Contents

List of Tables

List of Figures

List of Symbols

∇	“nabla”, the gradient operator
δ	”delta”, the Laplacian operator
$\emptyset$	“null” symbol, the empty set
$\cap$	“cap”, set intersection
$\cup$	“cup”, set union
$\mathbb{R}$	the set of real numbers
$\mathbb{I}$	the set of integers
$\mathbb{D}$	the distance transform in the image space
$\mathbb{C}$	the set of complex numbers
$\mapsto$	“maps to”
$\subseteq$	subset
$\subset$	proper subset
$\supseteq$	superset
$\supset$	proper superset
$\otimes$	convolution
∂	partial derivative
$\overset{\text{def}}{=}$	defined as
$\vec{\ }$	used with other symbols indicates a multi-valued or vector-valued quantity
$\mathcal{D}$	“caligraphic D”, the dissimilarity measure
$\mathcal{M}$	“caligraphic M”, a metric
$\mathcal{F}$	“caligraphic F”, the Fourier transform or the set of functions of polynomial invariants
$\mathcal{N}$	“caligraphic N”, a neighborhood
$\mathcal{P}$	“caligraphic P”, the set of polynomial invariants
$\mathcal{G}$	“caligraphic G”, a low-level computer vision algorithm or process
$\sum$	sum

$\int$	integral
$\in$	“element of”
$\notin$	“not an element of”
$\forall$	“for all”
$\exists$	“there exists”
$\Rightarrow$	“implies”
$\Leftrightarrow$	“implies and is implied by”
$\sqrt{}$	square root
$\cdot$	“dot”, the dot product of two quantities or a set of unspecified parameters in a function
∞	infinity

Chapter 1

Introduction

1.1 Segmentation and vision

Dividing an image into meaningful parts—a process called *image segmentation*—serves a number of purposes. First, it offers a means of consolidating the vast array of information that is typically present in a visual field. It can help to distinguish areas of importance from those of less importance, and it thereby provides a means of focusing attention. Second, segmentation gives a level of structure for subsequent analysis that reflects not only basic spatial relationships but the deeper structure of the image. Segmentations are also interesting for the boundaries that they generate. For instance, dividing the image into pieces that correspond to projections of three-dimensional objects in a scene provides contours of those objects. Those contours offer a great deal of insight into the shapes of those three-dimensional objects.

In computer vision a number of very impressive algorithms rely on an underlying segmentation. The analysis of shape, for instance, often assumes that boundaries of objects are known (Kimia and Zucker 1992, Brechbühler, Gerig and Kübler 1992, Blum and Nagel 1978, Zahn and Roskies 1972). "Edges" are assumed in a variety of algorithms for stereopsis, pattern recognition, and shape from shading. In real life, however, images are rarely presented as sets of boundaries and finding the "right" boundaries can be difficult.

Segmentations are of great practical importance. In medical imaging, segmentations are useful in the interactive visualization of MRI and CT data (Yoo, Neumann, Fuchs, Pizer, Cullip, Rhoades and Whitaker 1992). Models of anatomy can be constructed from patient data once the organs have been accurately distinguished from one another and the surrounding tissue (Gerig, Martin, Kikinis, Kubler, Shenton and Jolesz 1991, Kikinis, Cline, Altobelli, Halle, Lorensen and Jolesz 1992). Image registration is performed based on key points, boundaries, and possibly patches. Objective measurements of the size or volume of anatomical structures can be made once the appropriate tissue has been identified.

Segmentation plays a plays a pivotal role in natural vision as well. The human visual system can group parts of an image together based on such basic attributes as luminance (Marr 1982), color, texture (Beck 1972, Beck 1982), disparity, and motion. While some models of early vision place segmentation as a higher level process that is performed *after* the detection of basic attributes, other evidence suggests these levels are not so distinct (Grossberg and Mingolla 1985). Ramachandran (1992) explains:

> My first point concerns the role of *image segmentation* in perception. According to David Marr's (1982) principle of modular design, certain elementary functions such as stereopsis, motion, color, etc. are mediated relatively early in visual processing by specialized modules, whereas segmentation of the visual scene into separate objects is assumed to be

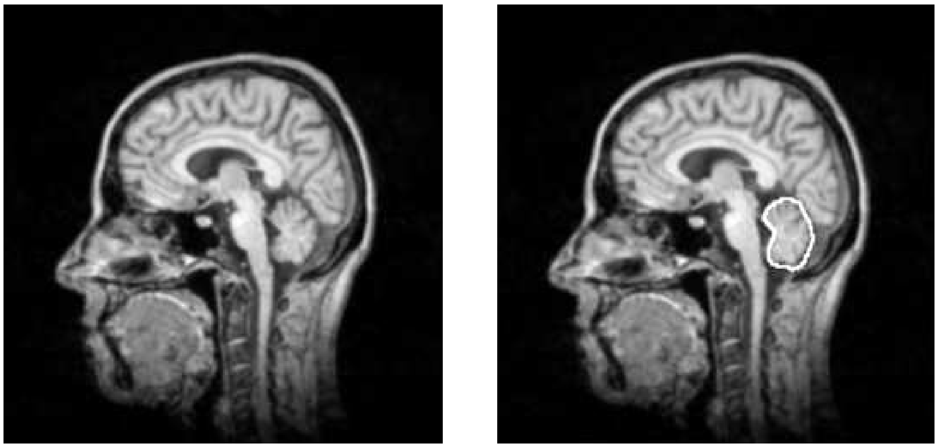


Figure 1.1: A slice of MRI image of a head shows that segmentations which are "easy" for our visual systems can be difficult for computer algorithms. The white outline in (b) identifies the cerebellum. Despite the ease with which the (untrained) human identifies this region as a distinct object, there appears to be no simple, consistent rule for deciding which pixels lie inside the object and which lie outside.

> a more complex process that can actually use the output of these early vision modules. Since the modules perform their functions prior to image segmentation, the argument goes, one can successfully model them and study them experimentally without worrying about segmentation. Contrary to this view our evidence suggests that image segmentation can influence a number of early visual processes such as stereopsis, structure from motion, motion correspondence, and shape-from-shading. The implication is that early vision modules are not autonomous—they interact significantly with each other and with segmentation.

1.2 The problem

The design of strategies or algorithms for segmenting images is extremely difficult. The challenge is two-fold. First, a large number of visual tasks are objectively difficult. This is because images are noisy and unpredictable, and because digital images contain a great deal of information, much of which irrelevant. Second, the human visual system, which often serves as benchmark for artificial vision, is very effective and not well understood.

The example in Fig. 1.1a shows a sagittal slice of an MRI image that was measured from a human head. Figure 1.1b shows a white curve that was drawn by hand in order to enclose a particular structure within the head image. This structure happens to be the cerebellum, but one needs to know nothing about anatomy in order to realize that this is a distinct piece of the image. What algorithm could reach such a conclusion? The intensity values associated with brightness in this picture range from about 350 to 900 within the cerebellum and from 200 to 500 in the area surrounding it. The same organ is easily identifiable across any number of such images from different patients of different shapes and sizes. How does one decide in each case that a pixel with value of 350 lies inside the organ, while another pixel with a value of 500 lies outside?

Among the various images that challenge computer vision, the MRI image of Fig. 1.1 is probably one of the easier ones. The contrast is high, and there is little noise. A useful general purpose segmentation algorithm would have to face problems that are significantly more difficult than this one.

1.3 Strategy

This dissertation proposes a segmentation strategy which relies on the local homogeneity of image structure. Segments are regions that exhibit local homogeneity (little variation) in image structure, while the boundaries of segments are places of local inhomogeneity (great variation). Visual features or structures are places of inhomogeneity, and thus the problems of structure extraction and segmentation are linked; visual structures are the boundaries of segments.

Image structure is quantified via a set of measurements. Image measurements are *physical observables* and therefore have an associated *aperture* or *scale*. This scale takes the form of a local neighborhood operator that performs a local averaging at every point in the image. The notion of scale introduces a free parameter into all image measurements; all measurable image structures necessarily have an associated scale (or scales).

This dissertation explores an *anisotropic* averaging process which is sensitive to structures that are present in the image. This is a nonlinear diffusion process which incorporates a variable-conductance term. The structures that shape this process have an associated scale, and this scale changes as the process evolves. The result is a process that combines information at a range of scales and helps to reduce small scale noise while retaining accurate information about the shape of large image segments.

The generalization of these principles to incorporate multi-valued, geometric descriptions of images provides a great deal of flexibility in defining the "structures" that control the diffusion process. "Patches" or segments are characterized by higher-order geometry, and the boundaries of these patches provide a means of detecting geometric structures which have visual importance.

1.4 Thesis and contributions

This dissertation presents a framework for constructing processes or algorithms which aid in image segmentation. I call this class of processes *geometry-limited diffusion*. My thesis is that *the geometry-limited diffusion framework that is described in Chapts. 3 and 4 offers a reliable and accurate means of segmenting images and characterizing visual structures. This method addresses many of the challenges of image segmentation and structure detection and promises to be a powerful tool in artificial vision.*

The task of making *accurate* measurements is not simply a matter of geometric precision. Later chapters will show that the one of the primary challenges facing low level computer vision algorithms is noise, or irrelevant luminance fluctuations. Methods for making accurate measurements of image structure must be not only geometrically correct and procise but also statistically reliable; they must be insensitive to noise. The thesis asserts that the diffusion processes described in this dissertation exhibit this reliability. The thesis also asserts the usefulness of a *framework*, as opposed to a specific technique or algorithm. In order for the framework to be useful, it must provide a strategy, or set of rules, for constructing good algorithms. The framework that I propose, geometry-limited diffusion, derives from a principled, coherent strategy and, when applied to specific examples, produces algorithms which perform well.

This work makes several contributions to the fields of image processing and computer vision. First, I propose a segmentation strategy that is based on the local homogeneity of image structure, and I suggest that interesting visual structures can be described as places that exhibit local inhomogeneity in geometric structure. Thus, structures can be detected by segmenting images and characterizing the boundaries of those segments. Different types of visual structures are detected by using different representations of local image structure. Second, I modify the variable-conductance diffusion processes proposed by previous authors to include a notion of scale. The result is a multiscale diffusion process that is mathematically well posed and appears to provide more reliable results. Third, I generalize these diffusion equations so that they can apply to multi-valued images, and I develop the notion of *dissimilarity*, a generalized gradient magnitude operator, which controls the conductance in this multi-valued setting. Fourth, I show how this framework applies to geometric descriptions, in particular sets of image derivatives. I develop several examples and show how a first-order diffusion process can be used to detect ridges and valleys while a second-order diffusion process can produce middles as well as edges of objects. Fifth, I develop an experimental method for evaluating low-level computer vision algorithms that relies on Monte Carlo simulations. I use this method to analyze the behavior of several different diffusion processes. I quantify (among other things) the relative advantage of the anisotropic and multiscale diffusion process over Gaussian blurring. Finally, I show that these diffusion processes can be tuned to capture frequency information and that systems of frequency-tuned diffusion equations can be used to segment images on the basis of texture.

1.5 Overview

This dissertation offers several types of analysis in support of the thesis. First, it introduces the concepts of image geometry and scale and then presents *anisotropic diffusion* as a natural extension of isotropic scale space. It offers a single basic strategy for segmentation that relies on the homogeneity of local image structure, and it develops geometry-limited diffusion from that strategy. Chapter 2 gives the background to this work, which includes image geometry, scale space, and anisotropic diffusion. Chapter 3 presents a form of anisotropic diffusion that incorporates information at multiple scales. This multiscale process produces more reliable results in the presence of correlated and uncorrelated noise.A number of examples are shown and discussed, and a mathematical analysis is used to describe the behavior of this process. Chapter 4 describes a generalization of anisotropic diffusion to multi-valued images. This framework combined with finite representations of local image structure results in geometry-limited diffusion. Examples for first and second-order shape representations are discussed, and a proof of the geometric invariance of this system is given. Chapter 5 gives a systematic empirical analysis which provides further insight into the behavior and limitations of this approach and shows that the geometry-limited diffusion compares favorably with Gaussian blurring. Chapter 6 demonstrates the flexibility of the general framework by showing how it can apply to other diverse segmentation approaches including texture and supervised classification. Some details concerning the implementation of this method are discussed in the Appendix.

Chapter 2

Background

2.1 Image segmentation

An image f is a smooth mapping from the image domain $I \subset \mathbb{R}^n$ to some subset of the set of real numbers, $L \subset \mathbb{R}$, so that $f : I \mapsto L$, where the range L is called the *intensity* or sometimes *luminance*. The graph of this function is the *intensity surface*. By this definition images are assumed to be "greylevel" like those obtained from a black and white camera or from a single channel of a color camera. (Generalizations to multi-valued images are discussed in Sect. 4.1.)

For the purposes of this work, a digital image is a discrete sampling of such a continuous function. Digital images are assumed to have finite resolution in the domain I. Figure 2.1 shows a shaded three-dimensional rendering of the intensity surface associated with the head image of Fig. 1.1. This surface has a special property; it is the graph of a function. There are three axes, x, y, and L, but the intensity axis L is distinct and assumed to be incommensurate with the spatial axes. There is an "up" associated with this surface, and the "height" at a particular point on this surface is its distance from the plane defined by $L = 0$.

Loosely defined, an image segmentation divides the image into parts or pieces. The exact specification of these pieces might depend on the particular application or task. As a formal definition that will serve the purposes of this discussion, I propose the following:

Definition 1 *An* image segmentation *is a partitioning of the image space I into a finite number of nonintersecting contiguous open subsets $S_1, S_2, \ldots \subset I$, such that the closure of the union of these subsets is the entire image space. That is, $S' = S_1 \cup S_2 \cup \ldots$ and $\widehat{S'} = I$, where $\widehat{S'}$ is the closure of S'. The complement of this union, $I - S'$, is the set of* boundaries *associated with that segmentation.*

2.2 Discrete versus fuzzy

Definition 1 describes a "discrete" segmentation; it identifies each point in the image according to its membership in a particular patch or segment. A discrete segmentation $\mathcal{S}$ is essentially a mapping from the image space to one a set of discrete values which identify the segments, $\mathcal{S} : I \mapsto V \cup B$, where V is a set with finite members that consist of one label for each segment, and B is a special label which indicates a boundary.

Alternatively one could consider *fuzzy* segmentations. Segments would no longer be conventional sets but fuzzy sets (Foster 1992, Ferraro and Foster 1992) which are characterized by graded membership functions. There are number of possibilities for describing such a segmentation. One is $\mathcal{S} : I \mapsto P^m$, where $P \subset [0, 1]$ is *degree* of membership and m is the number of individual segments.

The relative merits of these two approaches warrant discussion. While easily defined in the

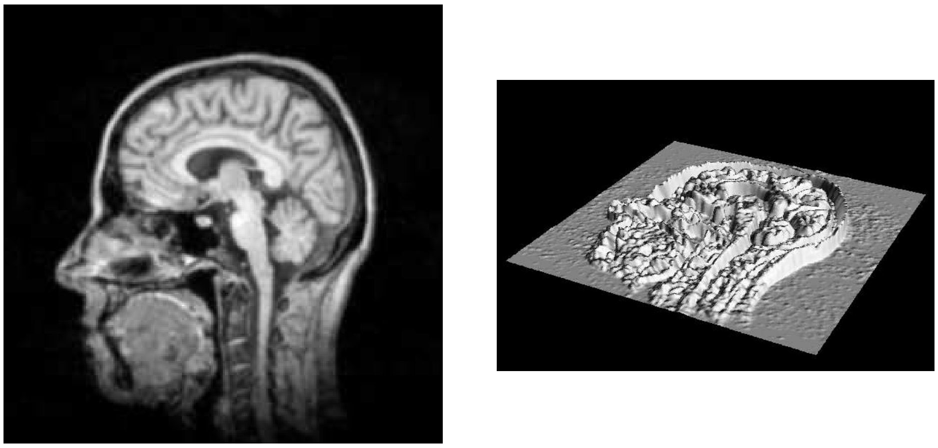


Figure 2.1: The MRI head image has an associated intensity surface.

abstract sense, fuzzy segmentations pose some practical difficulties. First, one of the motivations for segmenting images was to consolidate information, but the fuzzy segmentation as described here promises to do just the opposite. At each pixel there is an entire array of values, and the length of this array is proportional to the number of individual segments in a image— potentially a large amount data. Second, it is not yet clear from the literature what to do with these fuzzy segmentations. For instance, the total volume of a segment might be computed by integrating its associated membership function. But what does it mean to integrate fractional membership functions of a concrete physical object—like a car? What is the geometry of a fuzzy curve? Pizer and others (Pizer, Coggins and Burbeck 1991, Pizer, Burbeck, Coggins, Fritsch and Morse 1992, Pizer, Burbeck and Morse 1992, Morse 1992) have formulated a theory of shape that incorporates greylevel functions, but the practical implications of this work are under study.

The concept of discrete segmentation also has some serious shortcomings. While there are a number algorithms that rely on discrete segments or the precise locations of boundaries, it is not clear that boundaries warrant such precision. We shall see in later sections that vision, by its nature as a physical process, requires that all measurements entail a finite aperture. Such measurements should always include tolerances, but tolerances are rarely computed and are almost never included in subsequent analyses. *A majority of algorithms fail to recognize the inherent uncertainty associated with structures in images.*

A reasonable test of the legitimacy of any segmentation strategy is its usefulness in solving visual problems. In this light the following questions are relevant. What are the tolerances associated with boundaries or segments? How useful are tolerances in solving visual problems? Are decisions made on the basis of fuzzy segmentations "better" than those that result from discrete segmentations?

This discussion is a terse outline of what promises to be a deep and ongoing debate in the field of computer vision. I do not attempt to resolve the issue of discrete versus fuzzy segmentations in this dissertation. The methods presented here appear to be useful for both discrete and fuzzy segmentations. Most of the results and discussion in the dissertation assume the a discrete segmentation is the goal. However, the concept of tolerances in computing boundaries is addressed more extensively in the empirical analysis of Chapt. 5.

2.3 Example: edges

Given the task of dividing a discretely sampled image into a finite number of meaningful regions, there is a wide range of possible strategies. One reasonable approach is to group nearby pixels on the basis of similarity in intensity values. In medical imaging, for instance, different tissue types can sometimes be characterized by their response to specific imaging processes. In "real world" images like those that might be obtained from a video camera or projected onto a retina, luminance can indicate the normal direction (in the three dimensional scene) of a smooth surface relative to a light source as well as the physical properties of reflecting surfaces. The boundaries associated with this type of segmentation are often called "edges," and much of the computer vision literature focuses on these visual structures.

In this approach boundaries between regions can be identified as areas where the luminance f at each pixel is different from that of pixels nearby. One measure of this difference is the gradient vector. For a two dimensional continuous image $f(x, y)$ the gradient vector is

$$\nabla f = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) f = \left(\frac{\partial f}{\partial x} \ \frac{\partial f}{\partial y}\right) = (f_x \ \ f_y), \tag{2.1}$$

where I use the lowered indices x and y to denote derivatives with respect to the x and y coordinates, respectively. The gradient magnitude is

$$|\nabla f| = \left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2\right)^{\frac{1}{2}} = ({f_x}^2 + {f_y}^2)^{\frac{1}{2}}. \tag{2.2}$$

In a digital image that is sampled on rectilinear grid, the derivatives are typically approximated by finite differences of nearest neighbor pixels. The gradient magnitude associated with a pixel at a point x_n, y_m is

$$|\nabla f(x_n, y_m)| = \left(\left(\frac{f(x_{n+1}, y) - f(x_{n-1}, y)}{2}\right)^2 + \left(\frac{f(x, y_{m+1}) - f(x, y_{m-1})}{2}\right)^2\right)^{\frac{1}{2}}. \tag{2.3}$$

The gradient is a vector, and it points in the direction of maximum increase in intensity; it points in the "uphill" direction. The magnitude of this vector is the rate of change of intensity associated with a small movement in the direction of maximal change. Using the gradient, there are a number of options for identifying divisions or boundaries between regions. Perhaps the simplest is to threshold the gradient magnitude. Pixels that coincide with gradients higher than a specified threshold serve as boundaries to separate other regions.

This segmentation approach suffers from two basic problems:

- A number of interesting image segmentations are not easily described in terms of differences in luminance. The human visual system, for example, is capable of distinguishing patches that have the same overall luminance but differ in some other properties such as texture or depth (Ramachandran 1992, Beck 1966, Beck 1982).
- The gradient magnitude is not a reliable metric when measured in the presence of unwanted, small scale, luminance fluctuations. The gradient, as well as all other differential measurements made on images, is sensitive to noise.

Both of these problems are examined in the sections that follow. The first problem is addressed by including the higher-order geometric structure of the intensity surface. This leads to a wide range of possibilities for segmentation and a general framework for constructing image segmentations. The

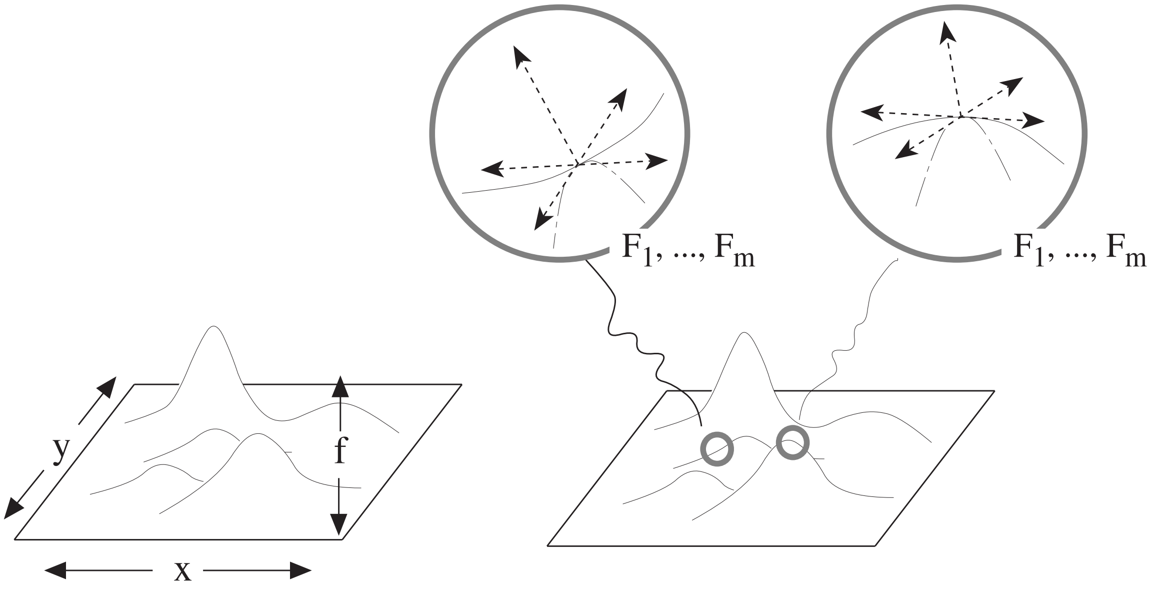


Figure 2.2: The intensity surface is the graph of an image. A finite set of measurements at every point in the image space represent the local shape of the intensity surface. The cutaways illustrate that each place on the surface has a local surface patch and a set of descriptors which describe that patch (to within some approximation).

second problem requires the introduction of a low-pass filtering process *in combination with* the differential measurements, which leads to the concept of scale-space.

2.4 Image geometry

2.4.1 Local shape

Each point in the image has a corresponding point on the intensity surface, and each point on that surface has a local structure which is defined over some neighborhood. This is the *local shape* of an image. I am purposely vague here because I want to allow for a wide range of possibilities for characterizing this structure.

For example, Koenderink (1984) describes the "deep structure" of images and argues that local image structure is captured by the differential structure of f. Thus, one option is to capture the local shape at each point in the image by measuring the derivatives of the intensity surface up to order N. These derivatives describe the local Taylor series approximation at each point on the intensity surface. The strategy of encoding local surface structure as sets of derivatives is adopted throughout most of this dissertation, but there are other possibilities. In Chapt. 6, for example, I explore an alternative representation of local shape that resembles a local frequency or Fourier analysis.

Generally, useful descriptions of local shape can be represented or "encoded" as a finite set of scalar values. At each point in the image there is a set of values $F_1, \ldots, F_m$ that results from measurements of the local image structure (Fig. 2.2). The inserts depict the differences in local image structure. The arrows represent the scribbling of a hypothetical geometer. This geometer makes a set of measurements at each point in the image (this is not a "real time" algorithm). He records these measurements as a list of numbers, $F_1, \ldots, F_m$. This collection of measurements is a *geometric description* of the image.

The specification of contiguous regions in the image domain such that the regions depend on the local shape properties or the geometric description of the intensity surface is a *geometry-based image segmentation*. An example of this kind of segmentation is that of Sect. 2.3, which entails partitioning an image into regions that are based on slowly varying luminance, where the resulting boundaries form luminance "edges." These boundaries are detected as areas of high gradient magnitude.

Another example is the segmentation of images using extrema in height and the flanks or hillsides associated with these "peaks" and "pits." Such segmentations have been created on the basis of gradient directions and watersheds (Serra 1982, Rosin, Colchester and Hawkes 1991). Second-order properties (properties that involve second-order derivatives), such as the mean curvature of the intensity surface or the second derivative in the gradient direction, can also serve as the basis for such segmentations.

2.4.2 Local homogeneity as a segmentation strategy

In this dissertation I consider segmentations that incorporate the following principle:

Principle 1 *Segmentations group image neighborhoods into regions based on the local homogeneity or similarity of the* local shape *across those neighborhoods.*

In the example of Sect. 2.3 local shape is implicitly reduced to the height of this intensity surface—a zeroth-order property. In that case measuring the similarity among points consists of measuring the intensity at various places in the image and then comparing these values in local neighborhoods. For higher-order properties of local shape the notion of similarity between the shape of different surface patches is ambiguous because there are many features that can be compared.

Two important decisions must be made before one constructs a geometry-based image segmentation. The first decision is the particular *geometric description* that is to be measured and compared. The space of possible surface shapes is infinite, yet we wish to compare surface shapes in some computable, i.e., finite, manner. Thus, representations of local surface structure must be finite; comparisons are made between approximating functions to the intensity surface at each point in the image. A shape-based segmentation algorithm requires a specification of this finite approximation. It is not enough to define a general strategy for representing surface shape, such as that of approximating the local structure by a set of image derivatives. In practice one must decide exactly which are used to represent local shape properties, e.g., derivatives of what order.

The second decision is the nature of the comparison, or the *metric*, that is used in determining similarity. Such a metric quantifies the differences in shape between two surface patches. For instance, one might choose to describe the local shape of the surface using second-order properties. Measurements that capture second-order information are the three mixed partial derivatives of order two:

$$\begin{aligned} F_1 &= f_{xx}, \\ F_2 &= f_{xy} = f_{yx}, \quad \text{and} \\ F_3 &= f_{yy}. \end{aligned}$$

Given this shape decomposition, it is necessary to quantify the *degree of difference* in two quadratic functions. There are a variety of possibilities (Koenderink 1990).

Regions are places in an image where variations in geometric descriptions are small within local neighborhoods, and boundaries are the loci of places where variations in geometric descriptions are large. With this definition, it can be difficult to decide the amount of variation that constitutes a boundary. How large is *large*? I do not address this question directly but point out that there are several widely used methods for making this decision in a discrete fashion, if indeed a discrete segmentation is the goal. One option is to quantify the variation in local shape and then threshold

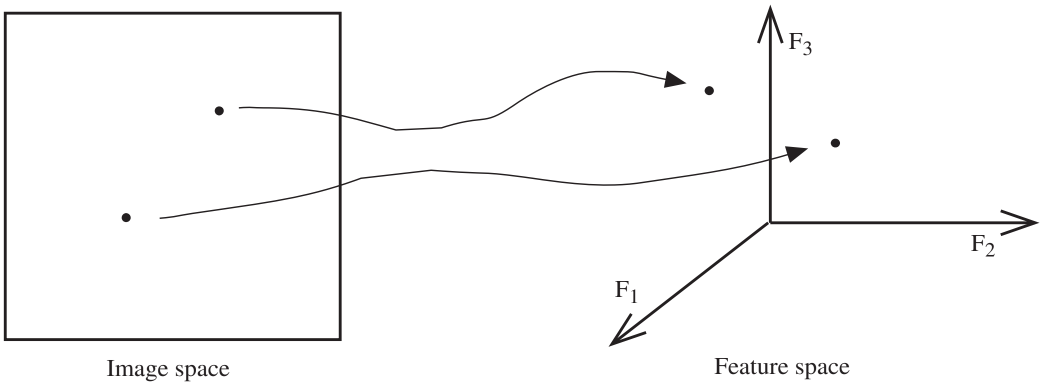


Figure 2.3: A set of geometric features, F_1, F_2, F_3, associate each point in the image with a position in a geometric feature space. Each position in the space represents a different surface shape, and the distance (not necessarily Euclidean) between points in this space provides a metric for quantifying the differences in local image structure. This metric will provide a means of computing the local inhomogeneity of image structure.

this measure. Points where geometry changes at rates above the threshold are defined as boundaries. Another option is to specify boundaries as critical points in the rate of change of local geometry, as does the Canny (1983, 1987) edge operator.

2.4.3 Patch boundaries and visual structures

In order to quantify differences in local shape, I choose to represent shape as a position in a finite-dimensional *feature space.* In statistical pattern recognition (Duda and Hart 1973) objects can be classified based on a set of measurements, called a *pattern* or *feature vector.* A set of features has an associated feature space, which is the space consisting of all possible feature vectors. Thus, the features span the feature space. Typically, classifications are made on the basis of proximity in the feature space. This approach can be used to classify points (or pixels) in images by assigning a number of values or *features* at each point in the image and using the structure of that feature space to group or classify points in the image. There is a feature vector assigned to each point in the image space, so the features form a mapping from the image space to the feature space as in Fig. 2.3.

For geometry-based image segmentation the geometric description determines this feature space. The similarity of shape between two points in the image is inversely related to the distance between the two corresponding points in this feature space. *Dissimilarity* is defined as the amount of variation between geometric descriptions within a neighborhood of a point in I. Thus, boundaries are regions of high dissimilarity.

A number of visual structures can be expressed as the boundaries between geometric patches. This is true of luminance "edges" described above. Ridges or *creases* [1] can be expressed as the boundaries of geometric patches that contain similar first-order information. Laplacian zero crossings (also considered a type of "edge") are the boundaries of regions that are grouped on the basis of the sign of mean curvature. There is duality between visual structures such as creases and edges, and the geometric properties of the patches that they enclose. Segmentation is a method for "detecting" or "extracting" the structures that correspond to the boundaries of such patches. The appropriate geometry and dissimilarity measures should depend on the kind of visual structures one wishes to characterize.

---

[1] A crease is a ridge or a valley.

2.5 Scale-space

2.5.1 Gaussian kernels

Differential measurements of digital images can be obtained through some neighborhood operator or kernel. That is, derivatives of digital images are computed by a sampling process that includes some weighted averaging over a local neighborhood. This averaging is done instead of the infinitesimal limits associated with conventional derivatives of continuous functions—such limits do not exist on the discrete grid. Koenderink (1984) and others (Koenderink and Doorn 1990, Witkin and Tenenbaum 1983) argue, based on mathematical principles, that the sampling function associated with image measurements should be the Gaussian. Gaussian kernels have also been shown to be excellent models of some receptive fields in natural vision (Young 1986, Young 1987, Young 1990). In the continuous case the Gaussian kernel, normalized to have a unit integral over an n-dimensional image space, is

$$G(\vec{x}, s) = \frac{1}{(2\pi s^2)^{\frac{n}{2}}} \exp\left(\frac{-\vec{x}\cdot\vec{x}}{2s^2}\right). \tag{2.4}$$

Neighborhood operators such as the Gaussian introduce a free parameter which controls the size of the neighborhood. The size of the sampling function controls the size of the structures that can measured. For the Gaussian the continuous scale parameter s allows one to create a dense set of blurred versions of the original image, called a “scale space.” Koenderink argues that different kinds of measurements require kernels of different sizes or scales.

Derivatives of images can be made using a convolution. For two functions $f(\vec{x})$ and $g(\vec{x})$,

$$g(\vec{x}) \otimes \frac{\partial}{\partial x_i} f(\vec{x}) = \frac{\partial}{\partial x_i} g(\vec{x}) \otimes f(\vec{x}), \tag{2.5}$$

where “$\otimes$” denotes a convolution over the image space. The Nth partial derivative of an image f is measured at a scale s using the kernel

$$\left(\frac{\partial^N}{\partial {x_1}^{l_1} \ldots \partial {x_n}^{l_1}}\right)_s f(\vec{x}) \equiv \left(\frac{\partial^N}{\partial {x_1}^{l_1} \ldots \partial {x_n}^{l_1}} G(s)\right) \otimes f(\vec{x}), \tag{2.6}$$

where $\sum_{i=1}^{n} l_i = N$.

Florack (1992) points out that the Gaussian kernel has an associated class of integrable (but not necessarily differentiable) functions that define a *distribution*. Derivatives on such distributions are defined in the “weak” sense by integration with a kernel or *test function*. In many cases where classical differentiation is not possible such test functions provide an alternate means of measuring derivatives.

The Gaussian kernel (as well as the modified Bessel function) is essentially a low-pass filter, where the bandwidth of the filter is determined by the choice of scale in the Gaussian kernel. Indeed, Blom (1992) shows that in the case of Gaussian noise the application of Gaussian blurring can greatly improve the signal to noise ratio of image measurements that consist of derivatives of the image intensity. However, as successive sections will show, spatial averaging in computing image measurements has some drawbacks. These drawbacks will motivate a class of nonstationary blurring processes that are sensitive to image structure. First, it is important to understand Gaussian blurring as an iterative process.

2.5.2 Diffusion

The convolution of an image with a test function is equivalent (given the appropriate integrability conditions) to blurring the original image before taking the measurement. The Gaussian kernel is the fundamental solution to the diffusion equation (also called the heat equation), where scale replaces the time or evolution parameter, $t = \frac{s^2}{2}$.

$$\nabla \cdot c \nabla f = \frac{\partial f}{\partial t} \tag{2.7}$$

The constant c is the conductance, which controls the rate of blurring with respect to the time parameter. Viewed in this way, $f(\vec{x}, t)$ is a density function that describes the amount of a "substance" at each point in the image space. For images, that substance is whatever units are associated with L, in most cases those of luminance. The units in this equation are often chosen so that $c = 1$, and the conductance does not appear explicitly. The full solution $f(\vec{x}, t)$, with initial conditions $f(\vec{x}, 0) = f(\vec{x})$, is the continuous scale space mentioned above, and each time slice is a version of the original image after some amount of blurring. This allows one to view Gaussian blurring as a continuous process that evolves from the original image.

Lindeberg (1991) applies a set of constraints to the measurement of discrete images and shows that the necessary blur kernels are modified Bessel functions. He shows that the modified Bessel function is the fundamental solution of a discrete version of (2.7) computed on a grid. In the limit, as the spacing of the grid becomes infinitesimal, solutions to the discrete version approach the continuous case (Petrovski 1967), and the modified Bessel functions approach Gaussian kernels.

2.6 Example: Gaussian blurring

The example in Fig. 2.4 helps to illustrate some of the limitations of uniform Gaussian blurring. Figure 2.4 is constructed by drawing two overlapping equiluminant circles on a black background and then adding Gaussian white noise with a standard deviation at each pixel (the image contains 256×256 pixels) that is equal to the intensity difference between the circles and the background. The image is noisy, and yet the outline of the shape formed by the overlapping circles is easily discernible to the human observer. This visual boundary includes not only the smooth outline of the original circles, but the sharp cusp that is formed by their intersection. Figure 2.4b shows the gradient magnitude of this image that is computed taking differences between nearest neighbor pixels as in (2.3). Although there are some areas of particularly high values that coincide with the borders of the original circles, these areas are incomplete and difficult to reliably characterize.

Figure 2.5 shows the effects of two levels of Gaussian blurring in the image and the gradient magnitude of these blurred images. The effects of the random noise have been dramatically reduced, and the gradient images are more coherent. However, the gradients have formed relatively wide bright bands that introduce some ambiguity or uncertainty regarding the precise location of the boundaries in question. One could interpret the fuzzy response of the gradient measure at a pixel for a given scale as a graded membership function which indicates the likelihood of a boundary at that pixel. It is conceivable (but not trivial) that one could develop a set of fuzzy segments on the basis of these fuzzy boundaries. The width of these bands in the gradient image suggests a tolerance on the locations of boundaries that result from these measurements.

This analysis of blurred derivatives is rather unsatisfying. Refer to Fig. 2.5, and see that the sharp cusps where the circles meet were apparent in the original image but are less sharp in the blurred images. The sharp points associated with these cusps degrade very rapidly with Gaussian blurring, even though they might be important to the overall shape created by the two circles. These "tolerances" would indicate that we cannot know the location of those cusps very accurately. This contradicts subjective impressions of this image—the cusps *appear* well defined. The actual location of those cusps does not even generate a particularly "high" membership value. Regardless of how one decides to incorporate tolerances, the boundaries at the cusps in this image blur as easily as the noise.

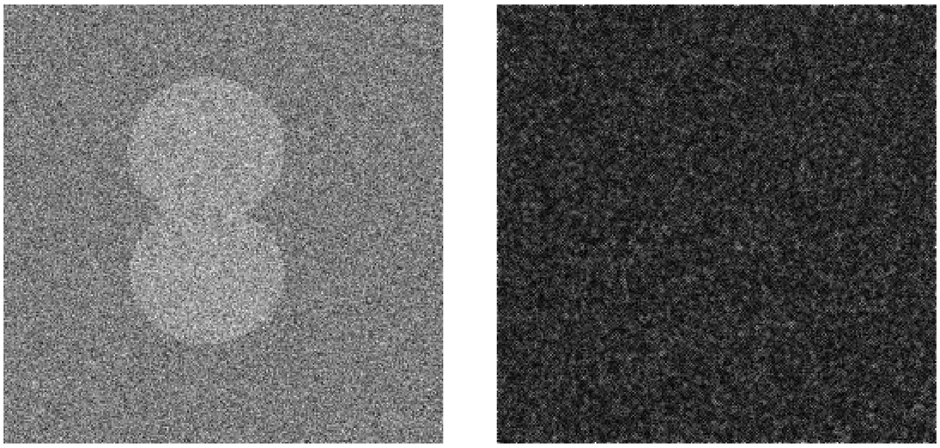


Figure 2.4: A noisy white figure (a) on black background shows the sensitivity of (b) the gradient magnitude to small scale noise. Despite the difference of average intensity between foreground and background in (a), the gradient magnitude in (b) shows little coherent structure. The derivative measurements behave as high-pass filters which serve to accentuate small-scale noise.

Figure 2.6 shows the Canny edges computed on this same figure at four different scales. At the point where the noise is sufficiently reduced in order to produce a coherent object, the cusps are no longer sharp. The indiscriminate reduction of small scale information associated with uniform blurring can remove important information about visual structures that have both small scale and large scale components. Ideally one would like to improve the tolerance of gradient measurements at places in the image where the large tolerances associated with larger Gaussian kernels lead to poor decisions about the location of the boundary. This fact is one the primary motivations for the research described in this dissertation.

2.7 Variable-conductance diffusion

The conductance term c in (2.7) controls the rate at which the density function f "flows" from areas of high concentration to areas of low concentration. Variable-conductance diffusion uses a variation on the heat equation which allows the conductance to vary over space as in (2.8).

$$\nabla \cdot c(\vec{x}, t) \nabla f = \frac{\partial f}{\partial t} \tag{2.8}$$

Grossberg et. al. (1984, 1985) have proposed variable-conductance diffusion as a model of low-level human vision.

Anisotropic diffusion as proposed by Perona and Malik (1990) and others (Nordstrom 1990, Shah 1991) is an alternative to the uniform scale space described by the diffusion equation. More specifically, anisotropic diffusion incorporates a variable-conductance term which limits the flow of intensity according to local gradient information. For this reason it is also called *edge-affected diffusion*. Several authors have proposed edge-affected diffusion in the form of the equation

$$\nabla \cdot g(|\nabla f|) \nabla f = \frac{\partial f}{\partial t} \ , \tag{2.9}$$

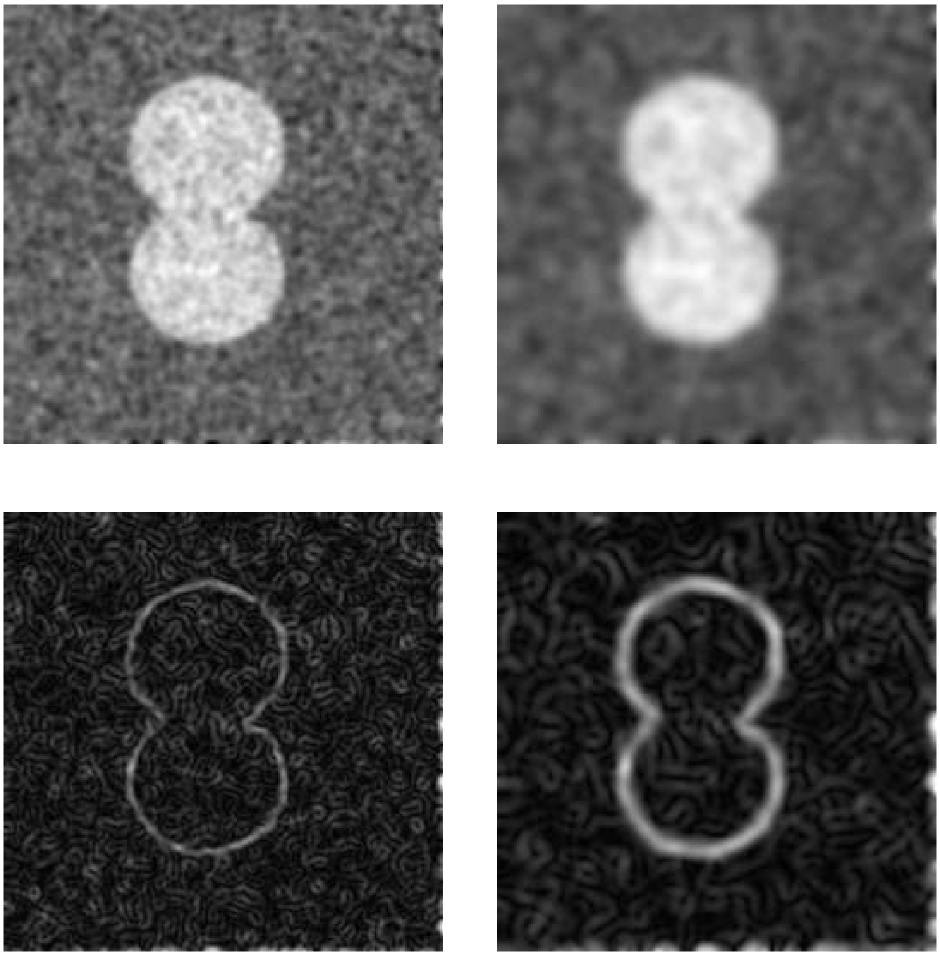


Figure 2.5: The noisy image of Fig. 2.4 with intensity (top) and gradient magnitude (bottom) conducted at scales 2.0 (left) and 4.0 (right).

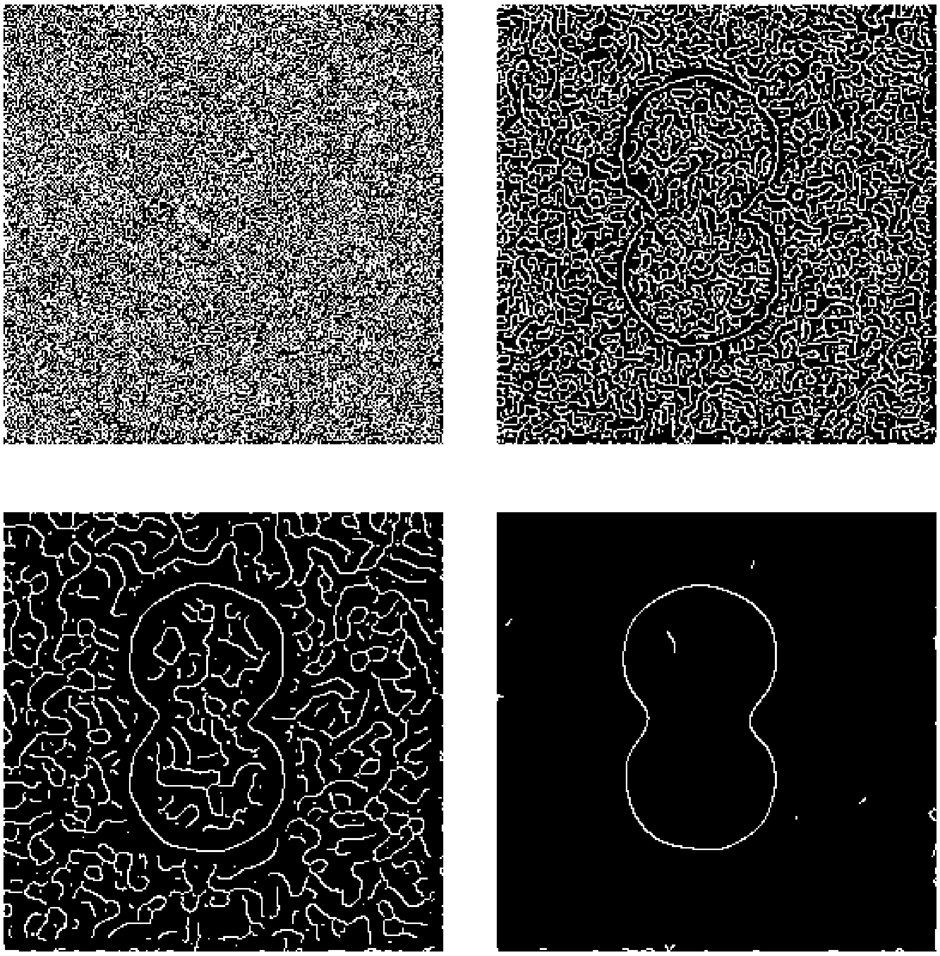


Figure 2.6: From the top left: Canny edges computed at scales 1.0, 2.0, 4.0 and 8.0.

where g, the conductance modulating term, is a bounded, positive, decreasing function of $|\nabla f|$, and t is the time or evolution parameter [2].

In this case the conductance $g(|\nabla f|)$ depends explicitly on the gradient magnitude of the function itself. Thus Eq. (2.9), unlike (2.7) and (2.8), is nonlinear. As a result, solutions are not described by any explicit convolution with a kernel. Perona and Malik have shown that (2.9) has the effect of limiting blurring near edges and also increasing the steepness or gradient of edges that are sufficiently steep in the initial image. The solutions of this equation can be thought of as a kind of anisotropic or nonuniform scale space. When the original image is used as the initial condition, this process produces, over time, a set of smooth regions with nearly constant luminance separated by sharp edges.

The behavior of Eq. (2.9) depends on an appropriate choice of $g(\cdot)$. Some constraints that yield well-behaved solutions are that $g(\cdot)$ be positive and bounded, and that in order to limit diffusion at edges $g(\cdot)$ should be monotonically decreasing. Such a conductance function insures the maximum principle: *Extrema in the solutions of this equation exist only along the boundaries, or the initial conditions.* Perona and Malik (1990) provide further constraints on $g(\cdot)$ that enable (2.9) to exhibit the "edge enhancing" behavior. They propose two possibilities,

$$\begin{aligned} g(|\nabla f|) &= \exp\left(\frac{-|\nabla f|^2}{2k^2}\right) \text{ and} && (2.10) \\ g(|\nabla f|) &= \frac{1}{1+\frac{-|\nabla f|^2}{k^2}}. && (2.11) \end{aligned}$$

Both of these functions introduce a conductance parameter k. This parameter controls the effect a given gradient value will have on the conductance. Ideally, k should be chosen to reflect the range of gradients in the image or possibly the gradients in a local neighborhood of every point.

The anisotropic diffusion equation introduces contains two free parameters. One is the total elapsed time, t. Solutions to (2.9) exist at a range of times and, assuming that subsequent analysis requires a single image, one must choose a time t, which indicates a single time slice from the entire solution. The parameter t controls the amount of diffusion, and it resembles the scale parameter s in Gaussian blurring in terms of both its mathematical function (for the special case of $g(|\nabla f|) = 1$, Eq. (2.9) is isotropic diffusion, i.e., blurring with Gaussian kernels of $s^2 = 2t$) and its affect on the resulting image (this is examined in more detail in Chapt. 5). Greater values of t and s result in "smoother" images.

The second free parameter is the conductance k. This parameter controls the degree to which edges control the flow of intensity. For very large values of k edges have little influence over diffusion, and anisotropic diffusion resembles Gaussian blurring. For many of the examples in this paper k will be expressed as some multiple of the root mean squared (rms) gradient magnitude in the image (Gerig, Kubler, Kikinis and Jolesz 1991). This has the effect of making the process independent of the particular units in which one chooses to express intensity and thereby allows a small range of k values to apply to a wide variety of images.

The choice of these free parameters t and k is an important issue in the use of anisotropic diffusion. A subjective examination of examples suggests that these choices will depend on the image that is being analyzed. One of the goals of the research presented here is to understand how sensitive the process is to these parameters, to quantify their effects on the output, and to construct diffusion processes that are less sensitive to these choices. These issues will be addressed

---

[2] The absolute value in $g(|\nabla f|)$ raises some questions about differentiability. I consider only conductance functions which are differentiable under such circumstances. For example, when using a Euclidean norm, functions of the form $g(|\nabla f|) = h(|\nabla f|^2)$ are differentiable as long as h is differentiable. This differentiability requirement will serve as a constraint when I generalize the conductance to include multi-valued functions as in Chapt. 4.

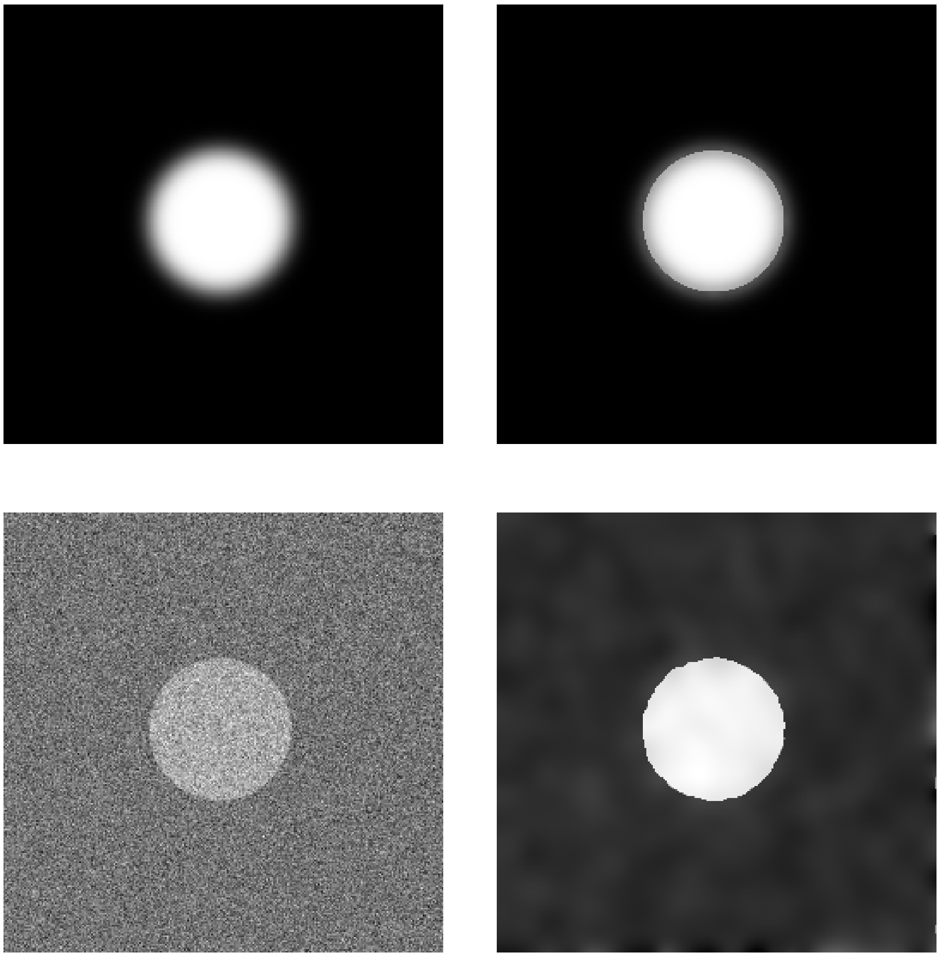


Figure 2.7: A disc with blurring (top) and noise (bottom) before (right) and after (left) anisotropic diffusion; 100 iterations with a k value of $2.0\ (\nabla f)_{\mathrm{rms}}$.

in subsequent chapters.

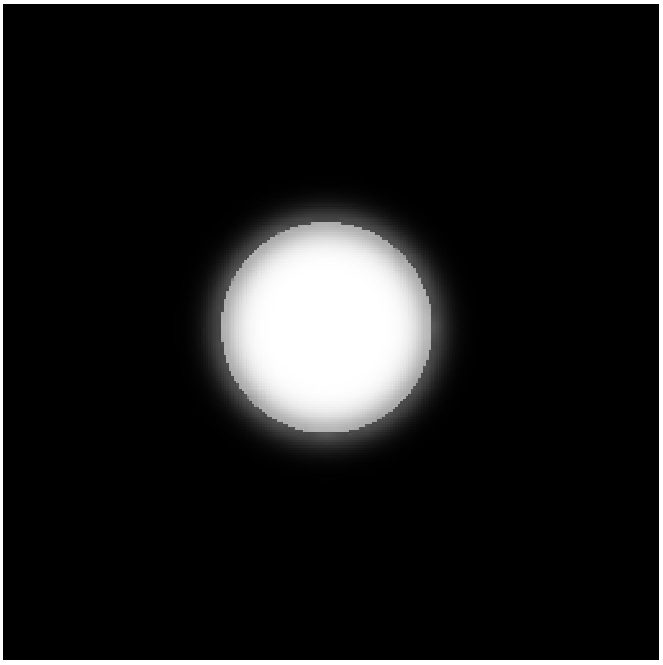
Figure 2.7a,b show the effects of the edge-affected diffusion process on a blurred disk. The algorithm employs the method of finite differences as described in Appendix A. It starts with the initial image and proceeds through discrete time steps. The gradient magnitudes are computed using nearest neighbor differences in orthogonal directions. The boundary conditions are chosen to be adiabatic: the derivative of intensity is zero in the direction perpendicular to the boundaries of the image. The time increments are chosen so as to maintain numerical stability (Perona and Malik 1990, Petrovski 1967), and the total elapsed time is chosen in order to accentuate the edge enhancing effects of the algorithm.

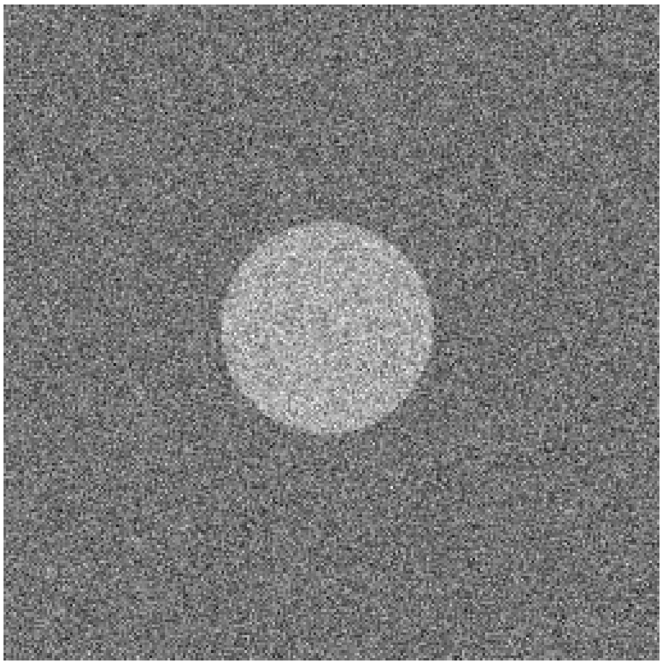
The edge-affected diffusion can blur images, reduce unwanted noise, and preserve (or even enhance) boundaries. The images that result tend to have steep, distinct boundaries that adhere to the shapes of objects in the original image. Solutions with adequate values of t offer more reliable gradient measures than in the original image (a quantitative justification is offered in Chapt. 5) and facilitate the use of Canny edges or even gradient thresholding as reliable indicators of boundaries.

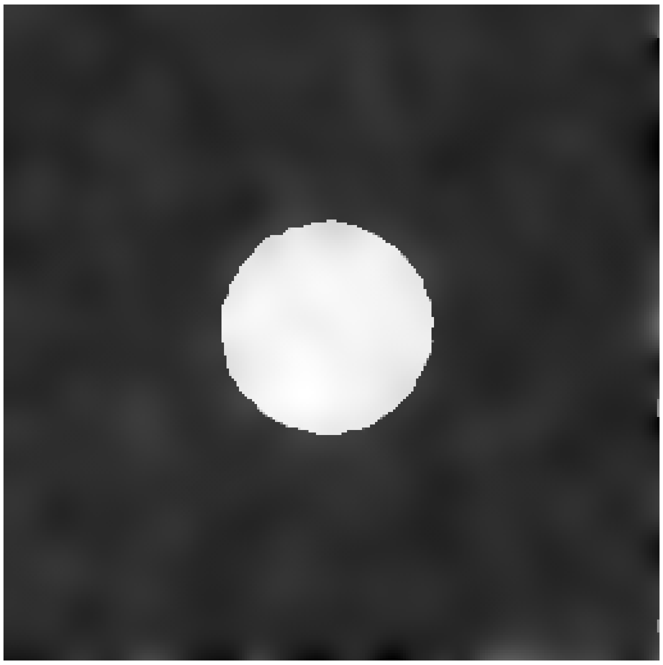
2.7.1 Regularization

It has been shown that the process described in (2.9) bears a strong mathematical resemblance to a class of relaxation algorithms that seek piecewise constant approximations to images (Nordstrom 1990, Snyder, Han, Bilbro, Whitaker and Pizer 1992). These algorithms typically define energy functions that penalize approximating images for differing too much from the input image and for having intensity variations other than steps at edges. Minimizing such energy functions using iterative techniques such as gradient descent yields processes that are similar, if not identical, to (2.9). This paper discusses a framework for generalizing this process in order to produce piecewise constant approximations to higher-order information. Although I discuss primarily nonuniform diffusion equations as in (2.9), I do so with the understanding that this framework has similar implications for a broad range of regularization problems that are solved by means of a constrained smoothing process.

While the anisotropic diffusion (and associated minimization procedures) offers some improvement to Gaussian blurring, it does not entirely settle the issue of scale. The anisotropic diffusion process as described here implicitly assumes that small-scale measurements are useful indicators of the presence of boundaries. The next chapter demonstrates the limitations of this assumption and introduces a "multiscale" diffusion process to alleviate this shortcoming.







2.7.1 Regularization

It has been shown that the process described in (2.9) bears a strong mathematical resemblance to a class of relaxation algorithms that seek piecewise constant approximations to images (Nordstrom 1990, Snyder, Han, Bilbro, Whitaker and Pizer 1992). These algorithms typically define energy functions that penalize approximating images for differing too much from the input image and for having intensity variations other than steps at edges. Minimizing such energy functions using iterative techniques such as gradient descent yields processes that are similar, if not identical, to (2.9). This paper discusses a framework for generalizing this process in order to produce piecewise constant approximations to higher-order information. Although I discuss primarily nonuniform diffusion equations as in (2.9), I do so with the understanding that this framework has similar implications for a broad range of regularization problems that are solved by means of a constrained smoothing process.

While the anisotropic diffusion (and associated minimization procedures) offers some improvement to Gaussian blurring, it does not entirely settle the issue of scale. The anisotropic diffusion process as described here implicitly assumes that small-scale measurements are useful indicators of the presence of boundaries. The next chapter demonstrates the limitations of this assumption and introduces a "multiscale" diffusion process to alleviate this shortcoming.