should be separated from “low values,” e.g., by detection of local maxima or by thresholding on gradient magnitude.

Consider, for simplicity, the one-dimensional case, and assume that the gradient is approximated by a central difference operator. More sophisticated approaches exist, but they will face similar problems. It is well-known that the selection of step size leads to a trade-off problem: A small step size leads to a small truncation error in the discrete approximation, but the sensitivity to fine-scale perturbations (e.g., noise) might be severe. Conversely, a large step size will, in general, reduce this sensitivity, but at the cost of an increased truncation error. In the worst case, a slope of interest can be missed and meaningless results be obtained if the difference quotient approximating the gradient is formed over a larger distance than the object considered in the image. See figure 1.1 for an illustration.

Although we shall here mainly be concerned with static images, the same kind of problem arises when dealing with image sequences. Similarly, models based on spatial derivatives ultimately rely on the computation of derivative approximations from measured data.

1.3.4. Scale

The problem falls back on a basic scale problem, namely that objects in the world and details in images, only exist as meaningful entities over limited ranges of scale, in contrast to certain ideal mathematical entities like “point,” “line,” “step edge,” or “linear slope,” which appear in the same way at all scales of observation.

A simple example is the concept of a branch of a tree, which makes sense only at a scale from, say, a few centimeters to as much a few meters. It is meaningless to discuss the tree concept at the nanometer or

---

4An interesting philosophical question in this context concerns whether or not the scale property should be attributed to the actual physical objects themselves or just to our subjective way of perceiving and categorizing them. For example, a table made out of wood certainly has a fine-scale texture with underlying fibral and molecular structures that we usually suppress when dealing with it for everyday purposes. Obviously, such fine-scale properties will always be there, but we almost always automatically disregard them. One may speculate that such an organization at multiple scales may be one way of simplifying the representation of our extremely complicated environment into a hierarchical structure to cope with it efficiently.
1.3. The nature of the problem

the kilometer level. At those scales it is more relevant to talk about the molecules that form the leaves of the tree, or the forest in which the tree grows. Similarly, it is only meaningful to talk about a cloud over a certain range of coarse scales. At finer scales it is more appropriate to talk about the individual droplets, which in turn consist of water molecules, which consist of atoms, which consist of protons and electrons etc.

This fact is well-known in the experimental sciences. In physics, the world is described at several levels of scales, from particle physics and quantum mechanics at fine scales, through thermodynamics and solid mechanics dealing with every-day phenomena, to astronomy and relativity theory at scales much larger than those we are usually dealing with. The physical description depends strongly on the scale at which the world is modelled. In biology, the study of animals can only be performed over a certain range of coarse scales. An organism looks completely different seen through a microscope when individual cells become visible.

These examples demonstrate that the scale concept is of crucial importance if one aims at describing the structure of the world, or more specifically the structure of projections of the world to two-dimensional data sets. As Koenderink (1984) has emphasized, the problem of scale must be faced in any image situation. The extent of any real-world object is determined by two scales, the inner scale and the outer scale. The outer scale of an object or a feature may be said to correspond to the (minimum) size of a window that completely contains the object or the feature, while the inner scale may loosely be said to correspond to the scale at which substructures of the object or the feature begin to appear.

In a given image, only structures over a certain range of scales can be observed. This interval is delimited by two scales; the outer scale corresponding to the finite size of the image, and the inner scale given by the resolution. For a digital image the inner scale is determined by the pixel size, and for a photographic image by the grain size in the emulsion.

1.3.5. Multi-scale representation

While these qualitative aspects of scale have been well-known for a long time, the concept of scale has been very hard to formalize into a mathematical theory. It is only during the last few decades that tools have been developed for handling the scale concept in a formal manner. A driving force in this development has come from the need for developing robust algorithms in image processing, computer vision, and other fields related to automatic signal processing.

A methodology that has been proposed for handling the notion of scale in measured data is by representing measured signals at multiple scales. Since, in general, no particular levels of scale can be pre-supposed without strong a priori knowledge, the only reasonable solution is that
the visual system must be able to handle image structures at all scales. The main idea of creating a multi-scale representation of a signal is by generating a one-parameter family of derived signals, where fine-scale information is successively suppressed. Then, a mechanism is required that systematically simplifies the data and removes finer-scale details, or high-frequency information. This operation, which will be termed scale-space smoothing, must be available at any level of scale.

![Image of multi-scale representation](image)

**Figure 1.2.** A multi-scale representation of a signal is an ordered set of derived signals intended to represent the original signal at different levels of scale.

Why should one represent a signal at multiple scales when all information is anyway in the original data? A major reason for this is to explicitly represent the multi-scale aspect of real-world data. Another aim is to suppress and remove unnecessary and disturbing details, such that later stage processing tasks can be simplified. More technically, the latter motivation reflects the common need for smoothing as a pre-processing step to many numerical algorithms as a means of noise suppression.

### 1.4. Scale-space representation

A methodology proposed by Witkin (1983) and Koenderink (1984) to obtain such a multi-scale representation of a measured signal is by embedding the signal into a one-parameter family of derived signals, the scale-space, where the parameter, denoted scale parameter $t \in \mathbb{R}_+$, is intended to describe the current level of scale.

---

footnote text: $\mathbb{R}_+$ denotes the set of real non-negative numbers, and $\mathbb{R}_+ \setminus \{0\}$ the corresponding set excluding the zero point.
1.4. Scale-space representation

1.4.1. Scale-space for one-dimensional signals: Gaussian smoothing

Let us briefly review this procedure as it is formulated for one-dimensional continuous signals: Given a signal \( f : \mathbb{R} \to \mathbb{R} \), the scale-space representation \( L : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R} \) is defined such that the representation at “zero scale” is equal\(^6\) to the original signal

\[
L(\cdot; 0) = f(\cdot),
\]

and the representations at coarser scales are given by convolution of the given signal with Gaussian kernels of successively increasing width

\[
L(\cdot; t) = g(\cdot; t) * f.
\]

In terms of explicit integrals, the result of the convolution operation ‘*’ is written

\[
L(x; t) = \int_{\xi=-\infty}^{\xi=\infty} g(\xi; t) f(x - \xi) \, d\xi,
\]

where \( g : \mathbb{R} \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R} \) is the (one-dimensional) Gaussian kernel

\[
g(x; t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.
\]

Figure 1.3 shows the result of smoothing a one-dimensional signal to different scales in this way. Notice how this successive smoothing captures the intuitive notion of fine-scale information being suppressed, and the signals becoming gradually smoother.

1.4.2. Diffusion formulation of scale-space

In terms of differential equations, the evolution over scales of the scale-space family \( L \) can be described by the (one-dimensional) diffusion equation

\[
\partial_t L = \frac{1}{\tau} \nabla^2 L = \frac{1}{\tau} \partial_{xx} L.
\]

In fact, the scale-space representation can equivalently be defined as the solution to (1.5) with initial condition \( L(\cdot; 0) = f(\cdot) \).

This analogy also gives a direct physical interpretation of the smoothing transformation. The scale-space representation \( L \) of a signal \( f \) can be understood as the result of letting an initial heat distribution \( f \) evolve over time \( t \) in a homogeneous medium. Hence, it can be expected that fine-scale details will disappear, and images become more diffuse when the scale parameter increases.

\(^6\)The notation \( L(\cdot; 0) = f \) stands for \( L(x; 0) = f(x) \ \forall x \in \mathbb{R}^N \).
Figure 1.3. The main idea with a scale-space representation of a signal is to generate a one-parameter family of derived signals in which the fine-scale information is successively suppressed. This figure shows a signal that has been successively smoothed by convolution Gaussian kernels of increasing width. (Adapted from Witkin 1983).

Figure 1.4. Schematic three-dimensional illustration of the scale-space representation of a one-dimensional signal.
1.4. Scale-space representation

1.4.3. Definition of scale-space: Non-creation of new structure

For a reader not familiar with the scale-space literature, the task of designing a multi-scale signal representation may at first glance be regarded as somewhat arbitrary. Would it suffice to carry out just any type of “smoothing operation”? This is, however, not the case. Of crucial importance when constructing a scale-space representation is that the transformation from a fine scale to a coarse scale really can be regarded as a simplification, so that fine-scale features disappear *monotonically* with increasing scale. If new artificial structures could be created at coarser scales, not corresponding to important regions in the finer-scale representations of the signal, then it would be impossible to determine whether a feature at a coarse scale corresponded to a simplification of some coarse-scale structure from the original image, or if it were just an accidental phenomenon, say an amplification of the noise, *created by the smoothing method—not the data*. Therefore, it is of utmost importance that artifacts are not introduced by the smoothing transformation when going from a finer to a coarser scale.

How should this property be formalized? When Witkin (1983) introduced the notion of scale-space, he was concerned with one-dimensional signals. He observed that the number of zero-crossings in the second derivative decreased monotonically with scale, and took that as a basic characteristic of the representation. In fact, this property holds for derivatives of arbitrary order, and also implies that the number of local extrema in any derivative of the signal cannot increase with scale. From this viewpoint, convolution with a Gaussian kernel possesses a strong smoothing property.

![Figure 1.5](image-url)

**Figure 1.5.** Since new zero-crossings cannot be created by the diffusion equation in the one-dimensional case, the trajectories of zero-crossings in scale-space (here, zero-crossings of the second derivative) form paths across scales that are never closed from below. (Adapted from Witkin 1983).
1.4.4. Uniqueness of the Gaussian

Later, when Koenderink (1984) extended the scale-space concept to two-dimensional signals, he introduced the notion of causality, which means that new level surfaces must not be created when the scale parameter is increased. Equivalently, it should always be possible to trace a grey-level value existing at a certain level of scale to a similar grey-level at any finer level of scale. By combining causality with the notions of homogeneity and isotropy, which essentially mean that all spatial points and all scale levels must be treated in a similar manner, he showed that the scale-space representation of a two-dimensional signal by necessity must satisfy the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (\partial_{xx} + \partial_{yy}) L. \quad (1.6)$$

Since convolution with the Gaussian kernel $g: \mathbb{R}^2 \times \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$

$$g(x, y; t) = \frac{1}{2\pi t} e^{-\left(x^2 + y^2\right)/2t} \quad (1.7)$$

describes the solution of the diffusion equation at an infinite domain, it follows that the Gaussian kernel is the unique kernel for generating a scale-space. This formulation extends to arbitrary dimensions.

Figure 1.6. The causality requirement means that level surfaces in scale-space must point with their concave side towards finer scales; (a) the reverse situation (b) must never occur.

A similar result based on slightly different assumptions, was given by Yuille and Poggio (1986) concerning the zero-crossings of the Laplacian of the Gaussian. Related formulations have been expressed by Babaud et al. (1986), and by Hummel (1987).

Another formulation was stated by Lindeberg (1990), who showed that the property of not introducing new local extrema with increasing scale by necessity lead to the Gaussian kernel if combined with a semi-group structure on the family of convolution kernels.

Florack et al. (1992) have elegantly shown that the uniqueness of the Gaussian kernel for scale-space representation can be derived under weaker conditions, by combining the semi-group structure of a convolution operation with a uniform scaling property over scales.
1.4. Scale-space representation

A notable similarity between these (and other) results is that several different ways of choosing scale-space axioms give rise to the same conclusion. The transformation given by convolution with the Gaussian kernel possesses a number of special properties, which make it unique. From the similarities between the different scale-space formulations, it can be regarded as well-established that within the class of linear transformations, the scale-space formulation in terms of the diffusion equation describes the canonical way to construct a multi-scale image representation.

An extensive review of basic properties about scale-space and related multi-scale representations is given in chapter 2. Before proceeding to the next subject, let us consider two every-day analogies concerning the need for multi-scale representation.

1.4.5. The scale parameter delimits the inner scale of observation

With respect to the notions of inner and outer scale, increasing the scale parameter in scale-space has the effect of increasing the inner scale of an observation. To appreciate the usefulness for this type of operation, consider the well-known printing method called dithering. It is used for producing impressions of grey-level information when printing images using only one colour of the ink (typically black). One of the most common techniques is to produce a pattern of very small black discs of different size. While the original image usually does not contain such structures, by averaging this intensity pattern over a local spatial neighbourhood, the effect will be the impression of a grey-level corresponding to grey-tone information. In this respect, this printing method makes explicit use of the multi-scale processing capabilities of our vision system.

1.4.6. Symbolic multi-scale representation

Referring to the analogies with other fields of science, the need for multi-scale representation is well understood in cartography. Maps are produced at different degrees of abstraction. A map of the world contains the largest countries and islands, and possibly, some of the major cities, while towns and smaller islands appear at first in a map of a country. In a city guide, the level of abstraction is changed considerably to include streets and buildings, etc. In other words, maps constitute symbolic multi-scale representations of the world around us, although constructed manually and with very specific purposes in mind.

It is worth noting that an atlas usually contains a set of maps covering some region of interest. Within each map the outer scale typically scales in proportion with the inner scale. A single map is, however, usually not sufficient for us to find our way around the world. We need the ability to zoom in to structures at different scales; i.e., decrease or increase the inner scale of the observation according to the type of situation at hand.
1.5. Philosophies and ideas behind the approach

1.5.1. Making information explicit

The scale-space theory constitutes a well-founded framework for handling structures at different scales. However, the information in the scale-space embedding is only implicit in the grey-level values. The smoothed images in the raw scale-space representation contain no explicit information about the features in them or about the relations between features at different levels of scale.

One of the main goals of this book is to present such an explicit representation called the scale-space primal sketch, and to demonstrate that it enables extraction of significant image structures in such a way that the output can be used for guiding later stage processes and simplifying their tasks. The treatment will be concerned with intensity images, the grey-level landscape, and the chosen objects will be blobs, that is, bright regions on dark backgrounds or vice versa. However, the methodology applies to any bounded function and is therefore useful in many tasks occurring in computer vision, such as the study of level curves and spatial derivatives in general, depth maps, and histograms, point clustering and grouping, in one or in several dimensions. Moreover, the underlying principles behind its construction are general, and extend to other aspects of image structure.

1.5.2. Scale and segmentation

Many methods in computer vision and image analysis implicitly assume that the problems of scale detection and initial segmentation have already been solved. Models based on spatial derivatives ultimately rely upon the computation of derivative approximations, which means that they will face similar scale problems as were described in the discussion about edge detection from gradient data in section 1.3.3. Although we shall here be mainly concerned with static imagery, the same type of problems arise also when dealing with image data over time. In other words, when computing derivatives from measured data, we in general always fall back to the basic scale problem of selecting a filter mask size\(^7\) for the approximation.

A commonly used technique for improving the results obtained in computer vision and other fields related to numerical analysis is by preprocessing the input data with some amount of smoothing or careful tuning of the operator size or some other parameters. In some situations the output may depend strongly on these processing steps. In certain algorithms these tuning parameters can be estimated; in other cases they are set manually. A robust image analysis method intended to work in

\(^7\)Observe that it is not the actual size of the filter mask that is important, but rather the characteristic length over which the difference approximation is computed.
autonomous robot situation must, however, be able to make such decisions automatically. How should this be done? I contend that these problems are in many situations nothing but disguised scale problems.

Also, to apply a refined mathematical model like a differential equation or some kind of deformable template, it is necessary to have some kind of qualitative initial information, e.g., a domain where the differential equation is (assumed to be) valid, or an initial region for applying the raw deformable template. Examples can be obtained from many "shape-from-X" methods, which in general assume that the underlying assumptions are valid in the image domain the method is applied to. A commonly used assumption is that of smoothness implying that the region in the image, to which the model is applied to, must correspond to, say, one physical object, or one facet of a surface. How should such regions be selected automatically? Many methods cannot be used unless this non-trivial part of the problem is solved.

How can we detect appropriate scales and regions of interest when there is no a priori information available? In other words, how can we detect the scale of an object and where to search for it before knowing what kind of object we are studying and before knowing where it is located. Clearly, this problem is intractable if stated as a pure mathematical problem. Nevertheless, it arises implicitly in many kinds of processes (e.g., dealing with texture, contours etc.), and seems to boil down to an intractable chicken-or-the-egg problem. The solution of the pre-attentive recognition problem seems to require the solution of the scale and region problems and vice versa.

The goal of this presentation is to demonstrate that such pre-attentive groupings can be performed in a bottom-up manner, and that it is possible to generate initial hypotheses about regions of interest as well as to give coarse indications about the scales at which the regions manifest themselves. The basic tools for the analysis will be scale-space theory, and a heuristic principle stating that blob-like structures which are stable in scale-space are likely candidates to correspond to significant structures in the image. Concerning scale selection, scale levels will be selected that correspond to local maxima over scales of a measure of blob response strength. (Precise definitions of these notions will be given later.) It will be argued that once such scale information is available, and once regions of interest have been extracted, later stage processing tasks can be simplified. This claim is supported by experiments on edge detection and classification based on local features.

1.5.3. Detection of image structure

The main features that arise in the (zero-order) scale-space representation of an image are smooth regions which are brighter or darker than
the background and stand out from their surroundings. These will be termed blobs. The purpose of the suggested representation is to make these blobs explicit as well as their relations across scales. The idea is also that the representation should reflect the intrinsic shape of the grey-level landscape—it should not be an effect of some externally chosen criteria or tuning parameters. The theory should in a bottom-up fashion allow for a data-driven detection of significant structures, their relations, and the scales at which they occur. It will, indeed, be experimentally shown that the proposed representation gives perceptually reasonable results, in which salient structures are (coarsely) segmented out. Hence, this representation can serve as a guide to subsequent, more finely tuned processing, which requires knowledge about the scales at which structures occur. In this respect it can serve as a mechanism for focus-of-attention.

Since the representation tries to capture important image structures with a small set of primitives, it bears some similarity to the primal sketch proposed by Marr (1976, 1982), although fewer primitives are used. The central issue here, however, is to represent explicitly the scales at which different events occur. In this respect the work addresses problems similar to those studied by Bischof and Caelli (1988). They tried to parse scale-space by defining a measure of stability. Their work, however, was focused on zero-crossings of the Laplacian. Moreover, they overlooked the fact that the scale parameter must be properly treated when measuring significance or stability. Here, the behaviour of structures over scale will be analysed in order to give the basis of such measurements.

Of course, several other representations of the grey-level landscape have been proposed without relying on scale-space theory. Let us also note that Lifshitz and Pizer (1990) have studied the behaviour of local extrema in scale-space. However, we shall defer discussing relations to other work until the suggested methodology has been described.

1.5.4. Consistency over scales

The idea of scale-space representation of images, suggested by Witkin (1983) has, in particular, been developed by Koenderink and van Doorn (1984, 1986, 1992), Babaud et al. (1986), Yuille and Poggio (1986), Hummel (1987), Lindeberg (1990, 1993), and Florack et al. (1992). This work is intended to serve as a complement addressing computational aspects, and adding means of making significant structures and scales explicit.

The main idea of the approach is to link similar structures (here blobs) at different levels of scales in scale-space into higher-order objects (here four-dimensional objects called scale-space blobs), and to extract significant image features based on the appearance and lifetime of the higher-order objects in scale-space. A basic principle that will be used is that significant image features must be stable with respect to variations in scale.
1.6. Relations to traditional applied mathematics

Another important point within the work is that the scale parameter is treated as being as equally important as the spatial and grey-level coordinates. This is directly reflected in the fact that the primitives in the representation are objects having extent not only in space and grey-level, but also in scale.

1.6. Relations to traditional applied mathematics

In principle, we are to derive information from image data by operating on it with certain operators. An obvious question to ask is then why this problem could not be seen as an ordinary standard problem in numerical analysis and be solved with standard numerical techniques? Let us point out several reasons as to why the problem is hard.

1.6.1. Modelling, simulation, and inverse problem

Traditional numerical analysis is often concerned with the simulation of mathematical or physical models (for example, formulated as discrete approximations to continuous differential equations, which are rather good descriptions of the underlying reality). The problems are usually well-defined, the models can often be treated as exact, and the errors involved in these types of computations are mainly due to discretization and round-off errors.

In computer vision the situation is different. Given a signal, the task is to analyse and extract information from it. We are trying to solve an inverse problem, where the noise level is generally substantially higher\footnote{A rule of thumb sometimes used in this context is that when derivatives of order higher than two are computed from raw image data, the amplitude of the amplified noise will often be of the same order of magnitude as the derivative of the signal or be even higher.} and the modelling\footnote{The geometry of image formation is quite simple and well understood, but our knowledge about the complicated physical phenomena (comprising reflections, etc.), and how to model them from a computational viewpoint, is still rather vague. In addition, we have the problem of representing the enormous variety of different situations that can occur in the real world, as well as the question of how cognitive aspects should be incorporated into the process.} aspect is still open. With a precise model of the illumination situation as well as the reflectance properties of the surfaces in the environment, one could conceive solving for the surface geometry based on the physical light characteristics. However, it is well-known that this problem of reconstructing the world is extremely hard, to a large extent because it is very difficult to formulate an accurate and physically useful model for the image formation process, but also because such a model would require much additional a priori knowledge in order to be computationally tractable. Although further attempts to explore the situation in more detail are being made (Forsyth and Zissermann 1989; Nayar et
al. 1990), most shape-from-shading and similar algorithms still rely on very restrictive simplifying assumptions.

1.6.2. Scale and resolution

Other aspects are those of scale and resolution. In numerical analysis the accuracy can often be increased by a refinement in the grid sampling. The selection of a larger grid size is mainly motivated by efficiency reasons, since exact equations are usually simulated. In computer vision algorithms the number of grid points used for resolving structures in a given image is sometimes very low, which makes a difficult problem even more difficult. This restriction can be relaxed, however, in an active vision situation, as will be developed in section 11.3.

A more serious problem is that of scale. In most standard numerical problems the inner scale is zero, which means that the smaller the grid size that is being used, the higher will be the accuracy in computations (compare again with the example in section 1.3.3). In easy problems, the solutions asked for contain variations taking place on essentially a single scale. Problems having solutions with variations on different scales are more complicated and require more advanced algorithms for their solution. Examples can be obtained from computer fluid dynamics, where turbulence and very thin boundary layers are known to lead to very hard numerical problems. These fine-scale phenomena cannot always be fully resolved by discrete approximations, and in fact some type of (sometimes artificial) smoothing (dissipative terms) is often required. When the fine-scale phenomena are not properly dealt with, they can interfere with and disturb the coarse-scale phenomena that usually are the ones of interest in, for example, design applications. Moreover, the occurrences of discontinuities in the solutions, which are also very frequent in image data, are known to complicate the situation further.

The idea with scale-space representation is to separate out information at different scales. Note, that this may be a difficult problem, since in general, very little or no a priori knowledge can be expected about what types of structures the visual system is studying, or at what scales they occur.

1.6.3. Interpreting the results

If an operator is applied all over an image, then it will at best give reasonable answers in those regions in which the underlying assumptions for the method are valid (provided that the operator size has been appropriately tuned). However, the operator also gives false alarms in regions where the assumptions are not satisfied. One could say that such a uniform application of an operator enforces an answer in every point even though any well-defined answer does not exist. In general, it is hard to
distinguish from the output of such an operation which responses can be trusted as correct and which ones should be rejected. Plain thresholding on the magnitude of the response is usually not sufficient. Therefore, a conservative strategy is to aim at deriving a sparse set of safe and reliable cues at the risk of "missing" a few that could be included rather than to try to compute "every" feature at the risk of including a large number of false responses. This is the motivation for trying to determine in advance where to apply\textsuperscript{10} refined operations.

1.6.4. Approximation and regularization

It is sometimes argued that the main aims of approximation theory have already been accomplished. Nevertheless, one is confronted with serious problems when applying this theory to irregular and noisy measurement data like those obtained from images. Some of the most basic problems concern how to determine a region in space appropriate for fitting a model to the data, and how one should tune the associated parameters (such as the filter weights). An approach that has been extensively used in computer vision during the last decade is regularization. This technique has been applied to a variety of reconstruction problems (see Terzopoulos 1986; Terzopoulos et al. 1987, 1988; Kass et al. 1987; Witkin et al. 1987; Blake and Zisserman 1987; Pentland 1990; Aloimonos and Schultmann 1990). The basic methodology is to define a functional, which is a weighted combination of different error criteria, and then try to compute the function within some restricted space that minimizes it. These methods often contain a large number of parameters but the theory usually gives little or no information about how they should be set without manual intervention, although attempts have been made to learn them from examples. In addition there is a verification problem, since the algorithm is forced to always find a solution within the given space. How does one determine whether that function resembles the answer we actually want (the answer to the original problem)? The solution to a regularized problem is, in general, not equal to the solution to the original problem, not even if the input data are exact. To summarize, both these types of methods require a careful setting of their associated parameters, as well as the regions in space to which they should be applied.

1.6.5. Principles behind the work

A basic intention behind the approach taken here is to pre-process the data and to derive context information from it in such a way that the

\textsuperscript{10}This is a problem arising mainly in an initialization phase of a reasoning process. When a time aspect is present, this problem is simplified, since context knowledge can be used for predictions about the future. It is generally argued that problems become easier once the boot-strapping step has been performed.
output from these types of operations can be well-defined. Although no claims are made that these problems have been solved, and even though further complications may appear on the way to the solution, I believe that the framework to be developed here represents a significant step toward posing the questions in a context where standard numerical techniques could be readily applied and give useful answers.

1.7. Organization of this book

The book deals with the fundamental problems that are associated with the use of scale-space analysis in early processing of visual information. More specifically some of the main questions it addresses are the following:

- How should the scale-space model be implemented computationally? The scale-space theory has been formulated for continuous signals, while realistic signals are discrete.

- Can the scale-space representation be used for extracting information? How should this be done?

- The scale-space representation in itself contains no information about preferred scales. In fact, without any a priori scale information all levels of scale must be treated similarly. Is it possible to determine a sparse set of appropriate scales for further processing?

- How can the scale-space concept interact with and cooperate with other processing modules?

- What can happen in scale-space? What is the behaviour of structure in scale-space? How do features evolve under scale-space smoothing? What types of bifurcation events can take place?

- Can cues to three-dimensional surface shape be computed directly from visual front-end operations?

The presentation is divided into four parts. We start by considering the basic theory of scale-space representation. A number of fundamental results on scale-space and related multi-scale representations are reviewed. The problem of how to formulate a scale-space theory for discrete signals is treated, as is the problem of how to compute image features within the Gaussian derivative framework.

Then, a representation called the scale-space primal sketch is presented, which is a formal representation of structures at multiple scales in scale-space aimed at the making information in the scale-space representation explicit. The theory behind its construction is analysed, and an algorithm is presented for computing the representation.
1.7. Organization of this book

It is demonstrated how this representation can be integrated with other visual modules. Qualitative scale and region information extracted from the scale-space primal sketch can be used for *guiding other low-level processes and simplifying their tasks*.

Finally, it is shown how the suggested method for *scale selection* can be extended to other aspects of image structure, and how *three-dimensional shape cues* can be computed within the Gaussian derivative framework. Such information can then be used for adapting the shape of the smoothing kernel, to reduce the shape distorting effects of the scale-space smoothing, and thus increase the accuracy in the computed surface orientation estimates.

1.7.1. Guide to the reader

As a guide to the reader it should be remarked that it is not necessary to read this book in a sequential manner. While the ordering of the chapters follows the bottom-up chain of processing levels in an imagined vision system, the chapters are written so that it should be possible to read them independently and still get the major ideas without having to digest the preceding chapters. The following table describes the mutual dependencies.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Contents</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Review of multi-scale analysis</td>
<td></td>
</tr>
<tr>
<td>3, 4</td>
<td>Discrete scale-space theory</td>
<td>(3, 4)</td>
</tr>
<tr>
<td>5</td>
<td>Computing derivatives in scale-space</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Feature detection in scale-space</td>
<td>(5)</td>
</tr>
<tr>
<td>7</td>
<td>The scale-space primal sketch</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Theoretical analysis of scale-space</td>
<td>(7)</td>
</tr>
<tr>
<td>9</td>
<td>Algorithm for blob linking</td>
<td>7, 8</td>
</tr>
<tr>
<td>10</td>
<td>Extracting salient image structures</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>Guiding processes with scale-space</td>
<td>7, 10</td>
</tr>
<tr>
<td>12</td>
<td>Summary and discussion of chapters 7–11</td>
<td>7–11</td>
</tr>
<tr>
<td>13</td>
<td>Scale selection</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Shape computation</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>Non-uniform smoothing</td>
<td>(14)</td>
</tr>
</tbody>
</table>

The level of presentation varies depending on the subjects. Some chapters are highly mathematical, while others are more descriptive. For a reader who wants to avoid the mathematics at first, I recommend chapters 7, 10, and 11 for getting the basic ideas of the approach. Then, it may be natural to proceed with chapters 6, 13, and 14, where straightforward descriptions can be found of how to use the scale-space methodology.
for different types of early visual computations. The basic scale-space theory underlying these chapters is described in chapters 3–5, which give a detailed mathematical analysis of scale-space theory for discrete signals, and chapter 8, which shows how the behaviour of image structures in scale-space can be analysed.

Now, in the form of a long abstract, a brief overview will be given of some of the main results presented in each of the different parts.

1.7.2. Part I: Basic scale-space theory

Chapter 2: Review of multi-scale analysis. A summary is given of basic properties of scale-space and related multi-scale representations, notably, pyramids, wavelets, and regularization. A number of special properties of the scale-space representation are listed, and the different multi-scale approaches are compared.

Chapter 3: One-dimensional discrete scale-space theory. Which convolution kernels share the property of never introducing new local extrema in a signal? Qualitative properties of such kernels are pointed out, and a complete classification is given.

These results are then used for showing that there is only one reasonable way to define a scale-space for one-dimensional discrete signals, namely by discrete convolution with a family of kernels called the discrete analogue of the Gaussian kernel. This scale-space can equivalently be described as the solution to a semi-discretized version of the diffusion equation. The conditions that single out this scale-space are essentially non-creation of local extrema combined with a semi-group assumption and the existence of a continuous scale parameter. Similar arguments applied in the continuous case uniquely lead to the Gaussian kernel.

The commonly adapted technique with a sampled Gaussian may lead to undesirable effects (scale-space violations). This result exemplifies the fact that properties derived in the continuous case might be violated after discretization.

Chapter 4: Discrete scale-space theory in higher dimensions. The one-dimensional scale-space theory is generalized to discrete signals of arbitrary dimension. The treatment is based upon the assumptions that (i) the scale-space representation should be defined by convolving the original signal with a one-parameter family of symmetric smoothing kernels possessing a semi-group property, and (ii) local extrema must not be enhanced when the scale parameter is increased continuously.

Given these requirements, the scale-space representation must satisfy a semi-discretized version of the diffusion equation. In a special case the representation is given by convolution with the one-dimensional discrete analogue of the Gaussian kernel along each dimension.
Chapter 5: Computing derivatives in scale-space. It is shown how discrete derivative approximations can be defined so that scale-space properties hold exactly also in the discrete domain. A family of kernels is derived which constitute discrete analogues to the continuous Gaussian derivatives, and possesses an algebraic structure similar to that possessed by the derivatives of the traditional scale-space representation in the continuous domain.

The representation has theoretical advantages compared to other discretizations of scale-space theory in the sense that operators which commute before discretization commute after discretization. Some computational implications of this are that derivative approximations can be computed directly from smoothed data (without any need for repeating the smoothing operation), and this will give exactly the same result as convolution with the corresponding derivative approximation kernel. Moreover, a number of normalization conditions are automatically satisfied.

Chapter 6: Feature detection in scale-space. The proposed methodology leads to a conceptually simple scheme of computations for multi-scale low-level feature extraction, consisting of four basic steps; (i) large support convolution smoothing, (ii) small support difference computations, (iii) point operations for computing differential geometric entities, and (iv) nearest neighbour operations for feature detection.

Applications are given demonstrating how the proposed scheme can be used for edge detection and junction detection based on derivatives up to order three.

1.7.3. Part II: Theory of the scale-space primal sketch

Chapter 7: The scale-space primal sketch. A representation is presented for making explicit image structures in scale-space as well as the relations between image structures at different scales. The representation is based on blobs that are either brighter or darker than the background. At any scale in scale-space grey-level blobs are defined at that scale. Then, these grey-level blobs are linked across scales into objects called scale-space blobs. The relations between these blobs at different scales define a hierarchical data structure called the scale-space primal sketch, and it is proposed that the volume of a scale-space blob in scale-space constitutes a natural measure of blob significance.

To enable comparisons of significance between structures at different scales, it is necessary to measure significance in such a way that structures at different scales are treated in a uniform manner. It is shown how a definition of a transformed scale parameter, effective scale, can be expressed such that it gives intuitive results for both continuous and discrete signals. The volumes of the grey-level blobs must be transformed in a similar
manner. That normalization is based on simulation results accumulated for a set of reference signals.

Chapter 8: Theoretical analysis of scale-space. It is demonstrated how the behaviour of image structures over scales can be analysed using elementary techniques from real analysis, singularity theory, and statistics.

The implicit function theorem describes how critical points form trajectories across scales when the scale parameter changes, and gives direct estimates of their drift velocity. Momentarily, the drift velocity may tend to infinity. Generically, this occurs in bifurcation situations only.

The qualitative behaviour of critical points at bifurcations is analysed, and the generic blob events are classified. A set of illustrative examples is presented, demonstrating how blobs behave in characteristic situations.

Chapter 9: Algorithm for blob linking. An algorithm is described for computing the scale-space primal sketch. It is based on detection of grey-level blobs at different levels of scale. On that output data an adaptive scale sampling algorithm operates and performs the actual linking of the grey-level blobs into scale-space blobs as well as the registration of the bifurcations and the blob events.

1.7.4. Part III: Applications of the scale-space primal sketch

Chapter 10: Extracting salient image structures. It is experimentally demonstrated how the scale-space primal sketch can be used for extracting significant blob-like structures from image data as well as associated scale levels for treating those. Such descriptors constitute coarse segmentation cues, and can serve as regions of interest to other processes.

The treatment is based on two basic assumptions; (i) in the absence of other evidence, structures, which are significant in scale-space, are likely to correspond to salient structures in the image, and (ii) in the absence of other evidence, scale levels can be selected where the blob response assumes its maximum over scales.

Chapter 11: Guiding processes with scale-space. It is demonstrated how the qualitative scale and region descriptors extracted by the scale-space primal sketch can be used for guiding other processes in early vision and for simplifying their tasks.

An integration experiment with edge detection is presented, where edges are detected at coarse scales given by scale-space blobs, and then tracked to finer scales in order to improve the localization. In histogram analysis, the scale-space primal sketch is used for automatic peak detection. More generally, such descriptors can be used for guiding the focus-of-attention of active vision systems. With respect to a test problem of detecting and classifying junctions, it is demonstrated how the blobs can
be used for generating regions of interest, and for providing coarse context information (window sizes) for analysing those.

Finally, it is briefly outlined how the scale-space primal sketch can be applied to other visual tasks such as texture analysis, perceptual grouping and matching problems. Experiments on real imagery demonstrate that the proposed theory gives intuitively reasonable results.

Chapter 12: Summary and discussion of the scale-space primal sketch approach (chapters 7–11). Basic properties of scale-space representation and the scale-space primal sketch are pointed out, and relations to previous work are described. A summary is given of the basic ideas, and a few alternative approaches are discussed.

1.7.5. Part IV: Scale selection and shape computation

Chapter 13: Scale selection. A heuristic principle for scale selection is proposed stating that local extrema over scales of different combinations of normalized scale invariant derivatives are likely candidates to correspond to interesting structures. The resulting methodology lends itself naturally to two-stage algorithms; feature detection at coarse scales followed by feature localization at finer scales. Support is given by theoretical considerations and experiments on blob detection, junction detection, and edge detection.

Chapter 14: Shape computation by scale-space operations. The problem of scale in shape-from-texture is addressed. The need for (at least) two scale parameters is emphasized; a local scale describing the amount of smoothing used for suppressing noise and irrelevant details when computing primitive texture descriptors from image data, and an integration scale describing the size of the region in space over which the statistics of the local descriptors is accumulated.

The mechanism for scale selection outlined in chapter 13 is used for adaptive determination of the two scale parameters in a multi-scale texture descriptor, the windowed second moment matrix, which is defined in terms of Gaussian smoothing, first-order derivatives, and non-linear pointwise combinations of these. This texture description can then be combined with various assumptions about surface texture in order to estimate local surface orientation. Two specific assumptions, “weak isotropy” and “constant area,” are explored in more detail. Experiments on real and synthetic reference data with known geometry demonstrate the viability of the approach.
Chapter 15: Non-uniform smoothing. Various generalizations of linear and rotationally symmetric Gaussian smoothing are briefly described.

A special approach of performing linear shape adaption in shape-from-texture is treated in more detail. It is demonstrated how an affine scale-space representation can be used for defining an image texture descriptor that possesses useful invariance properties with respect to linear transformations of the image coordinates.
Bibliography


Bibliography


D. Koller, K. Daniilidis, T. Thórhallson, and H.-H Nagel, “Model-based object tracking in traffic scenes,” in *Proc. 2nd European Conf. on Computer Vision*, (Santa Margherita Ligure, Italy), 1992.


P. Perona, “Steerable-scalable kernels for edge detection and junction analysis,” in Proc. 2nd European Conf. on Computer Vision, (Santa Margherita Ligure, Italy), pp. 3–18, May 1992.


Bibliography


Index

absolute error, 396
abstraction
  from parameter variations, 251, 294
accidental grouping, 184, 312
accommodation
  depth from, 298–299
active vision, 2
  active analysis of junctions, 285–300
  avoiding boundary effects, 121
adaptive refinement, 185
  continuous scale, 82–83
  in blob linking, 232
aerosol images
  analysis of, 300–304
affine intensity transformation, 152, 311
affine scale-space, 387–390
  interpretation, 389
  transformation property, 387
anisotropic diffusion, 59, 383
anisotropic smoothing, 119–120
anisotropy, 355
annihilation
  of blob, see blob annihilation
  of saddle-extremum pair, 197
aperture of observation, 41, 124
approximation theory, 21
area gradient, 352, 353, 358
autocorrelation function, 217
bandpass pyramid, 37–38, 55
base-level
  definition, 168
  of grey-level blob, 167
  registration 1-D, 227
  registration 2-D, 231
bifurcation, 172, 173, 185, 198
  multiple blob responses, 258
  set, 196
bifurcation events
blobs, 201–204
  critical points, 194–200
binomial kernel, 66–68, 80–82
  separable 2-D case, 117
blob
  requirements of, 165
blob annihilation, 172
    definition, 202
    extended neighbourhood search, 238
    registration of, 241
    weak candidate, 237
blob creation, 172
    definition, 202
    example, 203
    extended neighbourhood search, 238
    registration of, 241
    weak candidate, 237
blob detection, 159, 325–327, 371–374
  methods, 310–311
  scheme for, 252
blob events, 172, 173
  complex, 239
  for blob linking, 232
  generic types, 172, 202–204
  scale determination, 243
  strong conditions, 237
  weak conditions, 236
blob linking, 172, 177, 232–242
  basic algorithm, 240–242
blob merge, 172
  definition, 202
  registration of, 242
  strong candidate, 237
  weak candidate, 237
blob split, 172
  definition, 202
  example, 203
registration of, 242
strong candidate, 238
weak candidate, 237
blob-blob matching, 233–234
blob-edge matching, 273–277
blob-extremum matching, 234–236
blob-initiated edge focusing, 277–281
bright blob, 166, 167, 170
relation to dark blob, 169
cascade smoothing, 42, 46, 125, 128–132, 145
Fourier domain, 42
causality, 14, 47, 103, 106
central limit theorem, 53
characteristic length, 16, 32, 178
descriptor, 40
circulant matrix, 71
clustering, see grouping
colour segmentation, 282–285
commutative properties, 46, 47
affine scale-space, 388
feature detectors, 153
scale-space derivatives, 125–138
comparison
derivative approximations, 145
scale-space vs. pyramid, 58
conductivity, 120, 383
connected-component-labelling, 229, 230
consistent derivative approximation, 129, 145
continuous scale parameter, 82–83
contrast
of grey-level blob, 168
significance measure, 177
convolution matrix
circulant, 71
eigenvectors, eigenvalues, 72, 74
minors, 79
no real eigenvalues, 69, 70
covariance matrix, 40, 221, 354
creation
of blob, see blob creation
of new extrema
d example, 101–103
of saddle-extremum pair, 197
cross operator, 105
cubic spline, 57
curvature blob, 292–297
curvature of level curve, 151, 158–160, 291–293
in non-linear diffusion, 384
curve
multi-scale representation, 59
cusp singularity, 195
in scale-space, 210–212
dark blob, 169, 170
relation to bright blob, 169
data compression, 38
decreasing amplitude, 309–310, 318
deep structure, 187
delimiting saddle point, 201
definition, 168
illustration, 167
importance in bifurcations, 172
registration of, 230
delta function, 86, 136
density gradient, 352
density in scale-space
of local extrema, 217–226
depth-from-focus, see accommodation
derivative approximation, 8, 16, 123–148
discrete scale-space properties
necessity, 132
sufficiency, 132
implementation, 397
in feature detection, 154, 158
scale-space properties transfer, 119
difference of Gaussians, 53, 138
difference of offset Gaussians, 55
difference operator, 123, 127
1-D, 133
2-D, 133
backward, 139
commutes with scale-space smoothing, 128
second order, 133
sensitivity to step length, 8, 124
differentiability, 106
from scale-space axioms, 107, 131
infinite in scale-space, 45
differential geometric descriptor, 126, 149–161
scheme for computation, 136
Index

differential invariant, 149–152
differential singularities, 152–153
differential singularity
  drift velocity in scale-space, 214–217
diffuse L-junction, 329
diffuse step edge, 138, 329, 339
diffusion equation, 11
  discretization, 61, 100
  N-D, 43
  necessity for scale-space, 48
  semi-discretized, 94–97, 128
  N-D, 118
diffusion polynomials, 398
dimensional analysis, 49
dimensionless coordinate, 46
direct computation, 135, 349
directional derivative, 139
directional derivatives, 150–151
discrete analogue of Gaussian
definition, 86
  derivative approximation, 145
  filter coefficient generation, 396
  Fourier transform, 87
  kernel graphs, 139
  properties, 86
  separable case N-D, 115
  truncation, 396
  uniqueness, 84–86
  variance, 87
discrete iterations, 116
discrete scale-space, 61–122
  spatial isotropy, 117
  finite domain, 120
  Fourier transform, 112
  generating function, 112
  grids types, 121
  necessity, 109–111
  separability, 114–115
  sufficiency, 111–112
  unimodality, 114
  uniqueness, 85
discrete scale-space kernel, 96
  classification, 79–82
  definition, 63
  properties, 65–75
  distribution theory, 45
double asymmetric step edge, 340
double blob link
  strong candidate, 238
  drift velocity
  critical point, 190, 213
  curved edge, 216
  differential singularities, 214–217
  in blob linking, 243
  junction, 215
  parabolic curve, 217
  straight edge, 190
  unbounded in scale-space, 191
  zero-crossing of Laplacian, 217
dynamic shape, 187
eccentricity, 54
edge definition, 153–155
  discrete sub-pixel interpolation, 154–155
edge detection, 7, 153–158, 161, 162, 272–282, 339–343
discrete modelling, 138
  regularization, 57
edge focusing, 272–273, 277–281
effective grey-level blob volume
  definition, 184
effective scale, 180–182
  continuous signals, 219–220
  continuous vs. discrete, 223–225
  discrete signals, 220–223
equal contribution
  condition in pyramid, 35
equi-ripple design, 36
Expand, 37
explicit information, 2, 16
extrema
  density in scale-space, 181–182, 217–226
  extremum path, 47
  definition, 192
fast Fourier transform, 397
feature detection, 149–162
  scheme for, 145, 161
filter size
  pyramid, 36
  scale-space, 395
finite domain, 120
first moment vector, 40
five-point operator, 105
fixation, 295–297
fixed readout capacity, 54
focus-of-attention
fold singularity, 195
  in scale-space, 206–210
foreshortening effect, 351
Förstner operator, 332–334
Fourier spectrum
  self-similar, 346
Fourier transform
  definition, 42
  discrete Gaussian, 87
  discrete scale-space 2-D, 113
  non-uniform Gaussian, 390
  of Gaussian, 51
  pyramid, 37
  unimodality 2-D, 114
foveation
  definition, 286
  simulated, 291
Gabor function, 55, 59
Gaussian blob, 340, 385
Gaussian derivative, 126
  kernel graphs, 139
Gaussian derivatives
  biological vision, 54
Gaussian kernel, 11, 39
  discrete, see discrete analogue
    of Gaussian
  integrated, see integrated Gaussian kernel
  N-D, 39
  non-uniform, see non-uniform Gaussian kernel
Pólya frequency function, 89
sampled, see sampled Gaussian kernel
  special properties, 40–53
  uniqueness, 47–52
gaze transformation, 353
generalized binomial kernel, see bi-
  nomial kernel
generalized function, 124
generating function
  characterization of discrete scale-
    space kernel, 79
  definition, 67
  discrete scale-space 2-D, 113
  root of, 68
generic signal, 167
  discrete definition, 228
geometric coincidence, 273–276
grey-level blob, 166–169
  contrast, 168
  definition, 166–168
  detection algorithm, 227–232
  detection invariants, 229–230
  experiments, 174–177
  in scale-space, 171–177
  properties, 169
  relation Laplacian blob, 311
grey-level blob tree, 170–171
  computation, 232
grey-level blob volume
  definition, 168
  transformed, 183–184
  variation over scales, 183
grid types, 121
grouping, 6–7, 17, 165, 184, 250, 258,
  306–313
guide to reader, 23–28
  guiding visual processes, 271–306
Hadamard, 6
harmonic oscillator, 52
head-eye system, 285, 294
heat distribution, 43
Hermite function, 52
Hermite polynomial, 51
histogram, 287
histogram classification, 282–285
homogeneity, 14, 47
hypothesis generation, 256, 281, 285
hysteresis thresholding, 162
ideal low-pass filter, 36, 99
ill-posed problem, 5–6, 57
implementing scale-space smoothing,
  395–397
implicit function theorem, 189
infinite differentiability, 45
infinitesimal generator, 107
infinitesimal scale-space generator
  definition, 109
infinitesimal smoothing, 80
inner scale
  of image, 182
  interference, 182
  of object, 9
Index

of observation, 15
integrated Gaussian
derivative, 145
integrated Gaussian kernel, 97–98
integration scale
definition, 359–360
selection, 366
inverse problem, 19
iso-intensity linking, 309–310
isotropic pattern, 362
isotropy, 14, 47
junction classification, 285–300
basic scheme, 287–289
context information, 289
experiments, 295–297
scale problem, 289
junction detection, 158, 287, 291–293,
328–339
composed scheme, 337–339
junction types, 288
Laplacian operator, 169
discrete correspondences, 104
Laplacian pyramid, 37
large support diffusion smoothing, 350
large support smoothing, 148
level curve curvature, 159, 160, see
curvature of level curve
level curve interpolation, 154–155
level surface, 47
non-creation, 47
lifetime
in scale-space, 173, 174, 177
linear illumination gradient, 55
linear increase of receptive field size,
54
linearity, 40, 49, 63, 84, 90, 106
linking across scales, 18, 172, 177
iso-intensity vs. feature points,
309
local extrema, 159
non-creation, 47, 48, 63
local scale
definition, 359–360
selection, 366
locality, 106
localization
blob boundary improvement, 276
conflict with detection, 272
low-pass pyramid, 35
Markov process, 43
matching
blob-blob, 233, 293
blob-edge, 273, 276
blob-extremum, 234
proximity, 276
Voronoi, 274, 276
mathematical morphology, 59
maximum
region-based, 228
semi-weak, 228
strong, 228
weak, 104, 228
maximum principle, 44, 104, 119
non-linear diffusion, 384
merge
of blobs, see blob merge
Miller’s algorithm, 396
minimum, see maximum
minor, 76
mirror symmetry, 106
modified Bessel function, 85
generating function, 86
recurrence relation, 95
monotonic intensity transformation,
151
Morse function, 167, 195
multi-grid methods, 38
multi-index notation, 45
multi-scale representation, 9–10, 62
curve, 59
linear, 31–60
multiple blob responses, 258, 284, 313

N-jet representation, 51
algebraic properties, 125
basic filters, 139
computation of, 135
nearest-neighbour processing, 161
neighbourhood, 104, 119
nested hierarchy
of level curves, 232, 311
neurophysiology, 3, 54
nine-point operator, 118
noise sensitivity, 177
non-creation
of extrema, 63
of level curves, 103
of level surfaces, 47
of local extrema, 47, 48
of structure, 13
of zero-crossings, 47, 48, 63
non-enhancement
of local extrema, 44, 106, 128
non-enhancement of local extrema, 46
non-generic signal, 63
discrete definition, 228
non-linear diffusion, 59, 383–385
non-maximum suppression, 153, 273
non-uniform Gaussian kernel, 387
Fourier transform, 390
interpretation, 389
transformation property, 388
non-uniform smoothing, 383–393
normal distribution, 53
normalization, 48, 84, 90, 106
coordinates in scale-space, 250
discrete analogue of Gaussian, 136
discrete derivative approximations, 136
pyramid, 35
normalized anisotropy, 355
behaviour over scales, 366
normalized coordinate, 46, 319
normalized derivative, 46, 319
discrete implementation, 345–346
interpretation, 345–346
normalized differential entity
scaling property, 321
normalized scale-space extremum
definition, 324
invariance property, 391
scaling property, 325
numerical analysis, 19

object detection, 306
operator size, 7
organization of this book, 22
outer scale
of image, 182
interference, 182
of object, 9
oversampled pyramid, 38
parabolic differential equation, 53
parameter variation, see transformational invariance
peak detection, 282
perceptual organization, 6, see grouping
perspective projection
basic effects, 351–352
camera geometry, 352–353
Pi-theorem, 50
Plancherel’s relation, 346
point measurement, 41
pointwise non-linear combinations, 148, 350
Polya frequency function, 57
classification, 88
definition, 88
integrated, 97
sampled, 92
semi-group, 90
Polya frequency sequence, 57
classification, 79
definition, 78
normalized, 78
semi-group, 85
polygon approximation, 59
position effect, 351
positivity, 99, 106
in Fourier domain, 68–74
in spatial domain, 65–66
pyramid, 35
pre-scale-space family
derivative approximation kernels, 129
pre-scale-space property, 108
pre-scale-space representation, 107
derivative approximation, 130
differentiability, 107–108
primal sketch, 18, 307
primitive smoothing transformations
classification, 80
processing cone, 32
proximity matching, 274
pyramid, 33–38, 82
derivative approximation, 135
fixed scale sampling, 122
generation, 34, 53
oversampled, 38
properties, 38
quad-tree, 32–33
Index

reaction-diffusion equation, 59
reader’s guide, 23–28
receptive field, 54
recognition cone, 32
reconstruction
from bandpass pyramid, 38, 55
from zero-crossings, 60
of world, 19
recursive filtering, 80, 100
Reduce, 35
refinement, 83
continuous scale, 82
limit on depth, 239
local, 243
scale, 239
region of interest, 17, 289, 307
generation, 256
importance of, 271–272
regularization, 21, 57
and diffusion, 383
property of scale-space, 45
regularizing
property of scale-space, 57
relational tree, 170, 171, 232
relative error, 397
relative integration scale, 369
relative invariant, 152
reordering, 173, 174
repeated averaging, 61, 67
limit case, 97
representation, 2
resolution, 20, 32
retina, 54
review
multi-scale representations, 31–60
non-linear diffusion, 383–384
shape-from-texture, 350–353
total positivity, 76–79
rotational invariance
differential singularities, 153
scale-space primal sketch, 311
rotational symmetry, 52, 106
discrete scale-space
derivative approximation, 135
scale-space axiom, 49
saddle path
definition, 192
shared and non-shared, 201
sampled Gaussian
derivative, 137, 145
kernel, 137
sampled Gaussian kernel, 91–94
scale, 8–9, 20
in early vision, 16–17
scale invariance, 49
necessity from, 49–52
scale parameter, 10, 62
continuous, 84
scale problem
in early vision, 271
scale selection, 16, 317–348
assumption, 249
basic problem, 271, 285
blob detection, 325–327
guiding processes, 256
heuristic principle, 320–325
junction detection, 296, 328–332
junction localization, 332–337
parameter variations, 251
properties, 257
shape-from-texture, 360–363
scale-space, 10–15, 39–46
affine, 387–390
arbitrary dimensions, 39
behaviour of structures, 187–226
continuous signals, 39, 88–91
derivative approximation, 123–148
discrete, *see* discrete scale-space
discrete scale parameter, 82
implementation of, 395–397
infinite differentiability, 125
numerical approximations, 91–98
separability, 43
special properties, 40–53
uniqueness, 47–53
scale-space axioms, 47–52, 103–104
cascade smoothing property, 129
continuity, 107, 130
linear shift-invariant smoothing, 129
linearity, 90
list of, 106, 130
normalization, 90
semi-group, 90, 106
shift-invariance, 90
symmetry, 130
scale-space blob, 172–173
definition, 193
representative scale, 249
spatial representative, 250
scale-space blob tree, 173
scale-space blob volume
computation of, 242–243
definition, 194
normalization, 312–313
normalized, 194
significance measure, 177, 249
scale-space derivative, 45, 124–127
scale-space kernel, 63
scale-space lifetime, 173, 177
measurement, 178–182
significance measure, 177, 250
scale-space primal sketch, 16, 165–185
algorithm, 227–246
computation, 252
data structure, 185, 244–246
detecting image structures, 249–270
invariance properties, 311–312
motivation, 165
summary, 307–314
scale-space property, 132
derivative approximations, 127
scale-space representation, 108
scaling effect, 351
scaling invariance
differential singularities, 153
scale-space primal sketch, 311
scaling property
of scale-space, 44
Schrödinger equation, 52
Schwartz distribution theory, 124
second moment matrix, 40
composed method, 365
definition, 354
eigenvalues and eigenvectors, 355
interpretation, 354–356
junction detection, 333
linear transformations, 356–357
multi-scale, 359
non-uniform smoothing, 391
transformation property, 391
visualization, 358
segmentation, 16–17, 250, 256
basic problem, 271
histogram-based, 282
selective mechanism, 162
self-similarity, 49
semi-group, 41, 84, 90, 126
canonical, 49
discrete scale-space, 106
infinitesimal generator, 107
property of Gaussian, 41
scale-space axiom, 48, 49
violation, 92
separability, 43, 52, 114–116
Gaussian kernel, 43
scale-space axiom, 49
separable discrete scale-space
derivative approximation, 134
shape adapted smoothing
shape-from-texture, 385–387, 392–393
shape distortion, 54
junction detection, 293
non-uniform smoothing, 383, 385
shape-from-texture, 363–364
shape-from-texture, 304, 349–382
affine invariance, 392–393
composed method, 374
experiments, 368–371, 376–379
problem, 350
shape adapted smoothing, 385–387, 392–393
shift-invariance, 40, 49, 63, 84, 90
sign change, 76
sign-regularity, 77
signature
blob detection, 323
detection, 341
detection, 343
detection, 329
detection, 335
scale-space, 323
shape-from-texture, 366
significance
measure, 174, 177–184
ranking, 249, 250
singularity
Index

in scale-space, 197–200
singularity detection, 161
singularity set, 152, 196
in one-parameter family, 197
skew invariance, 159
slant, 353
small support derivative computations, 148, 350
smoothing, 20, 64
smoothing filter, 54
smoothing examples, 39
spatial extent
significance measure, 177
spectral density, 218
self-similar, 219
split
of blob, see blob split
steerable filters, 139
strong continuity, 108
structure, 47, 49, 63, 250
sub-pixel edge detection, 154–155
subsampling, 32, 34
support region
of grey-level blob, 167
definition, 168
of scale-space blob, 194
symmetry
pyramid, 35
texel grouping scale, 375
texture analysis, 304, see shape-from-texture
texture gradient, 350, 351, 353
Thom’s classification theorem, 195
three-kernels
scale-space properties, 74–75
three-point operator, 105
tilt direction
definition, 353
Toeplitz matrix, 69–71, 78
top point, 60, 188
total positivity, 76, 78, 88
transformational invariance, 251, 307–308
junction detection, 286
translational invariance, 106
differential singularities, 153
scale-space primal sketch, 311
truncation

error in approximation, 8
of Gaussian kernel, 395
two-stage scale selection, 337
uncertainty relation, 52, 53
degree detection, 272
uncommitted visual system, 54, 385
unidirectional pattern, 362
uniform rescaling, 150, 153, 311, 388
unimodality, 36, 65, 82, 99, 106
Fourier transform 2-D, 114
in Fourier domain, 68–74
in spatial domain, 65–66
pyramid, 35
uniqueness
of Gaussian, 14–15, 47–53
variance, 87
variation-diminishing, 76–78, 88, 89
vision, 1–3
visual cortex, 54
visual front-end, 3, 4, 31, 54, 106, 130, 148
Voronoi diagram
for matching, 274
wavelet representation, 55, 58, 60
weak isotropy, 357–358
weighted average, 41
well-posed problem, 6, 56
white noise, 182, 219
zero-crossing path, 47
zero-crossings
non-creation, 47, 48, 63
of Laplacian, 159
drift velocity, 217
degree localization, 161
information content, 59
non-creation, 103
related to extrema, 212–213
related to grey-level blobs, 169, 311