Chapter 1

Introduction

The driving problem of this dissertation is *image segmentation*, a process which identifies and labels objects in an image. Given an image, we ourselves can recognize objects. For example, the image in Figure 1.1 contains sunflowers. The flowers in the foreground are clearly recognized, including the individual petals. The background also contains flowers, but the individual petals are not so easily seen. A person could label the pixels, one at

Figure 1.1: Illustration of objects in an image

a time, with the appropriate object name, such as stem, petal, etc. Some pixels are not easily identified since objects are at times blurred together. For such a pixel, a person could provide some likelihood that the pixel is, for example, a stem or petal.

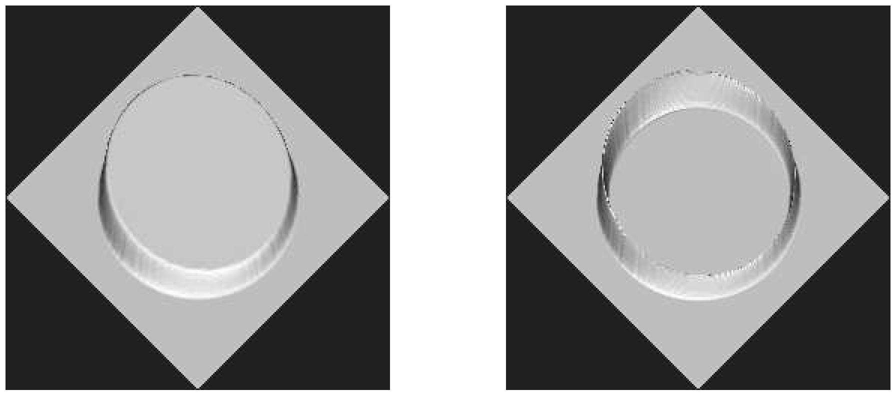
A pixel–by–pixel segmentation of an image that has thousands of pixels is a tedious and time–consuming task. I would like to automate the segmentation process to reduce the time it takes to segment, but in a way that models how the human visual system identifies objects. A full automation may be difficult. Humans can rely on previous experience and knowledge to aid in object recognition. Providing a computer with a knowledge base is certainly within reason, but is in the realm of artificial intelligence. Instead my goals are to process an image in the way that a front–end visual system[1] would, by using the local geometry induced by the intensity values of the image, and to create a representation of the objects that allows the user to explore the details of the image via an interactive computer system.

1.1 The Need for Ridges in Image Analysis

Methods for representing shapes of objects in gray–scale images have typically fallen into two categories: edge–based or region–based. Edge–based algorithms are developed under the assumption that large gradients of image intensity indicate the presence of an edge. The property of *edgeness* at a pixel is determined by measuring the *dissimilarity* between the pixel intensity and its neighbors' intensities, for example, by using the magnitude of the gradient of intensity. These algorithms additionally must handle edge orientation, edge strength, and edge connectivity. The method of edge detection essentially consists of following *ridges* of edgeness. Figure 1.2 illustrates this for a simple object. Many edge–based methods are deficient since the presence of noise can make it difficult to detect an edge and determine its orientation. Moreover, the characterization of the *global* structure and shape of an object by its boundary depends greatly on the correctness of the edge connectivity scheme.

Region–based algorithms are developed under the assumption that an object is locally homogeneous. The property of *interiorness* at a pixel is determined by measuring the *similarity* between the pixel intensity and its neighbors' intensities. Regions are essentially

[1]The term front–end is used by Florack (1993) and denotes the primary stage of a biological or artificial vision system. It is a syntactical system which establishes a representation of the input data in a format that can be interpreted easily by semantical systems in later stages of processing.



graph of intensity graph of magnitude of gradient of intensity

Figure 1.2: Edges as ridges of the magnitude of gradient of intensity

sets of positive measure whereas edges are sets of zero measure, so region–based algorithms tend to be less sensitive to noise. However, the region growing is still stopped at ridges of edge strength, and without the use of global information, there are still problems with noise sensitivity.

A good model for representing object shape should encapsulate the ideas of both edge– and region–based methods. The Blum medial axis construction for 2–dimensional binary objects was one of the first attempts at providing both region and edge information. The structure of the axis itself encodes the shape information of the object. The axis points are centrally located in an object. That is, each axis point pairs two opposing boundary points, called involutes, in the sense that a maximal disk (contained entirely inside the object) centered at the axis point touches the boundary tangentially in at least two points. The radii of the disks determine boundary locations in that the contour of the union of all the maximal disks is the boundary of the object.

Given a binary object, each point inside the object can be assigned its distance to the boundary of the object. The points outside the object are assigned a value of zero. The function corresponding to this assignment is called the *Euclidean distance transform* of the object. Ridges on the graph of the distance transform can be projected onto the plane to obtain medial–like structures. In Figure 1.3 the outline in the left picture is that of a binary object which resembles a face. The medial axis for the object is also shown. The right image

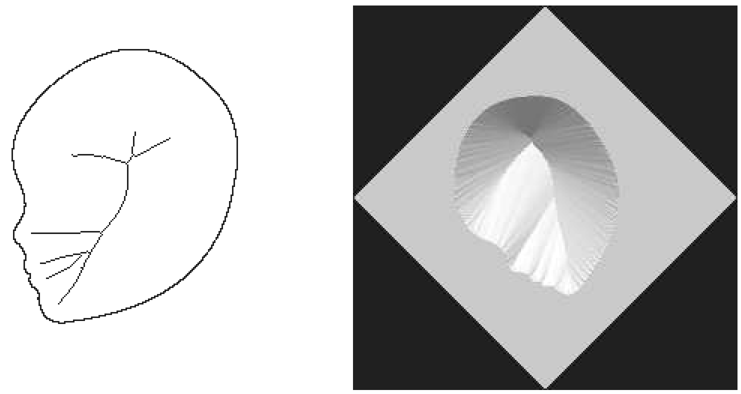


Figure 1.3: Binary object and medial axis, graph of distance function

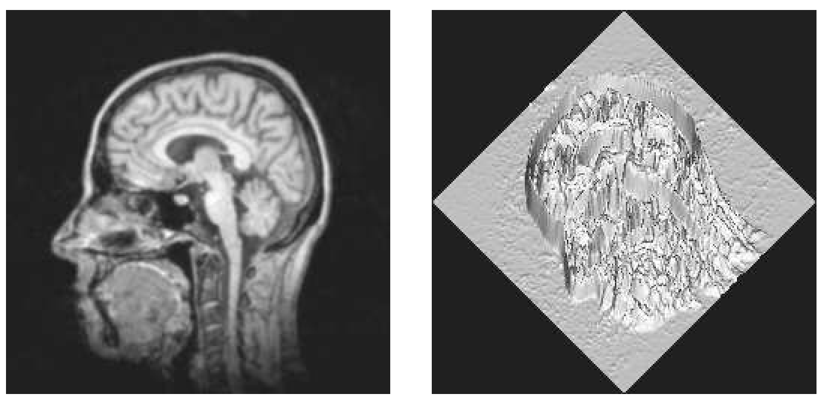


Figure 1.4: MR image and graph of its intensity values

is a shaded rendering of the graph of the Euclidean distance transform for the object.

The Blum medial axis construction requires a binary object whose edge locations are already known. Objects in gray–scale images are not binary and edge locations are not known, so the algorithm is of no practical use in this setting. However, for some types of images, the ridges of the intensity function may be medial–like. Figure 1.4 shows a Magnetic Resonance (MR) image and its corresponding intensity graph. Note how the ridges on the graph tend to be in the "middle" of objects such as the scalp or brain stem. The examples above clearly show that the analysis of images, and of objects in an image, necessarily requires the study of ridge structures of the underlying intensity surface.

In Chapter 2 I review many of the definitions for ridges that are found in the literature. Computational vision models require that medial structures should remain invariant under certain transformations of the spatial locations and intensities. For each ridge definition I point out which invariances the definition satisfies. The previous figures provide motivation for a ridge as a 1–dimensional structure in a 2–dimensional image. I also give extensions of the concepts so that d–dimensional ridge structures can be located within n–dimensional images. A comparison of the ridge structures produced by the different definitions is given by mathematical examples and by an application to the MR image of Figure 1.4.

1.2 Object Construction via Multiscale Methods

Ridges provide only partial information about the shapes of objects; they indicate only locations of object middles. A decision is needed about what radius or *object width* information to assign to the ridge points. Medial axes for binary objects capture the information about both central location and object width. Given the medial axis, the object can be reconstructed exactly. In gray scale images, it is not always clear precisely where object boundaries are. The object width values which are assigned to ridges should correspond to an *estimate* of how far the boundary of an object is from a ridge. Thus, the concept of *scale* of an object is needed, which is an indication of how wide an object is at any given central location. The scale need only be proportional to width, since measurements can always be made in units of the scale.

The introduction of a scale parameter in the image analysis is very powerful. Details in an image can occur at small scale or at large scale. Many early attempts failed at identifying objects, in particular methods which tried to locate edges of an object, because the measurements were made at the pixel scale. Small scale noise can easily interfere with the identification of edges. Moreover, the process of locating the entire closed contour of an object by following individual edge segments is bound to fail in that the contour is an entity which has a larger scale than its individual components. By analyzing an image at multiple scales, both local and global shape information can be extracted.

In Chapter 3 I present a general method for segmenting medical images which uses multiscale methods. The scale parameter is introduced by blurring the original image with a radially symmetric Gaussian kernel whose standard deviation is the scale. At each scale the blurred image is segmented into primitive regions. These regions are constructed by finding ridges on the intensity surface of the image, followed by associating pixels with each ridge via a process called *ridge flow*. The area/volume of the regions is approximately proportional to the scale of the blurring process.

The segmented image is represented by a tree structure, which naturally relates local information obtained at small scale to more global information obtained at large scale. The leaf nodes of the tree represent small–scale primitive regions. The interior nodes of the tree represent increasingly larger scale regions as the tree is traversed toward the root. The root node corresponds to a single region at essentially infinite scale. The interior nodes of the tree relate the regions in a way that reflects the natural object structure of the image. Objects are obtained as unions of subtrees of the hierarchy.

1.3 Scale Space: A Foundation for Image Analysis

The segmentation process described in the last section is based on a multiscale analysis, but it does not rely on information about *changes in scale*. That is, a scale of measurement is selected (the standard deviation of the Gaussian kernel), then the image is segmented at that fixed scale. As an analogy, the medial axis of a binary object stores central locations and object widths, but one can derive from these quantities information such as curvature of the object boundaries and rate of expansion or contraction of the object as the axis is traversed in a specified direction. These derived quantities are just as important in shape analysis as the medial locations and object widths. Moreover, the derived quantities are dependent on scale changes in the objects.

Successful algorithms for image analysis should be based therefore on making measurements in both space and scale. The collection of all pairs of spatial locations and scales will be denoted as *scale space*. The essential foundations of a scale space are that an image is a physical observable with an inner scale, determined by the resolution of the sampling device, and an outer scale, limited by the field of view. Basing the analysis on the model of a front–end visual system, I require that the methods preprocess input in a symmetric way (rotational and translational invariance), and they should have no preferred scale of measurement (scale invariance). These fundamental assumptions lead naturally to the idea that the scale parameter must be introduced by a possibly nonlinear blurring process. But even more importantly, they lead to the idea that measurements are only meaningful relative to the scale at which they are made. The most significant consequence is that the geometry of scale space is *not Euclidean.*

Chapter 4 provides the mathematical formalism for scale space as a geometric entity. All the familiar measurement tools in Euclidean geometry are derived in the scale space setting using tensor calculus and differential geometry. The basic concepts of distance, area, volume, curvature, and differentiation are developed. I also define what it means to be a ridge for a function defined on scale space. Just as the medial axis for a binary object can be viewed as a curve whose points have spatial and radial components, *cores* (formerly called *multiscale medial axes*) for gray scale images can be constructed as ridges in scale space. Although the material first appears unintuitive and perhaps formidable, further reflection will show that the quantities really do support the intuition about the concept of scale.