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

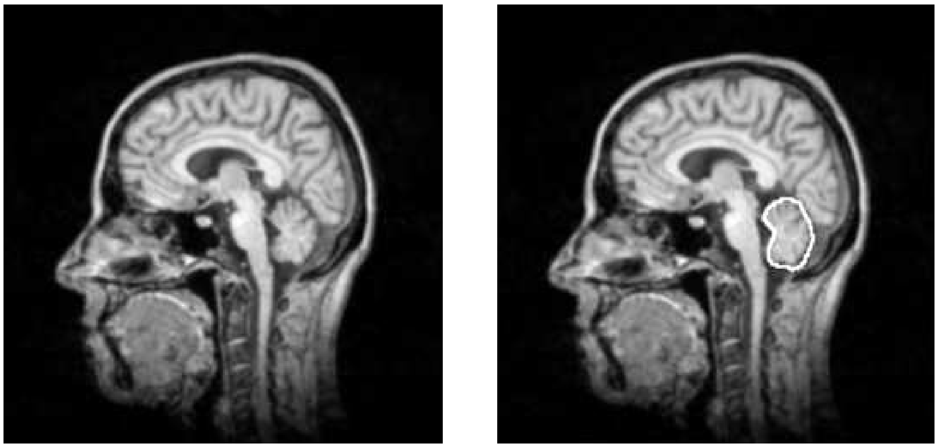


Figure 1.1: A slice of MRI image of a head shows that segmentations which are "easy" for our visual systems can be difficult for computer algorithms. The white outline in (b) identifies the cerebellum. Despite the ease with which the (untrained) human identifies this region as a distinct object, there 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

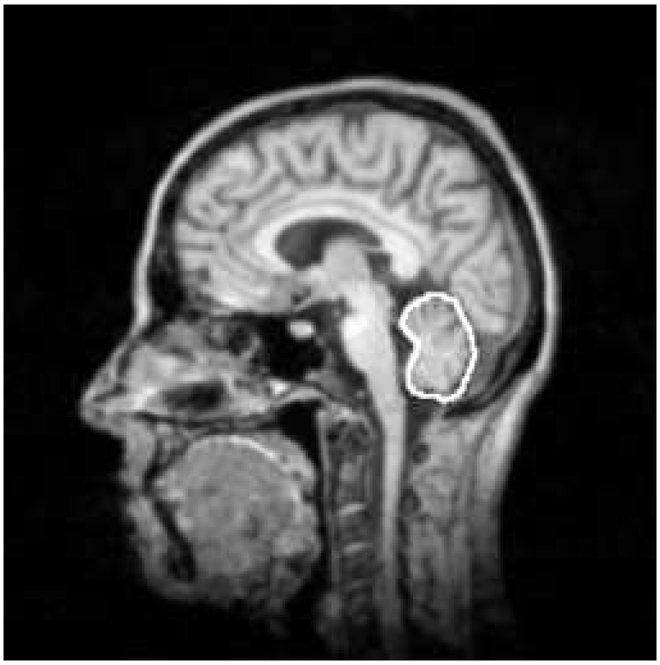
This dissertation presents a framework for constructing processes or algorithms which aid in image segmentation. I call this class of processes *geometry-limited diffusion.* My thesis is that *the geometry-limited diffusion framework that is described in Chapts. 3 and 4 offers a reliable and accurate 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.



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.