STATISTICAL MODELS FOR SHAPES AND DEFORMATIONS

by

Alyson Gabbard Wilson

Department of Statistics and Decision Sciences
Duke University

Date: _____
Approved:

_____
Dr. Valen Johnson, Supervisor

_____
Dr. Peter Müller

_____
Dr. Loren Nolte

_____
Dr. Dalene Stangl

_____
Dr. Robert Wolpert

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Statistics and Decision Sciences
in the Graduate School of
Duke University

1995

ABSTRACT

(Statistics)

STATISTICAL MODELS FOR SHAPES AND DEFORMATIONS

by

Alyson Gabbard Wilson

Department of Statistics and Decision Sciences
Duke University

Date: _____
Approved:

_____
Dr. Valen Johnson, Supervisor

_____
Dr. Peter Müller

_____
Dr. Loren Nolte

_____
Dr. Dalene Stangl

_____
Dr. Robert Wolpert

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Statistics and Decision Sciences in the Graduate School of
Duke University

1995

Abstract

Bayesian methods are valuable in image analysis because there is often *a priori* information that can contribute to the analysis of an image. This prior knowledge may be general (e.g., intensities vary smoothly across the image), or may be more specific (e.g., this is an image of a brain that may have a tumor). This dissertation develops flexible methods to incorporate prior knowledge from templates into algorithms for image analysis within a Bayesian framework. Classes of priors on landmark locations are developed that assign high probability to images "like" the template and low probability to images "unlike" the template.

The priors build on previous work in Bayesian image analysis by incorporating ideas from Markov random field priors and deformable template models. The prior models differ from standard applications of MRF models in that the sites in the fields represent image objects, and the random variables associated with the sites represent their locations. Another crucial idea grows from methods in computer vision research. Features in an image occur at a variety of scales, and to effectively model spatial and scale relationships, this variety of scales must be modeled. *Scale space* is an image representation that handles image structure at all resolutions simultaneously and allows efficient calculation using features at multiple scales. It also allows models to be specified that

are rotation, translation, and zoom invariant.

Other important aspects of the prior include quantifying feature similarity between images, locating landmarks within images, and measuring distances and spatial relationships between landmarks. The priors address these issues and incorporate the feature and location information into a hierarchical model. The hierarchical framework is natural for handling deformations and obstructions. Further, it allows the modeling of such properties as "the location of large-scale features is less variable than the location of small-scale features."

The priors on landmarks are used to perform automatic landmark identification. They also hold promise for tasks like automatic object recognition.

Acknowledgements

I would like to acknowledge the contributions of Dr. Dave Eberly, Dr. Steve Pizer, Dr. Liyun Yu, and Matt McAuliffe from the Computer Science Department at the University of North Carolina, Chapel Hill. I would like to thank Dr. Jim Bowsher for many helpful discussions. Special thanks go to Dr. Valen Johnson for the many hours and ideas he contributed to this work. And thanks to Greg Wilson for his love and support. The research in this dissertation is based upon work supported under a National Science Foundation Graduate Research Fellowship.

Contents

List of Figures

Chapter 1

Overview

Image analysis is concerned with the modeling, investigation, and display of "pictorial" data. The prototypical image is a photograph, which can be thought of as a continuous two-dimensional light intensity function $g(\mathbf{x})$, where $\mathbf{x}$ denotes a spatial coordinate and $g(\mathbf{x})$ is proportional to the brightness of the photograph at that point. The intensity function g can, however, describe energy from many sources, including electromagnetic radiation like infrared and X-rays, ionized high-energy particles, ultrasound, or pressure. Similarly, the spatial coordinate is not restricted to two dimensions, but can be n-dimensional.

A *digital image* is an image that has been discretized both in spatial coordinates and in brightness. The intensities associated with each spatial coordinate can be scalars, for example, gray levels, or vectors, as in the case of multi-spectra satellite images. Most image data is collected in the form of digital images; consequently, most image data can be thought of as a quantization of an underlying continuous process (Geman and Gidas 1991). The elements of the digital image array are called *image elements*, *picture elements*, or *pixels*.

A variety of techniques have been developed to investigate and analyze

images, including restoration, reconstruction, segmentation, and registration. *Restoration* enhances images that have been degraded by a known process. *Reconstruction* rebuilds images from projections, and is the process used to transform the raw data collected in computerized tomography (CT) to a digital image. *Segmentation* identifies and labels structures within images. *Registration*, or *matching*, establishes a point-by-point correspondence between two images.

Statistical techniques have been successfully applied in many areas of image analysis, especially those that are inherently "model-based." In restoration, for example, explicit models are derived that relate the "true" image to the observed image. Bayesian methods are especially useful, as one often has prior knowledge about the true image. This prior knowledge may be very general (e.g., intensities vary smoothly across the image), or may be more specific (e.g., this is an image of a brain that may have a tumor).

There is often an opportunity to incorporate specific prior knowledge into the analysis of medical images. When a diagnostic image is created, there is knowledge about what part of the body has been imaged and about why the image has been made. A template may be available in the form of a medical atlas, or as a cross-correlated high-resolution magnetic resonance (MR) or CT image of the same patient. The question addressed in this dissertation is how to use this type of prior information to enhance an image's analysis.

There are a variety of applications that can benefit from the effective use of template information. Multi-modal image analysis is an active research area in medical imaging. In clinical practice, a single photon emission computed tomography (SPECT) image might be made to explore tissue function and a

MR image to examine anatomy. In radiotherapy, for example, a CT scan is needed for dose distribution calculations, while an MR image is used to outline the target lesion (van den Elsen et al. 1993). By combining information from the two images, a correspondence can be established between structure and function. Appropriate use of a prior containing template information facilitates the registration of the images, the segmentation of the MR image, and the reconstruction of the CT scan. Consequently, the appropriate therapeutic dosage can be directed more precisely to the target lesion.

Segmentation is another active area of research. In MR images, the small-scale structure often has poor resolution. To divide the image into regions, expert human information must be incorporated; in essence, the technician segments the images based on "what ought to be there." This process can be automated by using prior information from an atlas to capture beliefs and confidence about small-scale structure. More generally, prior templates can be used for object identification.

Applications outside the medical field are also common. Short (1993) discusses the automated detection of rail surface defects. Phillips and Smith (1994) model human faces, with an eye toward developing aids to witness identification of criminal suspects. Grenander and Manbeck (1993) propose an automatic defect detection system for potatoes.

The most general goal of this dissertation is to develop flexible methods to incorporate prior knowledge from templates into algorithms for image analysis within a Bayesian framework. This is addressed by developing prior distributions on landmark locations and shapes. The priors assign high probability to images "like" the template and low probability to images "unlike" the tem-

plate. The priors are useful for a variety of tasks, but especially for automatic landmark identification, object recognition, and image registration.

As with any choice of prior, there must be a tradeoff between how well the class of priors can capture *a priori* beliefs, how easily the posterior distribution can be computed, and how appropriate the results are to the task at hand. Within these constraints, there are a number of specific objectives.

The priors should incorporate ideas about scale and space. Regions and features within images occur at a variety of scales. Some features, like boundaries, are local, while others, like middles, are more global. Images can be examined for small details or for large shapes. The prior must account for feature arrangement both in space and in scale. It should also be able to model properties like "the location of large-scale features is less variable than the location of small-scale features."

Additionally, the priors should be hierarchical. Phillips and Smith (1994) model human faces by looking first for the head, then the face, then the eyes, and so on. The location of the eyes is conditional on the location of the face, which in turn is conditional on the location of the head. A hierarchical structure provides a natural framework for the description of a template and of features at different scales. By using a branching structure, deformations and obstructions can be handled in a straightforward manner. The structure of the hierarchy may not be strictly decreasing in scale; if one is looking for a square, it may make sense to look for corners, which are small-scale features, first.

Finally, the priors should generalize easily. To be broadly applicable, the basic structures of the priors must work in n-dimensions, and they must be able to

model binary, gray-scale, continuous, or digital images. In addition, the models developed using the priors must be computable. Likelihood-based methods for image analysis, including maximum likelihood estimation and Bayesian models, are of great research interest, but have achieved little clinical use due to their long computation times.

These objectives are achieved by using a Markov random field (MRF) model to specify a prior on landmark locations. The model differs from the standard application of MRFs to imaging because the sites in the model represent objects within the template, and the associated random variables are the locations of the objects. Typically, the sites within MRF models are fixed at the pixels, and the random variables are pixel intensities. Hierarchical relationships between objects are specified using the neighborhood relationships between sites. Computations using the MRF model are straightforward because MRF models can be simulated using Markov chain Monte Carlo techniques.

Ideas about scale are incorporated using methods adapted from computer vision research. *Scale space* is an image representation that handles image structure at all resolutions simultaneously and allows efficient calculation using features at multiple scales. Using properties of scale space, the priors can be chosen to be rotation, translation, and zoom invariant. These invariances allow the prior to favor configurations of landmarks like those in the template without regard for orientation, location, or size. Scale space also provides a natural framework for the specification of likelihood functions that capture ideas like "medialness" and "cornerness."

The remainder of the dissertation proceeds as follows. Chapter 2 discusses both the "statistical theory of shape" and the basic paradigm of Bayesian

image analysis. Special attention is given to the types of priors commonly used in modeling shape and deformation. Markov random field models are presented, as well as priors that have been proposed to explicitly model templates. Chapter 3 develops tools that are useful in specifying a flexible class of priors, including ideas about scale, resolution, and multi-scale image analysis. Chapter 4 examines multi-scale methods useful for describing features and templates. The statistical development of the priors is presented in Chapter 5, where a multi-scale prior on landmarks within images is specified. Chapter 6 applies this prior to the automatic identification of structures in MR brain images. Chapter 7 summarizes the work and proposes extensions.

Chapter 2

Statistical Methods for Modeling Shapes and Images

This chapter examines both the "statistical theory of shape" and Bayesian image analysis. Both of these methodologies are useful for the stochastic modeling of shapes, images, and deformations. The statistical theory of shape focuses on the statistical analysis of point patterns in $\mathbf{R}^n$. Bayesian image analysis examines how prior information on the space of "true" images and likelihood information for the observed images can be combined into posterior inferences. The ability to assess posterior variability makes Bayesian image analysis a natural framework for addressing many problems of medical interest, such as the determination of the volume of regions within an image and the uncertainty associated with those volumes. These two statistical approaches to the modeling of landmarks, images, and deformations, are combined in Chapter 5 to create a new class of priors on landmarks.

2.1 Statistical Theory of Shape

Landmarks are a traditional method for describing shape and shape change. *Landmarks*, in their most general form, are a set of k labeled points in $\mathbf{R}^n$ used to describe a figure and its shape. *Shape* is taken to be those characteristics of a figure that are unchanged when the figure is translated, rotated, or scaled. Landmark data arise in many fields, including biological and medical settings (Bookstein 1991), archaeology, astronomy, cartography, manufacturing, geology, and geophysics. Goodall (1991) states:

> In some instances landmarks may refer to the same physical markers identifiable in more than one map, satellite image, X-ray, etc. In instances of biological *homology*, landmarks are uniquely defined locations, most often in the skeleton, that are identifiable across individuals, and in interspecies comparisons, linked by a presumed evolutionary pathway. At its most general, a set of landmarks is a set of labeled points, found in at least two figures, whose relative positions in the two or more figures has some scientific, educational, or artistic interest to us (pp. 285–286).

One of the simplest ways to compare two figures is to find a transformation that takes one to the other. To simplify matters, one often looks for an affine transformation. Because it is typically not possible to find an exact affine transformation, it is necessary to approximate the "best" transformation. Lele (1989) describes the typical procedure. Let A and B be two sets of k landmarks. Fix A as the reference figure. Select positive real-valued functions λ_i, the *loss functions*, for $i = 1, \ldots, k$. Translate, rotate, and scale B so that $\sum_{i=1}^{k} \lambda_i(d(i_A, i_B))$ is minimized, where $d(i_A, i_B)$ is the distance between landmark i in A and B.

The choice of the loss functions leads to different superimpositions of the figures on one another: for example, Bookstein's (1991) method, ordinary Pro-

crustes analysis, or weighted ordinary Procrustes analysis (Goodall 1991). It can be shown (Lele 1989) that by choosing the appropriate loss function, one can support almost any hypothesis about whether shapes differ. However, typically the question of interest is not *whether* the shapes differ, but *where* the shapes differ. The answer to this question also depends on the choice of loss function. Lele (1989) provides an example analyzing the shape difference between two triangles, where three "reasonable" choices of loss function lead to radically different conclusions.

A more robust method for comparing two figures requires the development of a *shape space*, or a representation of the shapes of sequences of labeled points. Following Kendall (1984), let the shape space of k labeled points in n-dimensional space be denoted by Σ_n^k. The development of the shape space Σ_2^3, the space for triangles in $\mathbf{R}^2$, illustrates the main ideas of the theory.

Consider the triple of points (x_1, x_2, x_3) in $\mathbf{R}^2$. The shape of the triangle defined by these three points is invariant to translation, rotation, and scale, but may change if the points are relabeled. If these points are considered to be in the complex plane, $\mathbf{C}$, then the set of points defined by the *complex affine transformations* $\{(wx_1 + z, wx_2 + z, wx_3 + z) : z \in \mathbf{C} \text{ and } w \in \mathbf{C}\backslash\{0\}\}$ all have the same shape. Adding the same complex number z to each x_i corresponds to translation; multiplying each x_i by the same non-zero complex number w corresponds to scaling and rotation.

If (x_1, x_2, x_3) is centered by subtracting the centroid from each coordinate, then the set of triangles with the same shape is the set of vectors of the form $\{(w(x_1 - \bar{x}), w(x_2 - \bar{x}), w(x_3 - \bar{x})) : w \in \mathbf{C}\backslash\{0\}\}$. (Each possible triangle (x_1, x_2, x_3) is associated with a unique centered triad, and the centering removes

the effect of translation.) This set is a complex line through the origin of complex two-dimensional space; the collection of such lines is called $\mathbf{CP}(1)$, or the complex projective one-space. In general, the shape space of k-ads of planar points is given by $\Sigma_2^k = \mathbf{CP}(k-2)$, where $\mathbf{CP}(k-2)$ is the *complex projective* $k-2$ *space* and can be thought of as the space of complex planes passing through the origin in $\mathbf{C}^{k-1}$ (Small 1988; Stoyan et al. 1987).

It is difficult to visualize complex projective spaces. However, $\mathbf{CP}(1)$ is topologically and smoothly equivalent to a sphere (Stoyan et al. 1987), and the shape of the original triangle (x_1, x_2, x_3) can be represented as a point on the sphere. Let (r, θ) be the polar coordinates of the complex number $\frac{2x_3-(x_1+x_2)}{x_2-x1}$ (the projective coordinate map in Stoyan et al. (1987)). Then $(\frac{1-r^2/2}{2(1+r^2/3)}, \frac{r\cos(\theta)/\sqrt{3}}{1+r^2/3}, \frac{r\sin(\theta)/\sqrt{3}}{1+r^2/3})$ are the Euclidean coordinates of a point on a sphere of radius 0.5 about the origin in $\mathbf{R}^3$; each triangle shape corresponds to exactly one point on the sphere. Stoyan et al. (1987) discuss where particular classes of triangles (equilateral, isosceles, right) fall on the sphere.

"One interpretation of the choice of radius $\frac{1}{2}$ for the sphere of triangle shapes is that it ensures that geodesic distance on the sphere corresponds to the Procrustes metric" on $\mathbf{C}^3$, the space of triad points (Stoyan et al. 1987). Bookstein (1986, 1991) and Small (1988) discuss further implications of this choice of metric and propose possible alternatives. With the Procrustes metric on $\mathbf{C}^3$, it can be shown that if x_1, x_2, and x_3 are independent and identically distributed bivariate normal points (with mean 0) with a variance-covariance matrix that is a scalar multiple of the identity matrix, then the shape of the triangle defined by (x_1, x_2, x_3) will be uniformly distributed on the sphere of shape coordinates (Small 1988).

A useful approximation for the shape space Σ_2^n is given in Goodall (1991). Let $x_1, x_2, \ldots, x_n \in \mathbf{C}$ be the complex coordinates of the landmarks of a planar figure. Let $u = (u_1, \ldots, u_{n-2})$ be defined as $u_i = \frac{x_{i+2}-x_1}{x_2-x_1}$. The u_i are called *Bookstein coordinates*. The Euclidean geometry of $u \in \mathbf{R}^{2(n-2)}$ is a useful approximation to the metric geometry of Σ_2^n; for $n = 3$, Bookstein coordinates are a stereographic projection of the spherical coordinates.

Applications of the statistical theory of shape typically investigate departures from randomness in a point pattern; there is usually less concern with explaining structure than with determining whether structure is present at all (Stoyan et al. 1987). Reviews of the statistical theory of shape from a morphometric perspective can be found in Bookstein (1986) and from a stochastic geometry perspective in Kendall (1989).

2.2 Bayesian Image Analysis

The Bayesian analysis of an image can be considered in the traditional framework of specifying a likelihood function and a prior distribution. The form of the likelihood is, of course, problem and task specific, but it is often well-described by the imaging modality or by a large set of training data. Many models appropriate to various imaging modalities and tasks are described in Geman (1988), Ripley (1991), and Karr (1991).

Consider the following general formulation of Bayesian image analysis (Ripley 1991; Geman 1988; Geman and Gidas 1991). There are a variety of problems that commonly arise when analyzing digital images. Besides digitization, three types of distortions are common: blur, noise, and sensor effects (Geman and Gidas 1991). *Blur* is the scattering caused by the medium itself (e.g., the

atmosphere or human body), by an unfocused camera lens, or by the motion of the medium, objects, or sensors. *Noise* is randomness introduced by the sensing and recording devices—for example, by film grain or current fluctuations in electronic scanners. Sensor effects, also called *radiometric distortion*, are nonlinear transformations of the data caused by the sensing and recording devices. Geman (1988) contains detailed examples that illustrate these distortions.

Let $g(\mathbf{x})$ denote the observed value at pixel $\mathbf{x}$, and $f(\mathbf{x})$ be the "true" signal that would be measured under ideal conditions. Then

$$g(\mathbf{x}) = \psi[\phi[K(f(\mathbf{x}))], \epsilon] \tag{2.1}$$

where K accounts for blurring, ϕ represents radiometric distortion, ϵ is a noise vector, and ψ defines the noise mechanism (e.g., additive or multiplicative). The noise may be spatially correlated, and may differ depending on what part of the image is considered.

In this context, the problem of restoration is to infer $f(\mathbf{x})$ from $g(\mathbf{x})$. Often (2.1) can be simplified to

$$g = Kf + \epsilon \tag{2.2}$$

where the blurring can be represented as a convolution and the noise is additive.

Given a model, like (2.2), for the degradation of $f(\mathbf{x})$ to $g(\mathbf{x})$ and a prior distribution for $f(\mathbf{x})$, by Bayes' rule

$$p(f|g) \propto p(g|f)p(f)$$

A point estimate for $f(\mathbf{x})$ might be the mode or the mean of $p(f|g)$.

2.2.1 Markov Random Field Priors

Once a likelihood has been specified, a prior distribution must be chosen on the space of true images. There have been many proposals for how to do this, but one of the most widely used classes of priors in Bayesian image analysis is the Markov random field (MRF). Typically in these models, the intensity at a pixel is modeled as depending on the intensities at nearby pixels. Examples of analyses using MRF priors can be found in Besag (1986), Besag et al. (1991), Ripley (1988, 1991), and Short (1993).

Besag (1974) introduces MRFs as a way to model spatial processes. MRFs allow distributions to be specified in terms of local conditional distributions rather than in terms of joint distributions. Specification of distributions in terms of local characteristics is often intuitively easier than specification of relationships on a global scale. Further, the conditional specification makes MRF models especially amenable to Markov chain Monte Carlo techniques for sampling from the posterior on the image scene. Specifying local conditional distributions, however, places restrictions on the form of the joint distribution.

To define a Markov random field model, let $\mathbf{S} = \{s_1, s_2, \ldots, s_N\}$ be a set of sites (e.g., a set of pixels on a digital image), and let $\mathbf{Y} = \{Y_s, s \in S\}$ be any family of random variables indexed by $\mathbf{S}$. Let Λ_s be the state space for Y_s, and define the configuration space $\Omega = \{\omega : \omega = (y_{s_1}, \ldots, y_{s_N}), y_i \in \Lambda_i\}$. Define a *neighborhood system* $\mathbf{N} = \{N_s, s \in S\}$ as any collection of subsets of $\mathbf{S}$ for which (1) $s \notin N_s$ and (2) $s \in N_r$ iff $r \in N_s$. A subset $C \subseteq \mathbf{S}$ is a *clique* if every pair of distinct sites in C are neighbors.

$\mathbf{Y}$ is a MRF with respect to a neighborhood system $\mathbf{N}$ if

$$p(Y_s = y_s | Y_r = y_r, r \neq s) = p(Y_s = y_s | Y_r = y_r, r \in N_s) \tag{2.3}$$

If $\mathbf{Y}$ satisfies (2.3) and

$$p(Y = \omega) > 0 \text{ for all } \omega \in \Omega$$

then the joint distribution of Y is a Gibbs distribution with respect to $\mathbf{N}$. A Gibbs distribution has the form

$$p(\omega) = \frac{1}{Z} \exp(-U(\omega))$$

where

$$U(\omega) = \sum_{C \in C} V_C(\omega)$$

Z is the *partition function* and is constant, U is the *energy function*, C is the set of cliques, and V_C is a *potential function* that depends only on those y_s for which $s \in C$.

According to Geman and Geman (1984), the utility of MRF for image modeling is that "priors are available with neighborhoods that are small enough to ensure feasible computational loads, and yet still rich enough to model and restore interesting classes of images." Small neighborhoods are not necessary, however; Johnson (1994) uses the whole image as the neighborhood, and chooses non-zero potentials on only a small number of cliques.

While MRFs have nice local properties, they often have undesirable global characteristics. In particular, the mode of many of the MRF priors is a single-color image. Another objection to these models is that correlations can be introduced between neighboring pixels that have very different values. Such contrast often occurs on the boundaries of objects, and without modification, MRF models can blur these boundaries. This is a serious problem, because segmentation is often a primary goal of image analysis.

To correct blurring across boundaries, Geman and Geman (1984) introduce a set of auxiliary variables called *line sites* between each pixel. The line sites are either "on" or "off." When the data indicate that a boundary is present, the line site is turned on and local correlations are broken. Line site models are useful in segmenting four to six level images, but are computationally intractable for continuous-valued images (Johnson et al. 1991). Johnson et al. (1991) extend binary-valued line sites to continuous-valued by letting the line site value correspond to the degree of correlation between adjacent pixels. Although this improves the sampling properties of the posteriors, it is not successful in imposing smoothness on the shapes of the estimated boundaries. Recent models incorporating line sites are found in Leahy and Yan (1991), Gindi et al. (1991), and Gindi et al. (1993).

A shortcoming of models that incorporate line sites is that segmentations do not follow readily from the estimated line sites. It is difficult to insure that line sites connect to form closed regions. Although it is easy to incorporate prior boundary information (turn on the line sites corresponding to the boundary), it is unrealistic to assume that the boundaries from a template correspond precisely to the boundaries in the blurred image. If the uncertainty is ignored, then false edges appear in the restored image. It is possible to modify the potential functions around the boundary; however, if all line sites are encouraged to form independently, ringing artifacts can occur (Johnson 1994). The modifications of the potentials must keep the number of activations down while encouraging connected, closed regions (to avoid smoothing "leaks"). This can be difficult within the local MRF structure (Johnson 1994).

2.2.2 Region-based Models

A variety of other prior models have been proposed for use in Bayesian image analysis. These priors can be grouped into four classes: region-based models, deformation models, boundary models, and differential geometry models. Those that are the most successful in modeling templates and deformations use information from a variety of scales in the image: deformation, as a rule, is not a strictly local phenomenon, and changes in shape must be accounted for both globally and locally.

Region-based models are exemplified by Johnson et al. (1993). Their method uses a MRF prior on region labels rather than on pixel intensities. The model assumes that an image is composed of an unknown number of intensity-varying objects or regions. A MRF model is used to specify the probabilities of all possible configurations of region identifiers, or equivalently, on all possible partitions of the image into regions. Prior structural information is incorporated into the specification of probabilities on configurations.

A Gibbs distribution is specified for the configuration of region identifiers, R. The neighborhood system for the distribution is taken to be the entire image. This neighborhood system could make computations intractable, as the conditional distribution of each pixel could depend on the values at all other sites; however, the potential functions are chosen so that the large neighborhood poses little computational problem.

Three general prior beliefs about region configurations are modeled. Large numbers of regions are discouraged; irregular object shapes are discouraged; disconnected objects are prohibited. To accomplish these goals, three types of potential functions are required.

The first potential function discourages large numbers of regions. The form of the function depends on the image itself, and on the anticipated number of regions, K, in the image. If there is little knowledge about the number of regions, a function of the form $V(R) = \alpha K$ might be appropriate. If there is knowledge that the number of regions is small, a quadratic potential of the form $V(R) = \alpha K^2$ might be appropriate. If the quadratic potential is used, then the probability that a new region is created during a pixel update is $\exp(2\alpha K + \alpha)$ times smaller than the probability that the pixel is added to an existing adjacent region. Although this potential depends on the entire image, it is easy to compute and requires only that the number of regions be tracked.

The second potential function encourages regular object shapes. A class of "regularity" cliques are defined so that each member of the class consists of a central pixel and a "ring" of surrounding pixels. The diameter of the ring is chosen to reflect the diameters of objects in the image.

Denote the class of cliques with diameter a as C_a cliques. The potential for each of the cliques is determined by the number of pixels in the outside ring that have the same region identifier as the center pixel. Recall that in a Gibbs distribution, positive potentials correspond to high-energy, low probability states. If at least τ pixels have the same identifier, then the clique is assigned potential zero; otherwise, it is assigned a potential of $\phi(\tau$ - number of pixels with matching labels). The clique is penalized if it does not have at least τ matching pixels in the outer ring. Each pixel in the image is contained in 5 C_1 cliques, 9 C_2 cliques, and so on. "Because the number of like neighbors can be stored and updated when the region identifiers change, this type of regularity clique is not only geometrically reasonable, but computationally manageable" (Johnson et al. 1993).

The final potential type prevents regions from splitting into two disconnected parts. Suppose that there is a region with a dumbbell shape, and that the connecting bar is one pixel wide. If a pixel in the bar should change regions, then the dumbbell would be split into two distinct regions, and all the region identifiers in one half of the dumbbell would have to change. This violates the Markovian property of the Gibbs distribution, so pixels in "bars" are not permitted to change. In practice, this is accomplished "by assigning infinite potentials to changes in region identifiers that result in disconnected regions" (Johnson et al. 1993). The infinite potential functions violate the positivity condition of the Hammersley-Clifford theorem (Besag 1974); however, "in this case an aperiodic irreducible Markov chain with equilibrium distribution and transition probabilities given by the implied Gibbs conditional distributions can be constructed using the Metropolis algorithm."

These three potentials capture general beliefs about the shapes of regions in the image. However, specific beliefs about the location, size, and shape of regions can also be modeled. Assume that a source of structural prior information is available (e.g., a standard anatomical atlas) and that the template has been segmented. Negative "pseudo-potentials" are assigned to pixels in the template where regions are encouraged, and positive pseudo-potentials are assigned where they are discouraged. The potential of a region in the reconstructed image is calculated by summing over the pseudo-potentials of its pixels. While this does incorporate structural prior information, it is not a flexible way of modeling template information, and it requires substantial computation when implemented in three dimensions.

2.2.3 Deformation Models

In situations where landmark data is available, a natural approach to the modeling of deformation is to model how the landmarks move from one image to the next. The methods described by Lele (1989) and summarized in Section 2.1 try to choose a single "best" transformation. Elastic deformation or *deformable template* models provide a broad class of alternatives.

One deformable template approach is to parameterize a warping to be applied to n-dimensional space and assign prior distributions to the parameters controlling the warp. Hanson (1993), for example, parameterizes the coordinate warping as a polynomial and assigns probability to the warp using a Gibbs distribution with $p(warp) = \exp(-(total\ strain\ energy))$. This encourages configurations that minimize the amount of twisting and stretching required by the warp.

Suppose that one takes the limiting case, where every pixel in a continuous image is taken to be a landmark. This approach to deformation models is examined by Amit et al. (1991) and Grenander and Miller (1994). Prior knowledge about the true image is contained in a template and in a distribution about how that template is warped. A two-dimensional Gaussian process is specified on the distance that each point in the template moves. The covariance function of the Gaussian process is chosen to provide continuous sample paths and computational convenience.

Deformation models are conceptually elegant, as they provide a direct interpretation of the observed image. One knows what part of the template was mapped to what part of the observed image. Amit et al. (1991) model human hands, and one knows that fingertips are mapped to fingertips. These models

can also provide a way to study variability in a family of images, perhaps classifying subgroups or identifying abnormalities (Amit and Kong 1994).

However, there are a variety of problems with deformation models. The models typically encourage smooth deformations, which can blur anomalies or extra regions. It is very difficult to take into account different prior beliefs about small-scale and large-scale features. Indeed, the deformations being used are generic and do not depend on the template being used (Amit and Kong 1994). In the Amit et al. (1991) model, the posterior samples are expensive to compute and cannot easily be generalized to more than two dimensions. The Grenander and Miller (1994) model requires calculation on a 4000 processor parallel computer. Further, the criterion that is used to match the template to an image is minimizing the mean square difference in intensity over all pixels. There is no guarantee that landmarks or specific points of interest are matched with any precision (Amit and Kong 1994).

Amit and Kong (1994) address some of these problems by proposing a computationally efficient template matching method "consisting of a graphical model of landmark points which describes their planar arrangement, together with local operators which identify candidates for the various landmarks in the data image." Their method is potentially useful for the automatic identification of landmarks, and has many features in common with the model proposed in Chapter 5. Both models employ a MRF with translation, rotation, and scale invariant potential functions to model differences in shape between configurations of points and a template. Amit and Kong (1994) use cliques with three points and potential functions of the form:

$$V(x_1, x_2, x_3) \quad = \quad (\log(\|x_2 - x_3\|/\|x_1 - x_3\|) - \log(\|t_2 - t_3\|/\|t_1 - t_3\|))^2 +$$

$$(\log(\|x_1 - x_3\|/\|x_1 - x_2\|) - \log(\|t_1 - t_3\|/\|t_1 - t_2\|))^2 +$$
$$(\log(\|x_2 - x_3\|/\|x_1 - x_2\|) - \log(\|t_2 - t_3\|/\|t_1 - t_2\|))^2$$

where t_1, t_2, and t_3 are the points corresponding to x_1, x_2, and x_3 in the template. They use features to identify candidate landmarks and then to calculate an optimal template warping, while the model in Chapter 5 uses features in the image to develop a likelihood function and to calculate posterior probabilities.

2.2.4 Boundary Models

Boundaries provide a natural way to describe the shape of an object. Many methods for boundary description have been proposed; details can be found in Ballard and Brown (1982) and Gonzales and Woods (1992). Boundary models can be seen as another approach to deformable templates. These models use a template to parameterize object boundaries and then assign probabilities that describe appropriate deformations of the boundaries. Examples can be found in Ripley and Sutherland (1990), Staib and Duncan (1992), Cootes et al. (1993), Grenander and Manbeck (1993), and Phillips and Smith (1994).

There are two basic frameworks that are used to model boundaries and boundary deformation. The first approach uses multi-purpose boundary descriptors, such as b-splines, polylines, or Fourier descriptors, and assigns prior distributions to these coefficients that model "allowable" deformations. For example, Staib and Duncan (1992) use an elliptical Fourier decomposition of the boundary and put a multivariate normal distribution on a truncated set of the Fourier coefficients. In general, multi-purpose descriptors make prior specification difficult, as the coefficients do not necessarily correspond to identifiable features of the shape. It is especially difficult to incorporate multi-scale

information within this framework.

The second approach to modeling boundaries is to create a specialized template for the object that needs to be deformed. Again, the template must be parameterized so that distributions can be assigned that allow appropriate deformations. Grenander and Manbeck (1993), for example, develop a model to automatically detect defects in potatoes. They begin by modeling the outer contour of a prototype potato as a cyclic graph defined by an ordered collection of n vectors. The entire contour may be rotated or scaled, and each vector may be individually rotated. Probabilities are assigned to the allowed deformations by letting (1) the overall rotation be uniform on $[0, 2\pi)$, (2) the overall scaling have a lognormal density, and (3) the individual rotation angles have a multivariate von Mises density. The parameters of these distributions are estimated from training data.

Phillips and Smith (1994) consider models for the human face. Separate templates are developed for the eyes, nose, mouth, and eyebrows. By using only 50–100 parameters, they achieve a substantial reduction in the complexity of the data represented. They propose a hierarchical framework with the following levels to relate the templates.

1. Find the boundary for the head.

2. Find the boundary for the face, inside the head boundary.

3. Determine the spatial arrangement of features within the face boundary.

4. Determine the scaling of the features, given their relative positions.

5. Determine the shapes of individual features, given their scale and location.

This sort of hierarchical framework is central to the approach for modeling general shape and scene deformation that is developed in subsequent chapters.

2.2.5 Differential Geometry Models

Another class of priors is based on differential geometry and multi-scale descriptions of object shape. It has been recognized for about 25 years that multi-scale approaches can improve image analysis. Various representations that incorporate scale and resolution have been proposed, including pyramids, wavelets, and multi-grid methods. One of the most promising representations, *scale space*, has been developed recently in the computer vision literature. Jackway (1993) and Lindeberg (1994) contain a summary of the recent research in this area.

Scale space was developed to model the processing at the front end of the visual system. Its basic premise is that the visual system must initially be able to handle image structure at all scales and resolutions (Lindeberg 1994). To accomplish this, the image to be analyzed is embedded in a one-parameter family of derived images, a scale space, with resolution as the parameter. Specific details about the construction of scale spaces are given in Chapter 3.

Llacer et al. (1992) create a scale space by applying a "Laplacian of Gaussian" filter, also called the *Marr-Hildreth* filter, to a signal. The Laplacian is defined as the trace of the matrix of second derivatives. The standard deviation of the Gaussian indexes scale. The zero-crossings images filtered in this way are often used to locate edges. The prior proposed in Llacer et al. (1992) gives low probability to images with zero-crossings that are a large distance away from those in a template; in other words, lower probability is given to

images with edges that are far from those in the template. This prior is not very successful even in one-dimensional images.

2.3 Summary

This chapter has discussed two types of statistical models for images and landmarks: those modeling landmarks and their configurations, and those modeling image intensities. The image intensity models fall into five classes, which include Markov random fields, region-based models, deformable templates, boundary models, and differential geometry models. Each have been used, with varying degrees of success, to model template information. The model that is developed in Chapter 5 incorporates the hierarchical specification used in Phillips and Smith's (1994) boundary model and the multi-scale ideas from the differential geometry models to build a flexible class of priors on landmarks and shapes. These models draw heavily on MRF models and deformable template models like that of Amit and Kong (1994).

Chapter 3

Multi-Scale Image Analysis

As introduced in Chapter 2, scale space was developed to model the processing at the front end of the visual system. The central idea is to create an image representation that can handle image structure at all scales and resolutions simultaneously. This is accomplished by embedding an image in a one-parameter family of derived images, with resolution or "scale" as the parameter. This representation is not "data-reducing" like a wavelet decomposition, but creates a highly redundant representation of the image that can be used to efficiently make calculations about features at multiple scales. The scale-space framework facilitates the incorporation of multi-scale information into a prior distribution.

The literature distinguishes *multi-resolution* image processing from *multi-scale* image processing. Multi-resolution image processing implies a reduction in sampling rate as the scale becomes coarser, and is strongly associated with the "pyramid" data structure. More information about multi-resolution processing can be found in Rosenfeld (1984).

At first, the task of creating a "multi-scale" representation of image data seems somewhat arbitrary. Considered more carefully, however, it is clear

that when constructing a scale-space representation it is critically important that the transformation from fine scale to coarse scale actually represent a simplification of the data, so that fine-scale features vanish monotonically with increasing scale. If new artificial structures could appear at coarse scales, it would be impossible to determine whether these structures arose from finer scale features or by accident—for example, by the amplification of noise.

Two broad classes of multi-scale image processing have been developed: the first uses *Gaussian* scale spaces, and the second *morphological* scale spaces. In a Gaussian scale space, no new level curves are created as scale increases; equivalently

> If for some scale level σ_0 a point $\mathbf{x_0}$ is a local maximum for the scale-space representation at that level (regarded as a function of the space coordinates only) then its value must not increase when the scale parameter increases (Lindeberg 1994, p. 103).

In morphological scale space, no new local extrema appear in the smoothed signal as scale increases (Jackway 1992).

3.1 Gaussian Scale Spaces

Witkin (1983) introduced the idea of scale space for continuous one-dimensional signals. Given a signal $f : \mathbf{R} \rightarrow \mathbf{R}$, the Gaussian scale-space image $L : \mathbf{R} \times \mathbf{R}^+ \rightarrow \mathbf{R}$ is defined so that the representation at "zero-scale" is the original signal,

$$L(x, 0) = f(x)$$

and the representation at coarser scales is given by the convolution of the signal with the Gaussian probability density function, G, with mean 0 and standard

deviation σ

$$L(x,\sigma) = f(x) * G(x;0,\sigma) = \int_{-\infty}^{\infty} f(\alpha)G(x;\alpha,\sigma)d\alpha$$

Witkin observed that the number of zero-crossings of the second derivative of a one-dimensional signal decreases monotonically with scale. Yuille and Poggio (1986) extend this result to any differential operator that commutes with the diffusion equation; specifically, in one dimension, this property holds for derivatives of arbitrary order. Since the local extrema of a signal correspond to the zero-crossings of its first derivative, the number of local extrema decreases monotonically with scale.

In more than one dimension, however, there is no non-trivial linear shift-invariant (convolution) kernel that never introduces new local extrema (Lifshitz and Pizer 1990). To generalize Gaussian scale space to more than one dimension, a different "monotonically decreasing feature" must be found. Koenderink (1984) derives a multi-dimensional Gaussian scale space using the property of *causality*, which is meant to capture the idea that "any feature at a coarse level of resolution is required to possess a (not necessarily unique) 'cause' at a fine level of resolution, although the reverse need not be true" (Koenderink 1984). As developed by Koenderink, the causality property implies that new level surfaces are not created as the scale parameter is increased, i.e., that local extrema are not enhanced and do not "pop up out of nowhere" (Lindeberg 1994).

If causality is combined with a prohibition of space-variant blurring, it can be shown that the derived family of images must satisfy the diffusion equation, with the initial condition that the derived image at scale zero is the initial

image. This is equivalent to

$$\begin{aligned} \frac{\delta}{\delta\sigma}L(\mathbf{x},\sigma) &= \sigma(\frac{\delta^2}{\delta x_1}+\ldots+\frac{\delta^2}{\delta x_n})L(\mathbf{x},\sigma), \qquad \sigma>0 \\ L(\mathbf{x},0) &= f(x) \end{aligned}$$

These equations are satisfied when the initial image is convolved with the (multivariate) Gaussian density function with mean 0 and variance-covariance matrix $\sigma^2 I$ or one of its derivatives. Scale is taken to be σ, the standard deviation of the Gaussian. If the image is thought of as an initial heat distribution, the scale space shows the heat distribution over time as diffusion occurs in a homogeneous medium.

Figure 3.1 shows a one-dimensional signal and slices from its Gaussian scale space at increasing scales. Notice that small-scale features are suppressed as the scale (σ) increases.

Other derivations of Gaussian blurring for the creation of a scale space are given in Lindeberg (1994). Whitaker and Pizer (1993) explore the consequences of allowing space-variant blurring. Lindeberg (1994) shows that to maintain the desirable properties of Gaussian scale spaces in discrete images, Gaussian blurring should be replaced with discrete convolution with $T(n;\sigma) = \exp(-\sigma)I_n(\sigma)$, where I_n are the modified Bessel functions of integer order.

3.2 Morphological Scale Spaces

One of the axioms used to derive Gaussian scale spaces is that all representations should be generated by convolutions of the original image with a kernel (linear shift-invariant smoothing) (Lindeberg 1990). If the assumption of linearity is relaxed, then it can be shown that another class of filters, morpholog-

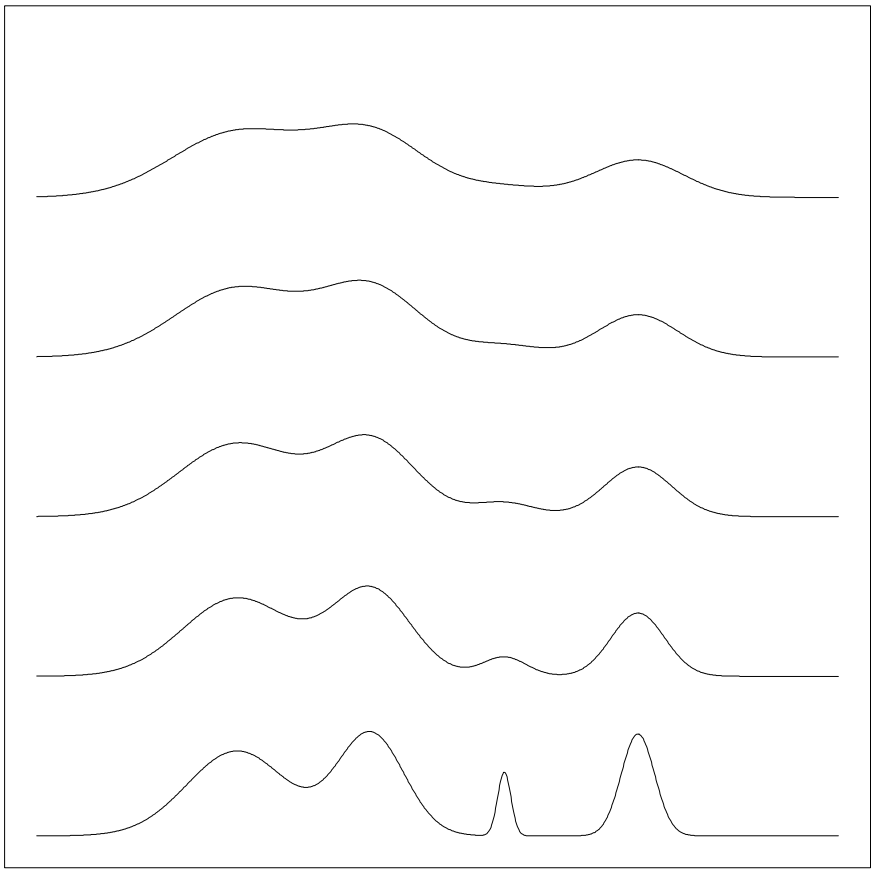


Figure 3.1: Gaussian Scale Space

ical filters, creates scale spaces with many of the same properties as Gaussian filters.

Morphological filters can be understood intuitively as follows. Think about a two-dimensional image, $f(x, y)$, as a surface in three-space $S = \{(x, y, z) : z = f(x, y)\}$. Take a ball of radius r, roll it over the surface S, and keep track of the surface traced out by the center of the ball. Intuitively, this new surface is smoother than the original, and the larger the radius of the ball, the smoother the "filtered" surface. Define smoothing by a ball of negative radius as rolling the corresponding positive radius ball under the surface.

Consider what happens at a hill (local maxima) in S when smoothing with a positive radius ball. If the ball touches the top of the hill, the hill appears at the same location in the output. If the ball is prevented from touching the top of the hill by surrounding hills, then the hill vanishes and cannot reappear when smoothing occurs with a larger radius ball. Formally, the number of local maxima is a monotonically decreasing function of r. By symmetry, the same property holds for local minima when smoothing with a ball of negative radius.

Smoothing by "rolling balls" can be formalized using ideas from gray-scale morphology. Smoothing by a positive radius ball corresponds to *dilation*; mathematically, the dilation of a signal f by a structuring element h is given as

$$f \oplus h = max_{s,t}\{f(s, t) - h(s - x, t - y)\}$$

where the functions f and h are taken to be negative infinity outside their domains. Smoothing by a negative radius ball corresponds to *erosion*; mathematically, the erosion of a signal f by a structuring element h is given as

$$f \ominus h = min_{s,t}\{f(s, t) + h(-(s - x), -(t - y))\}$$

where the functions f and h are taken to be negative infinity outside their domains. Notice that dilation and erosion are dual operators, i.e., $f \oplus h = -(f \ominus -h)$.

The multi-scale dilation-erosion of a signal f by a structuring element h is defined by Jackway (1993) as

$$(f * h_\sigma)(\mathbf{x}) = \begin{cases} (f \oplus h_\sigma)(\mathbf{x}) & \text{if } \sigma > 0 \\ f(\mathbf{x}) & \text{if } \sigma = 0 \\ (f \ominus h_\sigma)(\mathbf{x}) & \text{if } \sigma < 0 \end{cases}$$

where h_σ denotes a multi-scale structuring element, which in the case of the spherical structuring element is a ball of radius σ.

While the intuitive development has been given in terms of spherical structuring elements, the definition of multi-scale dilation-erosion holds for any structuring function $h_\sigma(\mathbf{x})$. A particularly useful class of structuring elements are given by the parabolic functions, $h_\sigma(\mathbf{x}) = \frac{1}{\sigma^2}\mathbf{x}'\mathbf{x}$, also called *rotationally symmetric quadratic structuring functions* (QSFs). The rotationally symmetric QSFs have many of the desirable properties of the Gaussian kernel, including shift invariance, separability (filtering can be decomposed into one-dimensional dilations), and closure with respect to dilation (if $q_A = x'Ax$, then $q_A \oplus q_B = q_{(A^{-1}+B^{-1})^{-1}}$). In addition, the scale space created using multi-scale dilation-erosion with rotationally symmetric QSFs displays "causality," as local extrema are monotonically decreasing in σ (Jackway 1992; van den Boomgaard 1992).

Figure 3.2 shows a one-dimensional signal (bold) and slices from its morphological scale space at several scales. Morphological scale space distinguishes between objects ("bright" or "white blobs") and background ("dark" or "black blobs"). "In effect the morphological scale space consists of two (tightly linked)

scale spaces: one for the object structure and one for the background structure" (van den Boomgaard 1992). As with the Gaussian scale space, small-scale features are suppressed as the magnitude of the scale (σ) increases.

3.3 Scale-Space Metrics

Scale space can be considered to be an $(n+1)$-dimensional space, where n dimensions are "space" and one is "scale." What is the appropriate way to measure distances within this space? Intuitively, the scale dimension is not commensurate with the spatial dimensions. In the context of Gaussian scale spaces, with $\sigma > 0$, Eberly (1994) suggests that measurements in scale space should be rotation, translation, and zoom invariant, where *zooming* is uniform magnification of both the space and scale variables. To accomplish this, an appropriate metric for scale space is

$$dp^2 = \frac{d\mathbf{x} \cdot d\mathbf{x}}{\sigma^2} + \frac{d\sigma^2}{\sigma^2} \tag{3.1}$$

where p denotes arc length, $\mathbf{x}$ denotes spatial coordinates, and σ denotes the scale coordinate. While this choice of metric is somewhat arbitrary, its invariance properties are useful in the development of the priors in Chapter 5.

With this metric, Gaussian scale space has hyperbolic geometry. The distance between any two points $(\mathbf{x_1}, \sigma_1)$ and $(\mathbf{x_2}, \sigma_2)$, with $\sigma_1 \leq \sigma_2$, is given as

$$d_{ss}[(\mathbf{x_1}, \sigma_1), (\mathbf{x_2}, \sigma_2)] = \log\left(\frac{\sigma_2}{\sigma_1} \frac{(1 + \sqrt{1 - (\rho\sigma_1)^2})}{(1 - \rho L + \sqrt{1 - (\rho\sigma_1)^2}}\right) \tag{3.2}$$

where

$$\rho = \frac{2L}{\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + L^2[L^2 + 2(\sigma_1^2 + \sigma_2^2)]}}$$

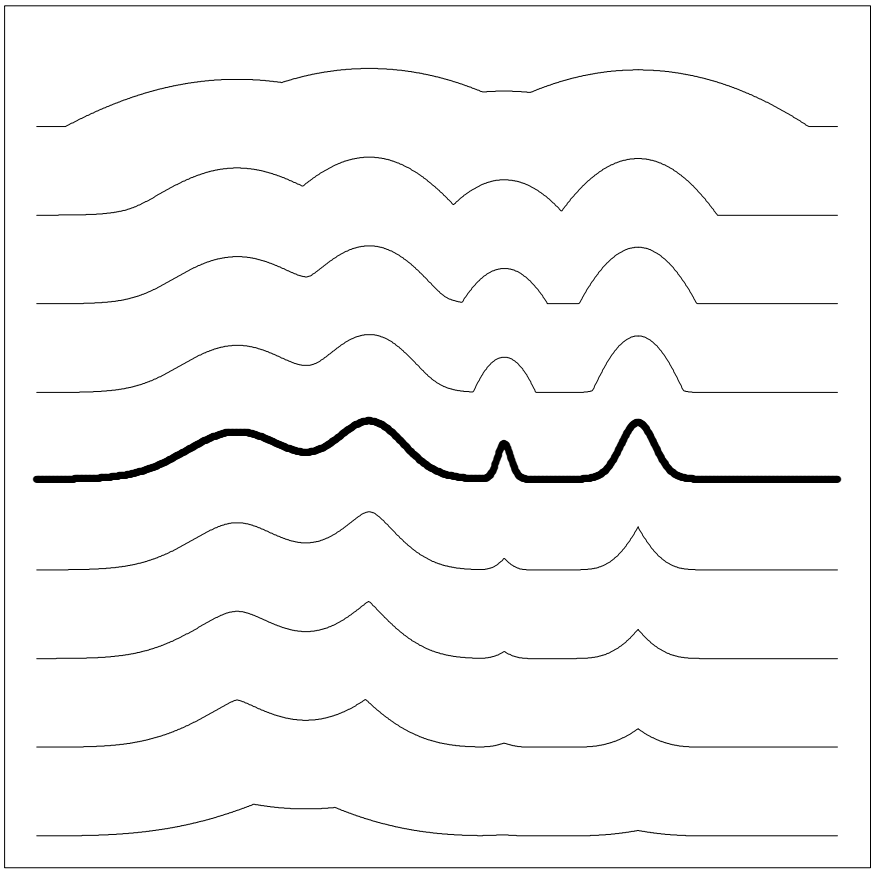


Figure 3.2: Morphological Scale Space

and $L = \|\mathbf{x_1} - \mathbf{x_2}\|$.

Consider the scale-space hypersphere of radius r centered at $(\mathbf{x_0}, \sigma_0)$. It can be shown that this is equivalent to the Euclidean hypersphere centered at $(\mathbf{x_0}, \sigma_0 \cosh(r))$ with radius $\sigma_0 \sinh(r)$. In other words, all of the points a distance r from $(\mathbf{x_0}, \sigma_0)$ lie on a circle centered at $(\mathbf{x_0}, \sigma_0 \cosh(r))$ with radius $\sigma_0 \sinh(r)$. Notice that these circles are not concentric as r changes, but that as $r \to 0, \cosh(r) \to 1$ and the center approaches $(\mathbf{x_0}, \sigma_0)$.

Consider the distance between two points $(x, \sigma) = (1, 1)$ and $(3, 1)$. In Euclidean space, the distance is 2. In scale space, the distance is $\log(\frac{1+\sqrt{2}}{\sqrt{2}-1}) \approx 1.762$. With this metric, the shortest distance between the two points is not a straight line, but is the arc of the circle centered at $(2, 0)$ with radius $\sqrt{2}$ that passes through $(2, \sqrt{2})$. This occurs because scale space is less dense at higher scales. These paths are illustrated in Figure 3.3.

One other consequence of this metric is required to develop the priors in Chapter 5. If the metric on scale-space were Euclidean, it would be a trivial task to find a point that is a given distance away from a specified point. Suppose that $x_0 = (1, 1, 1)$ and $x_1 = (2, 2, 2)$. Suppose also that $y_0 = (3, 2, 4)$, and it is necessary to locate y_1 to maintain the original relationship. Clearly, $y_1 = (4, 3, 5)$.

In scale-space, however, this calculation is more difficult. Regardless of the metric, distances are measured along curves with zero curvature. These zero-curvature curves are called *geodesics*. In Euclidean space, the differential equations determining the geodesic $\mathbf{x}(p)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and p is arc length, are given by the system of second order differential equations

$$\frac{\delta^2 \mathbf{x}}{\delta p^2} = 0$$

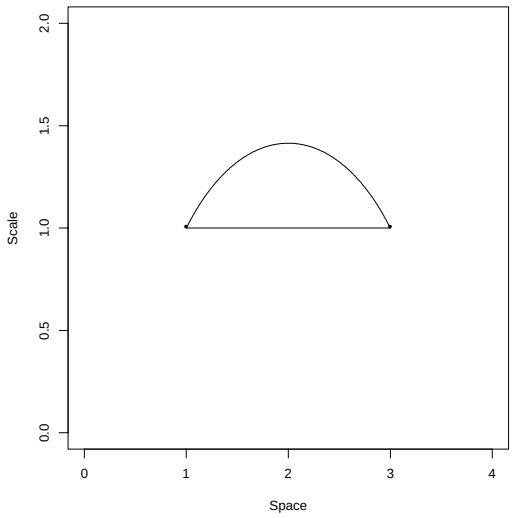


Figure 3.3: Shortest Distances

Integrating twice yields $x_i(p) = x_{i0} + x'_{i0}\ p$, where the initial position is $\mathbf{x_0}$ and the initial direction is $\mathbf{x'_0}$, where $\mathbf{x}' = \frac{\delta \mathbf{x}}{\delta p}$ and $(\mathbf{x_0}')^T(\mathbf{x_0}') = 1$.

In scale-space, however, using the metric given in (3.1), the differential equations specifying geodesics are (Eberly 1994)

$$\begin{aligned} \frac{d^2\mathbf{x}}{dp^2} &= \frac{2}{\sigma}\frac{d\sigma}{dp}\frac{d\mathbf{x}}{dp} \\ \frac{d^2\sigma}{dp^2} &= \frac{1}{\sigma}[(\frac{d\sigma}{dp})^2 - |\frac{d\mathbf{x}}{dt}|^2] \end{aligned}$$

These equations can be integrated to

$$\begin{aligned} \mathbf{x}(p) &= \mathbf{c} + r\ \tanh(\mathrm{p})\ \mathbf{u} \\ \sigma(p) &= r\ \mathrm{sech(p)} \end{aligned} \tag{3.3}$$

These equations specify a circle centered at $(\mathbf{c}, 0)$ with radius r and "north

pole" at $(\mathbf{c}, r)$ that lives in the plane spanned by $(\mathbf{0}, 1)$ and $(\mathbf{u}, 0)$. All of the constants, $\mathbf{c}$, r, and $\mathbf{u}$, are determined either by initial value data (The geodesic starts at $(\mathbf{x_0}, \sigma_0)$ and moves in the direction $(\mathbf{x_0}', \sigma_0')$ satisfying $[(\mathbf{x_0}')^T(\mathbf{x_0}') + (\sigma_0')^2]/\sigma_0^2 = 1.$) or by boundary value data (The geodesic starts at $(\mathbf{x_0}, \sigma_0)$ and ends at $(\mathbf{x_1}, \sigma_1)$.).

The relationships of the constants to the initial value data are as follows (Eberly, personal communication):

$$\begin{aligned} \mathbf{u} &= \frac{\mathbf{x_0}'}{|\mathbf{x_0}'|} \\ r &= \frac{\sigma_0^2}{|\mathbf{x_0}'|} \\ \mathbf{c} &= \mathbf{x_0} + r\ (\sqrt{1 - (\frac{\sigma_0}{r})^2})\ \mathbf{u} \end{aligned}$$

The relationships of the constants to the boundary value data are as follows:

$$\begin{aligned} L &= \|\mathbf{x_1} - \mathbf{x_0}\| \\ \mathbf{u} &= \frac{(\mathbf{x_1} - \mathbf{x_0})}{L} \\ r &= \frac{\sqrt{(\sigma_1^2 - \sigma_0^2)^2 + L^2[L^2 + 2(\sigma_1^2 + \sigma_0^2)]}}{2L} \\ \mathbf{c} &= \mathbf{x_0} + r\ (\sqrt{1 - (\frac{\sigma_0}{r})^2})\ \mathbf{u} \end{aligned}$$

The problem that needs to be solved is as follows. Given $(\mathbf{x_0}, \sigma_0)$ and $(\mathbf{x_1}, \sigma_1)$, find the geodesic distance d between them. Find the initial direction of the geodesic, D_0. Starting at $(\mathbf{y_0}, t_0)$, walk along a geodesic initially in the direction D_0 through a distance d to get to $(\mathbf{y_1}, t_1)$. Assume that $\sigma_0 \leq \sigma_1$.

From (3.3),

$$\begin{aligned}\frac{d\mathbf{x}}{dp} &= r\ \mathrm{sech(p)}^2\ \mathbf{u} = \frac{\sigma^2}{r}\mathbf{u} \\ \frac{d\sigma}{dp} &= -r\ \mathrm{sech(p)tanh(p)} = -\sigma(\sqrt{1-(\frac{\sigma}{r})^2})\end{aligned}$$

The initial direction is

$$D_0 = (x_0', \sigma_0') = (x'(0), \sigma'(0)) = (\frac{\sigma_0^2}{r}\mathbf{u}, -\sigma_0(\sqrt{1-(\frac{\sigma_0}{r})^2}))$$

where r, $\mathbf{c}$, and $\mathbf{u}$ are computed using the boundary value conditions. From the construction, $[(\mathbf{x_0}')^T(\mathbf{x_0}') + (\sigma_0')^2]/\sigma_0^2 = 1$.

The geodesic distance d can be calculated from (3.2). The initial direction for the "y-geodesic" is

$$(\mathbf{y_0}', t_0') = \frac{t_0}{\sigma_0}(\mathbf{x_0}', \sigma_0')$$

This vector is in the same direction as D_0, has initial point at $(\mathbf{y_0}, t_0)$, and has unit scale-space length $[(\mathbf{y_0'})^T(\mathbf{y_0'}) + (t_0')^2]/t_0^2 = 1$.

The equation of the circle containing the geodesic connecting (y_0, t_0) to (y_1, t_1) is

$$\begin{aligned}\mathbf{y}(p) &= \mathbf{c} + r\ \mathrm{tanh(p)}\ \mathbf{u} \\ t(p) &= r\ \mathrm{sech(p)}\end{aligned}$$

where

$$\begin{aligned}\mathbf{u} &= \frac{\mathbf{y_0}'}{|\mathbf{y_0'}|} \\ r &= \frac{t_0^2}{|\mathbf{y_0}'|} \\ \mathbf{c} &= \mathbf{y_0} + r\ (\sqrt{1-(\frac{t_0}{r})^2})\ \mathbf{u}\end{aligned}$$

The initial arc length parameter for $(\mathbf{y_0}, t_0)$, p_0, satisfies

$$\mathbf{y_0} = \mathbf{c} + r \tanh(p_0)\ \mathbf{u}$$

so

$$p_0 = \tanh^{-1}(-\sqrt{1 - (\frac{t_0}{r})^2})$$

The arc length parameter for $(\mathbf{y_1}, t_1)$ should be d units larger than p_0, so

$$p_1 = p_0 + d$$

The desired point is (Eberly, personal communication)

$$\begin{aligned} \mathbf{y_1} &= \mathbf{c} + r \tanh(p_1)\ \mathbf{u} \\ t_1 &= r \operatorname{sech}(p_1) \end{aligned}$$

In practice, most of these calculations do not need to be performed. Assume that $(\mathbf{y_0}, t_0)$ and $(\mathbf{y_1}, t_1)$ are in the same spatial and scale relationship as $(\mathbf{x_0}, \sigma_0)$ and $(\mathbf{x_1}, \sigma_1)$. Then it can be shown that

$$\begin{aligned} (\mathbf{y_1}, t_1) - (\mathbf{y_0}, t_0) &= \frac{t_1}{\sigma_1}((\mathbf{x_1}, \sigma_1) - (\mathbf{x_0}, \sigma_0)) \\ &= \frac{t_0}{\sigma_0}((\mathbf{x_1}, \sigma_1) - (\mathbf{x_0}, \sigma_0)) \end{aligned} \tag{3.4}$$

3.4 Summary

Scale space has proven to be a powerful tool in computer vision research. It allows easy manipulation of multi-scale features and measurements of an image. Exploiting its properties allows multi-scale descriptions of a template and transparent handling of affine transformations—essential components of the classes of priors developed in Chapter 5.

Chapter 4

Describing Templates

The crucial first step in specifying priors using template information is the development of a multi-scale description of the shape and features of a template. A template description provides a "most likely" shape against which others can be compared and evaluated. It also identifies landmarks, features at landmarks, and spatial and scale relationships among landmarks. Because shape is defined as those properties of the figure that are unaltered by rotation, translation, and uniform scaling, it is important to choose feature descriptors that have these properties. In some situations, it may also be desirable to choose descriptors that are invariant to monotonic or affine intensity transformations.

4.1 Transformations and Invariances

Affine transformations preserve parallel lines and equally spaced points. They include translation, rotation, scaling, and shear. Let $(x, y) = (X(u, v), Y(u, v))$ be the *output pixels* and $(u, v) = (U(x, y), V(x, y))$ be the *input pixels*. X and Y are called the *forward mappings* (of input to output), and U and V are the *inverse mappings* (Wolberg, 1990). Affine transformations can be formulated

as follows.

$$\begin{aligned} X(u,v) &= a_{11}u + a_{21}v + a_{31} \\ Y(u,v) &= a_{12}u + a_{22}v + a_{32} \end{aligned}$$

or equivalently

$$(x,y,1) = (u,v,1)\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} = (u,v,1)A$$

Notice that the inverse of an affine transformation is itself affine.

$$(u,v,1) =$$

$$(x,y,1)\frac{1}{a_{11}a_{22} - a_{21}a_{12}}\begin{pmatrix} a_{22} & -a_{12} & 0 \\ -a_{21} & a_{11} & 0 \\ a_{21}a_{32} - a_{31}a_{22} & a_{31}a_{12} - a_{11}a_{32} & a_{11}a_{22} - a_{21}a_{12} \end{pmatrix}$$

Translation has a forward mapping specified by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ T_u & T_v & 1 \end{pmatrix}$$

Rotation has a forward mapping specified by

$$A = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Scale has a forward mapping specified by

$$A = \begin{pmatrix} S_u & 0 & 0 \\ 0 & S_v & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

A translation followed by a rotation and a scale change is given by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ T_u & T_v & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} S_u & 0 & 0 \\ 0 & S_v & 0 \\ 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} S_u \cos(\theta) & S_v \sin(\theta) & 0 \\ -S_u \sin(\theta) & S_v \cos(\theta) & 0 \\ S_u(T_u \cos(\theta) - T_v \sin(\theta)) & S_v(T_u \sin(\theta) + T_v \cos(\theta)) & 1 \end{pmatrix} \tag{4.1}$$

When translation and rotation invariance are posited for features and distances in scale space, the transformations are assumed to happen only in the spatial coordinates. Uniform scale changes, however, happen both in the spatial and scale coordinates. This transformation is called *zooming* to distinguish it from scale changes that occur in space only.

Eberly (1994) describes zoom invariance by saying, "If scale measurements are made for an object, the scale measurements for a magnified version of the object should be the magnified measurements for the original object." Suppose the feature of interest is intensity, and let $f(x)$ be the zero-scale image. The intensity at any point in scale space is given by

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(x-u)^2) f(u) du$$

If the image is zoomed by a factor of μ, the intensity at any point in scale space is given by

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(x-u)^2) f(\mu u) du =$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\mu\sigma} \exp(-\frac{1}{2(\mu\sigma)^2}(\mu x - v)^2) f(v) dv$$

This is equivalent to evaluating the scale space of the initial image at $(\mu x, \mu\sigma)$.

This property is useful for measuring features in images. If a candidate image is a zoomed version of the template, and the features are zoom invariant, then the features from the template are found in the candidate image at the zoomed coordinates. The feature descriptors developed in this chapter are invariant to rotation and translation, and often to intensity changes and zooming. The priors developed in Chapter 5 share these invariances; consequently, the models that are developed can be applied regardless of the orientation, location, or size of an object within an image.

4.2 Medial Descriptions

One important class of feature descriptors relies on extracting medial or "middleness" structure. Blum and Nagel (1978) developed the *symmetric axis transform* (SAT) as a medial descriptor for binary images. The two components of the SAT are the *symmetric* or *medial axis* and the *radius function*. The medial axis consists of the centers of all circles within the object that are tangent to the object boundary at two or more points. The radius function is defined at each point along the medial axis as the radius of the tangent circle.

The SAT and the object boundary are equivalent descriptors; however, some properties (like object width and curvature) are easier to describe using the SAT (Blum and Nagel 1978). The medial axis itself is a connected structure with three types of points: normal, branch, and end. A *normal* point is one whose underlying circle touches the figure in exactly two separate contiguous sets; a *branch* point's circle touches in three or more separate contiguous sets; an *end* point's circle touches in one contiguous set. Objects can be partitioned using *simplified segments*, which consist of contiguous sets of normal points

terminated by a branch or end point. By further classifying simplified segments, it is possible to compare shapes independent of width, orientation, and object curvature.

There are two problems with the SAT. The first is that it is extremely sensitive to small distortions along the boundary. Small protrusions or indentations can create new axis branches. The second problem is that an object must be segmented before its SAT can be calculated. These problems can be addressed by another type of shape analysis, called *structural image description*, where features representative of objects in an image are extracted from the original intensity values.

The *multi-scale medial axis*, or *core* (Fritsch 1993; Burbeck and Pizer 1994; Pizer et al. 1994), models medial structure directly from image intensity values. The idea is to apply a filter to an image to create a function on scale space, $M(\mathbf{x}, \sigma)$, that describes how medial a point $\mathbf{x}$ is relative to object boundaries (Eberly 1994). The core is defined as a ridge of the medialness function in scale space.

A variety of medialness functions have been proposed, each tailored to specific imaging situations. The medialness functions all share the properties of translation, rotation, and zoom invariance. Fritsch (1993) uses $M(\mathbf{x}, \sigma) = -\sigma^2 \nabla^2 L(\mathbf{x}, \sigma)$, where ∇^2 denotes the Laplacian or trace of the matrix of second derivatives. This function is useful "for extracting anatomic objects with non-parallel sides, approximately uniform interiors, edges of fixed contrast polarity, and possibly low signal to noise ratio" (Pizer et al. 1994). Another proposal is $M(\mathbf{x}, \sigma) = -\sigma^2 \lambda$, where λ is the largest magnitude eigenvalue of the matrix of second derivatives of $L(\mathbf{x}, \sigma)$. This medialness function is "particularly

sensitive to objects with parallel sides and uniform interior intensity, such as a blood vessel" (Pizer et al. 1994).

Ridges are a generalization of local maxima. For a real-valued function $f(\mathbf{x})$, a local maximum for f occurs at $\mathbf{x}$ if $Df(\mathbf{x}) = 0$ and $D^2f(\mathbf{x})$ is negative definite, where D denotes derivative. A ridge requires that these properties hold in a smaller number of directions. Let v_k denote the eigenvectors of $D^2f(\mathbf{x})$ and λ_k the eigenvalues, with $\lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$. The conditions for being a d-dimensional ridge point are $v_k \cdot Df(\mathbf{x}) = 0$ and $\lambda_k < 0$ for $k = 1, \ldots, n-d$.

Because cores make explicit use of scale, they are more robust than the SAT to small indentations or protrusions along object boundaries. Further, they do not require prior segmentation of the object. Fritsch (1993) discusses the use of cores for image registration. The cores of an object resemble a skeleton, and consequently may be useful in establishing hierarchical relationships among the parts of an object.

4.3 Differential Geometry Descriptions

Structural image description has also given rise to a set of descriptors based on differential geometry. These descriptors were introduced briefly in Section 2.2.5. Differential geometry descriptors are composed of linear or non-linear combinations of scale-space derivatives. For ease of notation, let $L(\mathbf{x}, \sigma) = f(\mathbf{x}) * G(\mathbf{x}; \mathbf{0}, \sigma^2 I)$ denote the Gaussian scale space of an image $f(\mathbf{x})$ and let derivatives be denoted as $\frac{\delta^2}{\delta x_i x_j} f(\mathbf{x}) = f_{x_i x_j}(\mathbf{x})$. It should be noted that a single partial derivative, e.g., $L_{x_i}(\mathbf{x}, \sigma)$, does not contain meaningful geometric shape information, because its value depends critically on the coordinate system. Instead, interest focuses on descriptors that are invariant with respect to rotation

and translation, and often with respect to scale and intensity transformations.

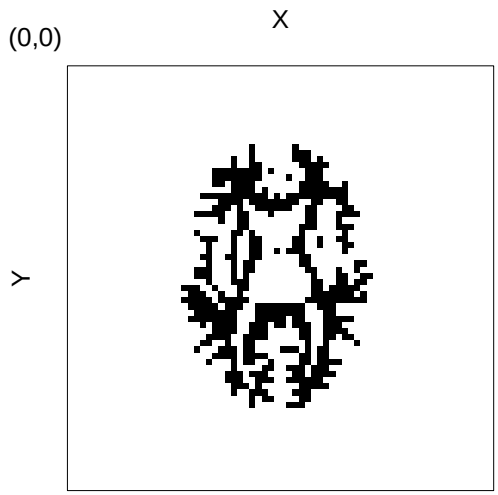
Figure 4.1 shows the white matter for the Hoffman brain phantom (HBP) (Hoffman et al. 1990) and the slice of the phantom's Gaussian scale space at scale 3.0. Figure 4.2 shows the first and second partial spatial derivatives at scale 3.0. The largest values of the derivatives are in white, the smallest in black.

4.3.1 Second Order Differential Invariants

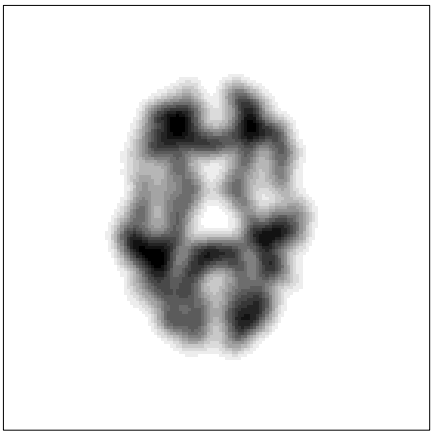
When one uses differential geometry descriptors, one is interested in describing local image properties. However, when working in scale space, the meaning of local depends on the scale where the descriptor is calculated. A "local" property at a large scale may depend on intensities from almost every pixel in the image.

Florack (1993) derives a set of five second order irreducible polynomial differential invariants. *Second order differential* means the expressions are composed of intensity, first derivatives, or second derivatives; *irreducible* means that the five functions are not polynomial combinations of one another; *invariant* means that the expressions do not depend on the coordinate system and are invariant under rotations and translations. Every second order polynomial differential invariant can be expressed through multiplication or addition of these invariants (Florack 1993). The invariants are (in terms of scale-space derivatives and assuming two spatial dimensions): L (intensity), $L_x^2 + L_y^2$ (squared gradient magnitude), $L_{xx} + L_{yy}$ (Laplacian), $L_{xx}^2 + 2L_{xy} + L_{yy}^2$ (deviation from flatness), and $L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy}$ (non-maximum suppression). Florack (1993) also derives complete sets of non-polynomial invariants.

Invariants are useful as descriptors because their values do not depend on



(a)



(b)

Figure 4.1: Original and Scale-Space HBP (a) White matter, scale 0; (b) White matter, scale 3.0

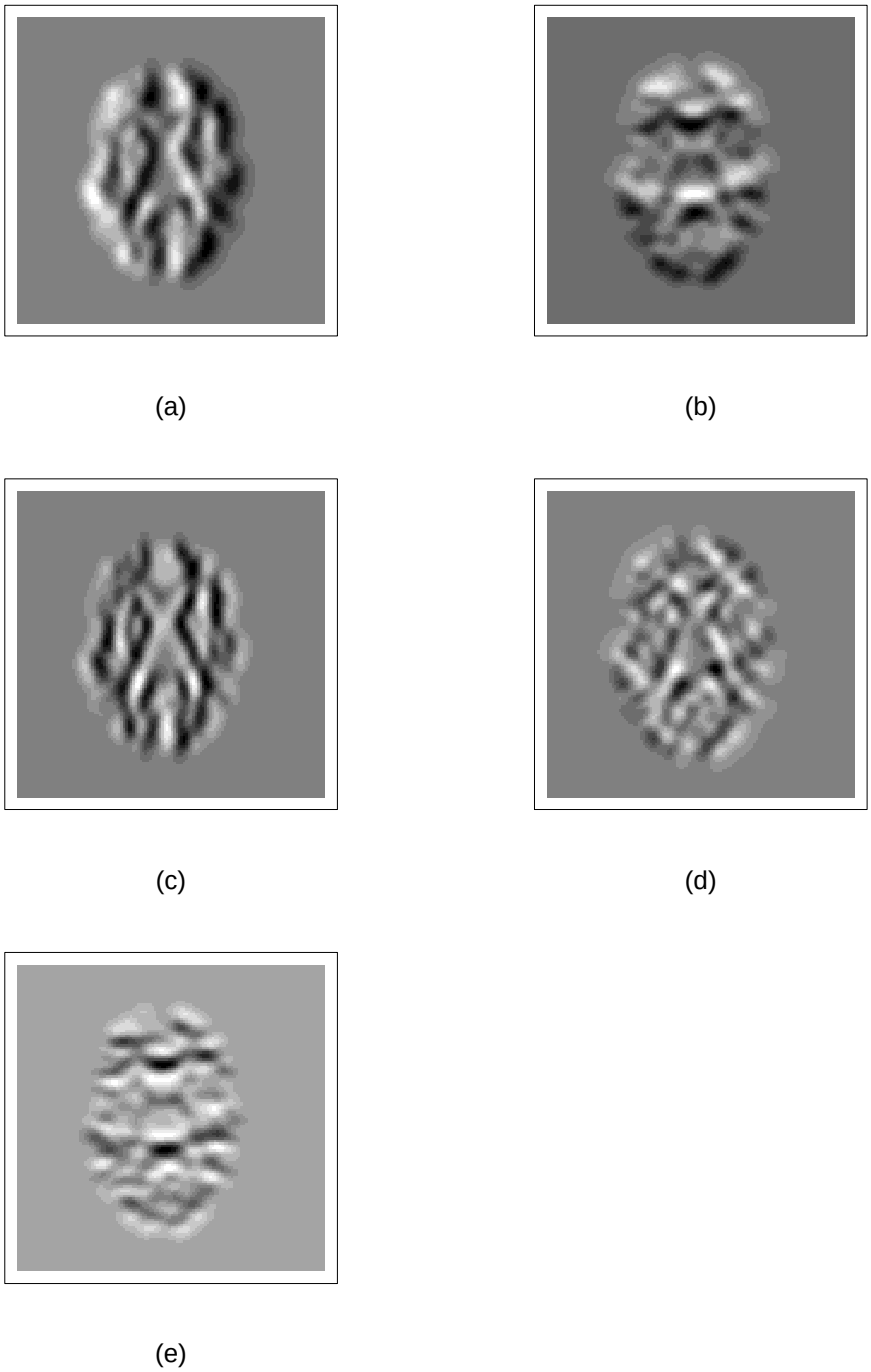


Figure 4.2: HBP Derivatives (a) X-derivative, scale 3.0; (b) Y-derivative, scale 3.0; (c) XX-derivative, scale 3.0; (d) XY-derivative, scale 3.0; (e) YY-derivative, scale 3.0

the coordinate system in which they are calculated. Most of the differential geometry descriptors that are described below are functions of this set of second order polynomial differential invariants.

4.3.2 Edge Detection

Finding boundaries and edges in images is a common task. If the regions in an image are described by boundaries, then edge finding is a natural precursor to segmentation. A simple edge descriptor is the squared gradient magnitude $L_x^2 + L_y^2$. Places in the image with high gradient magnitude correspond to places where intensities are changing sharply, which often indicates a change from one region to another. The squared gradient magnitude of the Hoffman phantom at scale 3.0 is given in Figure 4.3(a).

Another way to define edges in a continuous gray-level image is as the set of points where the gradient magnitude assumes a maximum in the gradient direction. This method is called *non-maximum suppression* (Canny 1986). Lindeberg (1994) formulates this method of edge-finding in terms of differential geometric descriptors and shows that "edges" should be located at the zero-crossings of

$$L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy} \tag{4.2}$$

where

$$L_x^3 L_{xxx} + 3L_x^2 L_y L_{xxy} + 3L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} < 0 \tag{4.3}$$

Notice that there is no need to estimate the gradient direction. Equations (4.2) and (4.3) are invariant with respect to translation and rotation, and, because one is looking for zero-crossings of (4.2), to uniform rescaling and affine intensity transformations as well. Figure 4.3(b) shows the non-maximum

suppression edge detector applied to the white matter of the Hoffman phantom at scale 3.0.

Zero crossings of the Laplacian have been used as edge-detectors (Marr and Hildreth 1980), but they do a poor job finding the edges of small indentations (Llacer et al. 1993). The Laplacian is invariant with respect to rotation and translation, and its zero-crossings are additionally invariant with respect to scale. The Laplacian of the Hoffman phantom at scale 3.0 is shown in Figure 4.3(c).

4.3.3 Level Curves

Level curves or *isophotes* are lines of equal intensity within an image. Examining the properties of isophotes is useful because isophotes are invariant under rotation, translation, and monotonic intensity transformations. Using only first and second derivatives, there are two isophote descriptors. The first is level curve, or isophote, curvature

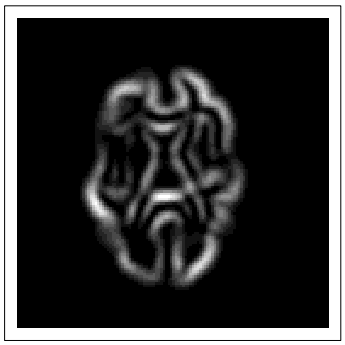
$$\frac{L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_y L_{xy}}{(L_x^2 + L_y^2)^{3/2}} \tag{4.4}$$

The second is flow line curvature,

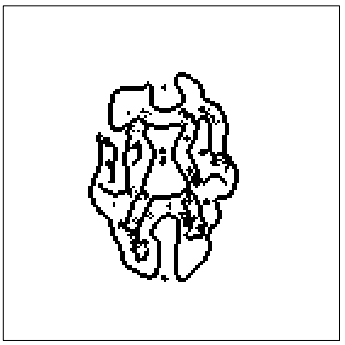
$$\frac{(L_x^2 - L_y^2)L_{xy} - L_x L_y (L_{yy} - L_{xx})}{(L_x^2 + L_y^2)^{3/2}}$$

where flow lines (also called streamlines or gradient integral curves) are the orthogonal trajectories of the isophotes.

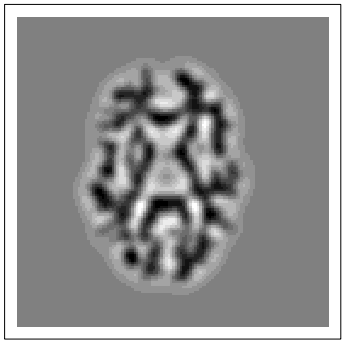
Figure 4.4 shows these two invariants calculated for the white matter of the Hoffman phantom at scale 3.0. Higher curvature corresponds to lighter gray levels. These features are not very useful in practice because their calculation is



(a)



(b)



(c)

Figure 4.3: HBP Edges and Boundaries (a) Squared Gradient Magnitude, scale 3.0; (b) Non-Maximum Suppression, scale 3.0; (c) Laplacian, scale 3.0

ill-conditioned. The denominator of these expressions is a power of the gradient magnitude, which goes to zero at local maxima and minima. The numerator should also go to zero at these points, resulting in an undetermined value. However, due to noise and numerical errors, what often happens is that very large positive or negative values appear at these points.

4.3.4 Corners and Junctions

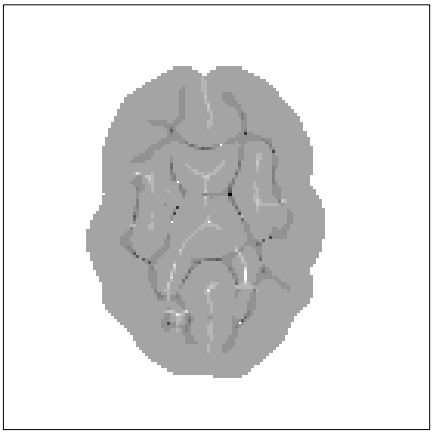
A descriptor that is commonly used for junction or corner detection is isophote curvature multiplied by gradient magnitude. Corners are typically characterized both by high curvature and by large gradient. To give a stronger response near edges, the isophote curvature is often multiplied by some power k of the gradient magnitude: often $k = 3$ is chosen so that the expression for the corner detector becomes

$$\tilde{\kappa} = L_y^2 L_{xx} - 2L_x L_y L_{xy} + L_x^2 L_{yy} \tag{4.5}$$

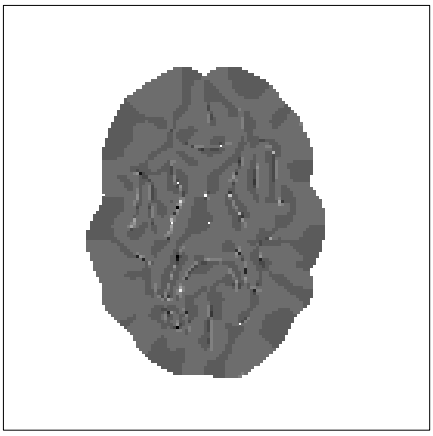
Notice that this is the numerator of (4.4). This expression is invariant with respect to rotation and translation. Further, if the x- and y-coordinate axes are rescaled by different factors s_x and s_y, $\tilde{\kappa}$ is multiplied by $s_x^2 s_y^2$. This means that any maximum of $\tilde{\kappa}$ remains a maximum under non-uniform scalings. "Corner candidates" are typically defined as local maxima or minima of $\tilde{\kappa}$. Figure 4.5 shows the corner detector calculated for the white matter of the Hoffman phantom at scale 3.0. Larger values have lighter gray levels.

4.3.5 Measures of Flatness

Other second order descriptors that are invariant with respect to rotation and translation include measures of "deviation from flatness." The "total deviation



(a)



(b)

Figure 4.4: HBP Curvatures (a) Isophote Curvature, scale 3.0; (b) Flow Line Curvature, scale 3.0

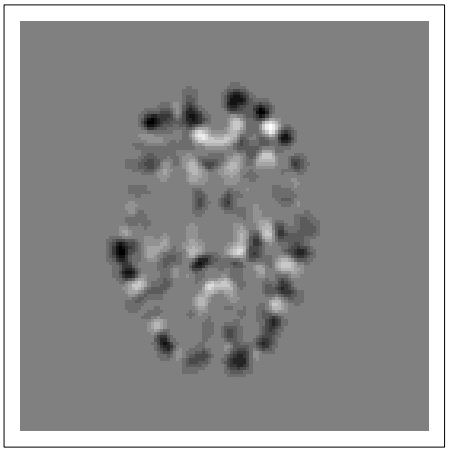


Figure 4.5: Corner Detector, scale 3.0

from flatness" corresponds to areas of the image where intensities are changing non-linearly. This descriptor is

$$L_{xx}^2 + 2L_{xy} + L_{yy}^2$$

The "maximal and minimal deviation from flatness" (ter Haar Romeny et al. 1991) are the eigenvalues of the second derivative matrix and have expressions

$$\frac{L_{xx} + L_{yy} \pm \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}}{2}$$

The eigenvectors corresponding to these eigenvalues are

$$(\frac{L_{xx} - L_{yy} \pm \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}}{2L_{xy}}, 1)$$

These eigenvectors can be useful in estimating rotation. The deviations from flatness are given for the Hoffman phantom at scale 3.0 in Figure 4.6.

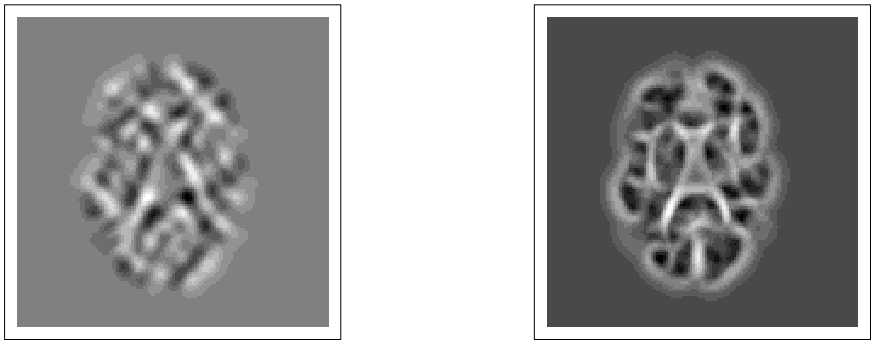
4.4 Choosing Features

As discussed above, there are a variety of features that can be used to describe characteristics of a template. It is important to examine what methods can be used to develop effective template descriptions by choosing both locations, features of interest, and geometrical relationships of interest.

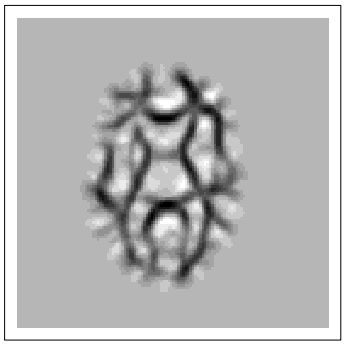
4.4.1 Previous Work

The previous work addressing these questions has focused primarily on pattern recognition as applied to image segmentation. Coggins and Huang (1993) develop the idea of *multi-scale geometric statistical pattern recognition*. Their goal is to be able to label points in an image based on distinctive geometric features. They apply various filters to the Gaussian scale space of an image to calculate differential geometric descriptors and create a *feature space*. Every point in the image is represented by a vector in the feature space that contains the value of multi-scale descriptors at that point. Training data is used to determine the normal distribution that best describes each segmentation class, and then the points $\mathbf{x}$ to be segmented are assigned to the class that minimizes the Mahalanobis distance between $\mathbf{x}$ and the class mean. This methodology is problematic because points in the same class do not tend to cluster in well-separated "blobs" in feature space, and assuming a normal distribution is typically a poor approximation.

Haring et al. (1993) address the same problem by using differential geometry



(a) (b)



(c)

Figure 4.6: HBP Flatness Measures (a) Total Deviation from Flatness, scale 3.0; (b) Maximum Deviation from Flatness, scale 3.0; (c) Minimum Deviation from Flatness, scale 3.0

shape descriptors to train a neural network to recognize clusters of features at various scales and hence to segment images. Although some success was achieved using this algorithm, it is difficult to draw general conclusions about either the distribution of features in feature space or appropriate combinations of features useful for describing templates.

4.4.2 Heuristics

In addition to the multi-scale statistical pattern recognition techniques, there are a variety of heuristic techniques for feature selection and template description that appear throughout the literature on multi-scale image analysis.

It is helpful if the feature of interest is invariant to translation, rotation, and scale changes, as these components are not involved in the definition of shape. In addition, features that are invariant to monotonic intensity changes are of special interest, as intensity is usually not of primary interest in a template. If the feature of interest is changing quickly at a particular location, the rapid change can also be quite useful in distinguishing that location. This is exploited in the specification of likelihood functions in Chapter 6.

To identify landmarks, Amit and Kong (1994) use local operators involving only rank relations between intensities in the neighborhood of the pixel. They found, for example, that local maxima appeared to be the most descriptive of the anatomy of the X-ray images they examined. Hierarchical relationships among landmarks were "introduced manually by connecting vertices that seemed to have an important geometric relation with each other."

Bookstein (1991) proposes three types of landmarks for use in biological morphometrics; however, with adaptation, his principles are useful for general

landmark selection. *Type 1* landmarks are located at the discrete juxtapositions of tissues, i.e., at points in space where three structures meet. Examples include the branching points of tree structures (for example, branching points of a multi-scale medial axis) or the bony sutures under the bridge of the human nose.

Type 2 landmarks are maxima of curvature or other local morphogenetic processes. Examples include the tips of protrusions and valleys of indentations; in three dimensions, one looks for bulges, dips, or saddle points. *Type 3* landmarks are extremal points. Examples include the endpoints of diameters, centroids, or medial axes. Notice that Bookstein's landmarks fit naturally with the medial description of templates, as points of interest correspond to branching and end points of medial descriptions.

Lindeberg (1994) proposes the following heuristic for identifying scales and locations of interest in an image.

> In absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data (p. 320).

If the input image is rescaled by a constant scaling factor, then the scale at which the maximum of many differential geometric descriptors is achieved is multiplied by the same factor. This selection criteria has also been used by Fritsch (1993) for identifying multi-scale medial axes (maxima over scale of the medialness function) and Pizer et al. (1994) for identifying cores.

Pizer et al. (1994) also propose the use of directed acyclic graphs for object description. They envision the nodes of the graph as occurring on cores and containing intensity statistics and boundary texture statistics; the arcs of the

graphs contain subfigure type (e.g., protrusion or indentation), location relative to a core, and angle relative to a boundary.

4.4.3 Specifying a Template Description

Clearly, there is no unique way to describe a template. The following techniques have proven useful in empirical work with the multi-scale priors for template information and are employed in subsequent chapters.

There are three components of a template description. The first identifies landmarks; the second identifies features of interest at the landmarks; the third specifies geometric relationships among the landmarks.

The steps needed to identify landmarks cannot be separated from those needed to identify features of interest in the template. The first step in the process is the identification of an "origin", which typically corresponds to the large-scale center of the template. The feature that identifies the origin is usually some measure of medialness. Next, n locations $(\mathbf{x_i}, \sigma_i)$ in scale space are chosen where there are features of interest. Locations of interest are typically at medial, corner, boundary, or junction points in the image and its background.

It has proven useful to start the identification of template landmarks by examining the cores of the objects in the image. In the example in Chapter 6, the two objects in the image are a brain ventricle and the brain itself. Identifying the cores provides a set of possible points of interest. Some of these points can be uniquely identified: for example, the ends of the brain core can be identified by local extrema in the corner detector (4.5). Other points are identified only in relation to the overall configuration of landmarks. For example, normal core points are found by identifying a ridge in medialness; the

"coreness" of any individual point depends on what features occur around it. Both types of points are useful in template description and prior specification.

In addition to cores, the corner detector and squared gradient magnitude are useful in suggesting points of interest. Local extrema of the corner detector correspond to junctions and core vertices; the squared gradient magnitude identifies likely boundary points. All of these feature descriptors must be "scaled" by multiplying by an appropriate power of scale to insure zoom invariance.

After identifying landmarks and features of interest, the "important geometric relations" among the points must be specified. I use several strategies for identifying these connections. The first is to establish a hierarchical branching structure that groups the locations of interest into "branches" that contain ordered sets of related features. This is intended to establish links among points whose relative locations are important to maintaining the shape of the template. Often, moving from the top of the branch to the bottom corresponds to moving from large-scale features to small-scale features. For example, one branch for a hand template might contain the large-scale center of the hand followed by the center of the index finger followed by the center of the index fingernail. Because boundaries tend to be well-localized in scale, I have found it useful to pair them with medial points in the branching structure, often using the boundaries as the ends of branches.

After specifying connections along a branch, links between branches must be included. For example, in the example in Chapter 6, the core for the brain is connected in several locations to the central core for the ventricle. Links between the boundary points at the ends of contiguous branches are helpful because the boundary points are not uniquely specified by the features.

A concrete example of a template built using these heuristics can be found in Chapter 6, where a distribution on landmarks is constructed using a brain ventricle template.

4.5 Summary

This chapter provides an overview of invariances, feature description, and methods for specifying template descriptions. Specifying a good multi-scale template description is a crucial first step in the development of a prior capturing template information. Because shape is defined as those properties of the figure that are unaltered by rotation, translation, and uniform scaling, it is important to choose feature descriptors that have these properties. Differential geometry descriptors offer a rich variety of feature descriptors that have these invariances. These descriptors are used in Chapter 6 to develop both prior distributions and likelihood functions.

Chapter 5

Multi-Scale Template Priors for Landmarks

This chapter develops a class of multi-scale prior distributions for the location of landmarks and objects within images. The priors incorporate ideas about landmarks from the statistical theory of shape and about neighborhoods and local conditional probability specification from MRF models for pixel intensities. The priors differ from standard applications of MRF models in that the sites in the fields represent image objects, and the random variables associated with the sites represent their locations. Another key feature of these priors is their ability to use information from a variety of resolutions. Because these types of priors allow the computation of posterior distributions of the locations of landmarks, they are inherently useful in such imaging tasks as automatic landmark identification, segmentation, and object recognition.

5.1 Motivation

Amit and Kong (1994) suggest that three problems must be addressed when developing deformable template models: the deformation model and its vari-

ability, the data term that drives the matching, and the computational tools and their limitations. These criteria have proven to be useful for the development of multi-scale priors on landmarks as well. The deformation model in this chapter is a MRF prior on the scale-space location of landmarks; the data term is a likelihood function that incorporates features at the landmarks; the computational tool is Markov chain Monte Carlo (MCMC), which has been widely used for calculating posterior distributions for models with MRF priors. The development of likelihood functions and MCMC methods are addressed in the context of an example in Chapter 6; this chapter focuses on the MRF prior.

Suppose that a template description like that proposed in Section 4.4.3 has been created. The prior uses only part of the information from the template description. Specifically, it requires the n scale-space locations and the neighborhood relationships between them. Information about the features at the landmarks is incorporated into the likelihood function.

The goal for specifying the prior is to create a probability distribution on sets of n scale-space locations that assigns collections of points describing figures "like" the template high probability and points describing figures "unlike" the template low probability. If a figure has the same shape as the template, it should be assigned the same probability as the template. In other words, translations, rotations, and zoomings of the template receive the same probability as the template.

A first attempt at developing such a prior might proceed as follows. Given a set of points $X = (\mathbf{x_1}, \ldots, \mathbf{x_n})$, translate X so that the mean of its spatial coordinates matches that of the template. Zoom X so that the sum of squares

difference of its spatial coordinates from their means is the same as the template. Rotate X so that it is in the same orientation as the template. After rotation, translation, and scale have been accounted for, calculate a measure of how different the shape of X is from the template.

The problem with this method can be illustrated by an example. Suppose that the template is a "normal" human hand and that the candidate image X is from a hand where the thumb is swollen to twice the normal size. If X is globally rescaled to the size of the template, the fingers will be too small, the thumb will be too large, and a large shape difference will be found. It would be more desirable to assess a single "penalty" for the thumb's size, but if the thumb is otherwise the right shape, to find no other shape differences. These types of "penalties" can be modeled by using a MRF model with appropriately chosen neighborhoods and cliques. Many elastic deformation models have implicit constraints that penalize large changes in the local lattice elements (Amit and Kong 1994); this is avoided using the MRF structure.

5.2 Sites, Random Variables, Neighborhoods, and Potential Functions

Following Geman and Geman (1984), to specify a MRF model, one must choose sites, random variables indexed by the sites, a neighborhood system over the sites, and potential functions on cliques of sites. The prior specified here uses the template description to specify both sites and a neighborhood structure. The sites are taken to be the nodes of the graph in the template description (i.e., the landmarks or places where features of interest are found). The neighborhood system is "nearest neighbor," meaning any node that can be reached

by traversing exactly one arc of the template description graph. Section 4.4.3 gives more details on the specification of template descriptions.

The random variables at the sites are not the features of interest, but rather the scale-space location of the landmark. This differs from the typical MRF models specified in image analysis. In Besag (1974) and all subsequent work, the locations of the sites are assumed to be known and are located at each pixel, while the random variables at the sites are pixel intensities or region labels.

The cliques used in the template priors depend on the precise structure of the template graph. In the the prior developed here, the only cliques with non-zero potential functions are those containing either one or two nodes. This, of course, can be generalized to allow more flexibility in the modeling of deformation.

In order to assign the same probability to any rotation, translation, or uniform rescaling of the template, it is necessary to have potential functions that are invariant to these transformations. One potential function that satisfies these requirements is given by Amit and Kong (1994) and is discussed in Chapter 2. I am interested in measuring distances in scale space rather than in Euclidean space; consequently, the potential functions I use employ scale-space distance rather than Euclidean distance.

The potential function for singleton cliques is intended primarily for cliques containing the scale coordinate of the scale-space location. The potential function has the form of the log of an inverse gamma distribution. (Because the potential functions are exponentiated in the Gibbs distribution, this is equivalent to adding independent inverse gamma distributions to the prior.)

The mean of the inverse gamma distribution for each scale is set as follows. For an image closely matching the template, calculate the moment of inertia. Normalize the moment of inertia by dividing by the total intensity, and take the square root of the resulting quantity. Let the result be denoted by $I^{(t)}$. If the template is zoomed by μ, $I^{(t)}$ is multiplied by μ; $I^{(t)}$ is invariant to uniform changes in intensity. For each landmark's scale in the template, calculate the ratio of the template scale and $I^{(t)}$. The mean of the inverse gamma is set to this ratio times the I calculated for the current image. The variances are set *a priori* to provide a reasonable range of values. This potential is rotation, translation, and zoom invariant. If the image is zoomed, I increases by the same factor, and the mean of the gamma distributions shifts appropriately.

There are three potential functions that are useful for cliques with two elements. The first potential function that I use measures the distances between pairs of points in the candidate image and compares them to distances in the template. Because scale-space distance is invariant to rotation, translation, and zoom, it satisfies the necessary transformational invariances. The potential is given as

$$V_{ij}^{(1)} = -k(d_{ss}(\mathbf{x_i}, \mathbf{x_j}) - d_{ss}(\mu_\mathbf{i}, \mu_\mathbf{j}))^2 \tag{5.1}$$

where μ_i and μ_j are the template locations of landmarks i and j, x_i and x_j are "candidate" locations in scale space, and $d_{ss}(x_1, x_2)$ is defined in (3.2).

There is an obvious problem with this potential function. Because it only involves distances between pairs of points, there is nothing to prevent the template from "folding" on itself. In other words, this potential does not preserve angles at the landmarks. In Euclidean space, a solution to this problem would be to compare distances between triples of points. This solution does

not apply in the same way in scale space.

Scale-space distance has five isometries, or transformations that preserve distance. These are translation in space by a constant vector, rotation in space (left multiplication by an orthogonal matrix of determinant 1), reflection in space (left multiplication by an orthogonal matrix with determinant -1), zooming (multiplication of space and scale coordinates by the same constant), and inversion with respect to a hyperellipsoid. The final isometry implies that

$$d_{ss}((\mathbf{x_1}, \sigma_1), (\mathbf{x_2}, \sigma_2)) = d_{ss}(\frac{(\mathbf{x_1}, \sigma_1)}{\|\mathbf{x_1}\|^2 + \sigma_1^2}, \frac{(\mathbf{x_2}, \sigma_2)}{\|\mathbf{x_2}\|^2 + \sigma_2^2})$$

If measurements involving angles or lengths are made at a particular point, the measurements remain the same even if the coordinates are translated, rotated, reflected, or zoomed.

As an example, consider the triangle containing the three points $(x, \sigma) = (1,1), (5,1), (1,4)$. Inverting with respect to a hyperellipsoid yields $(x, \sigma) = (\frac{1}{2}, \frac{1}{2}), (\frac{5}{26}, \frac{1}{26})$, and $(\frac{1}{17}, \frac{4}{17})$. Both triangles are plotted in Figure 5.1, where the second set has been zoomed by a factor of 6 to make the figures more comparable. (Recall that zooming does not change distances.) In scale-space, the distances between any pair of points are the same in each figure. It may be misleading to label these pictures as triangles, because the distances are measured along geodesic curves, which are arcs of circles, between the points; however, the "straight line" connection is the typical visualization used when constructing a neighborhood structure for a template description.

Inversion with respect to a hyperellipsoid has consequences for the potential function $V^{(1)}$, because it depends only on scale-space distance. Suppose that the core of a figure can be parameterized as $(x(t), y(t), \sigma(t)) =$

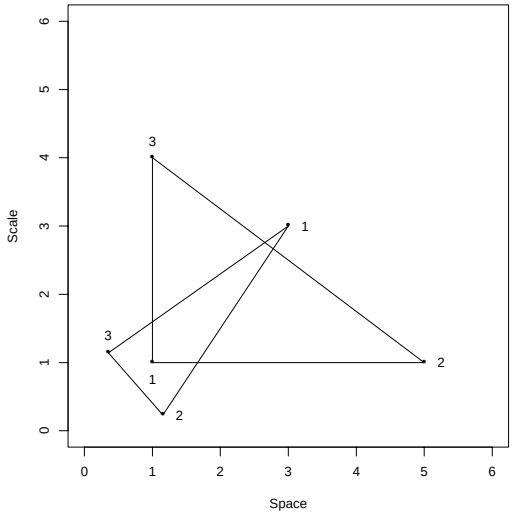
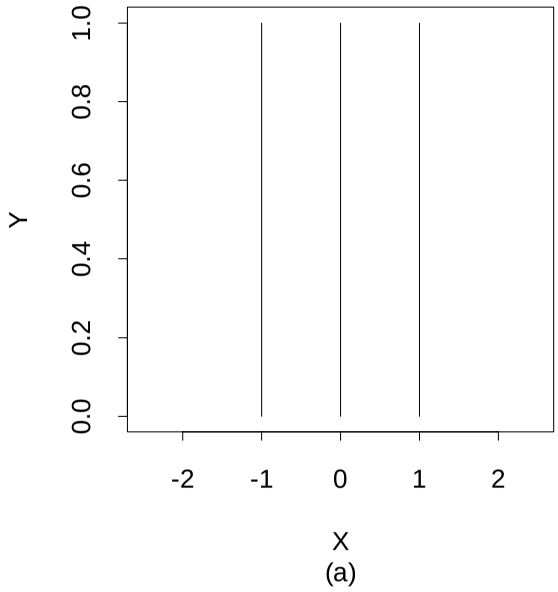


Figure 5.1: Scale-Space "Triangulation"

$(0, t, 1)$ for $t \in [0, 1]$. This is plotted in Figure 5.2(a), where scale is represented by the lines on either side of the spatial core. If the core is "inverted with respect to a hyperellipsoid," the resulting core can be parameterized as $(u(t), v(t), \gamma(t)) = (0, \frac{t}{1+t^2}, \frac{1}{1+t^2})$ for $t \in [0, 1]$. This is plotted, zoomed by a factor of 2, in Figure 5.2(b). For any $t_1, t_2 \in [0, 1]$, it is the case that the distance from $(x(t_1), y(t_1), \sigma(t_1))$ to $(x(t_2), y(t_2), \sigma(t_2))$ is the same as the distance from $(u(t_1), v(t_1), \gamma(t_1))$ to $(u(t_2), v(t_2), \gamma(t_2))$. This implies that no matter how the k landmarks and neighborhood structure are chosen for a particular template description, there exists another set of k points, not translations, rotations, reflections, or zoomings of the initial points, with the same scale-space distances between points in the structure.

The consequences of this isometry for potential function $V^{(1)}$ are, for ex-



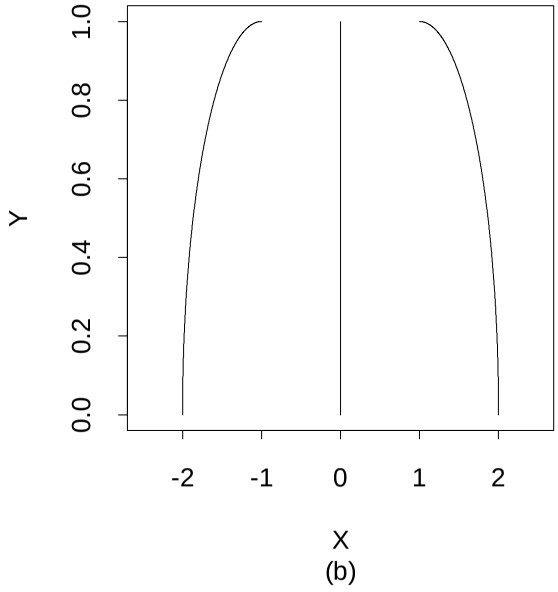


Figure 5.2: Inversion with Respect to a Hyperellipsoid

ample, that the triangles in Figure 5.1 receive the same probability, as do any translations, rotations, reflections, or zoomings of the triangles. In practice, this invariance should never cause a problem. The likelihood functions used in conjunction with these priors give higher probability to images with the correct features at the landmarks. For a particular type of image, it is extremely unlikely that appropriate features (e.g., boundaries, middles, corners) would exist at the appropriate locations and scales for both configurations of points. As a result, using triples of points and a potential function that compares the difference in their lengths to the lengths in the template may work well.

I have considered an alternative approach to modifying the potential function that still allows two-element cliques to be used. The modification preserves angles by removing rotation invariance. The motivation for the modified potential functions comes from the auto-normal model (Besag 1974). This model is a MRF with potential functions for two-element cliques of the form

$$V_{ij} = -k((x_i - \mu_i) - (x_j - \mu_j))^2$$

in the univariate case, and

$$V_{ij} = -k((\mathbf{x_i} - \mu_\mathbf{i}) - (\mathbf{x_j} - \mu_\mathbf{j}))'((\mathbf{x_i} - \mu_\mathbf{i}) - (\mathbf{x_j} - \mu_\mathbf{j})) \tag{5.2}$$

in the multivariate case, where $k > 0$ and μ is the expected value of x.

The intuition for potentials that can be used to measure shape difference comes from rewriting (5.2) as

$$\begin{aligned} V_{ij} &= -k((\mathbf{x_i} - \mathbf{x_j}) - (\mu_\mathbf{i} - \mu_\mathbf{j}))'((\mathbf{x_i} - \mathbf{x_j}) - (\mu_\mathbf{i} - \mu_\mathbf{j})) \qquad (5.3) \\ &= -k(\|x_i - x_j\|^2 + \|\mu_i - \mu_j\|^2 - \\ &\quad 2\|x_i - x_j\|\|\mu_i - \mu_j\| \cos(x_i - x_j, \mu_i - \mu_j)) \end{aligned}$$

where $\cos(x_i - x_j, \mu_i - \mu_j)$ is the cosine of the angle between the vectors $x_i - x_j$ and $\mu_i - \mu_j$. Think of μ as the template. This type of potential compares the squared distances between two points in a candidate image X and two points in the template μ and whether $x_i - x_j$ and $\mu_i - \mu_j$ are in the same direction.

From (3.4), if x_i and x_j are in the same scale-space relationship as μ_i and μ_j, then $(x_i - x_j) = c(\mu_i - \mu_j)$, and the cosine between $x_i - x_j$ and $\mu_i - \mu_j$ is one. This holds using both the scale and space components of $x_i - x_j$ and $\mu_i - \mu_j$ or just the space components. Let $sp(x)$ denote the space coordinates of a scale-space point x. By substituting scale-space distance for Euclidean distance, two other potential functions can be defined as

$$\begin{aligned} V_{ij}^{(2)} &= -k[(d_{ss}(\mathbf{x_i}, \mathbf{x_j}))^2 + (d_{ss}(\mu_\mathbf{i}, \mu_\mathbf{j}))^2 - \\ &\quad 2d_{ss}(\mathbf{x_i}, \mathbf{x_j}) d_{ss}(\mu_\mathbf{i}, \mu_\mathbf{j}) \cos(x_i - x_j, \mu_i - \mu_j)] \end{aligned} \tag{5.4}$$

$$\begin{aligned} V_{ij}^{(3)} &= -k[(d_{ss}(\mathbf{x_i}, \mathbf{x_j}))^2 + (d_{ss}(\mu_\mathbf{i}, \mu_\mathbf{j}))^2 - \\ &\quad 2d_{ss}(\mathbf{x_i}, \mathbf{x_j}) d_{ss}(\mu_\mathbf{i}, \mu_\mathbf{j}) \cos(sp(x_i - x_j), sp(\mu_i - \mu_j))] \end{aligned}$$

The potentials $V^{(1)}$, $V^{(2)}$, and $V^{(3)}$ are illustrated with the following example. Let $x_1 = (x, y, \sigma) = (0, 0, 1)$, $\mu_1 = (0, 0, 1)$, and $\mu_2 = (0, 1, 1)$ and consider the probability distribution of x_2 given x_1 as specified using each of the potential functions.

Figure 5.3 shows the distributions generated by potential function $V^{(1)}$. This potential relies strictly on scale-space distance and has no "preferred direction." Recall from Section 3.3 that the locus of points a distance $d_{ss}((0, 0, 1), (0, 1, 1)) = 0.962$ away from $(0, 0, 1)$ is a circle of radius $\sinh(0.962) = 1.12$ and center $(0, 0, \cosh(0.962)) = (0, 0, 1.5)$. Notice, however, that in the plots of X versus σ (scale) and Y versus σ that there is more scatter above $\sigma = 1.5$ than

below. Figure 5.4 shows why this is so. The circles in Figure 5.4 contain points that are equidistant from $(0, 1)$, and each circle increases the distance from $(0, 1)$ by an equal amount. The equidistant circles extend farther "up" in scale than down.

Figure 5.5 shows the distributions generated by potential function $V^{(2)}$. This function does have a preferred direction—in this case, toward $(0, 1, 1)$. There is correlation between Y and σ, which is not surprising if one looks in the area of $(y, \sigma) = (1, 1)$ in the Figure 5.3. Figure 5.6 shows the distributions generated by potential function $V^{(3)}$. This potential function considers only the spatial coordinates in the calculation of the direction cosine. The plot of X versus Y shows clearly a wedge of vectors that make an angle of approximately ninety degrees with the x-axis. The plot of Y versus σ clearly indicates that for Y near one, the scale is not restricted to one, but can move around the "equidistant" circles.

After specifying potential functions for a MRF model, it is important to insure that the resulting joint distribution can be normalized. As developed in Section 5.1.2, the prior on landmark locations is clearly improper. However, given the likelihood used in Chapter 6, if the additional prior constraint that scale is bounded away from zero is imposed, then the posterior distribution is proper. In practice, it makes sense to bound scale away from zero, because in digital images, there is really no information at the sub-pixel level.

5.3 Summary

Several properties of the prior on landmarks and shapes should be noted. The prior incorporates ideas about scale and space in several ways: (1) through the

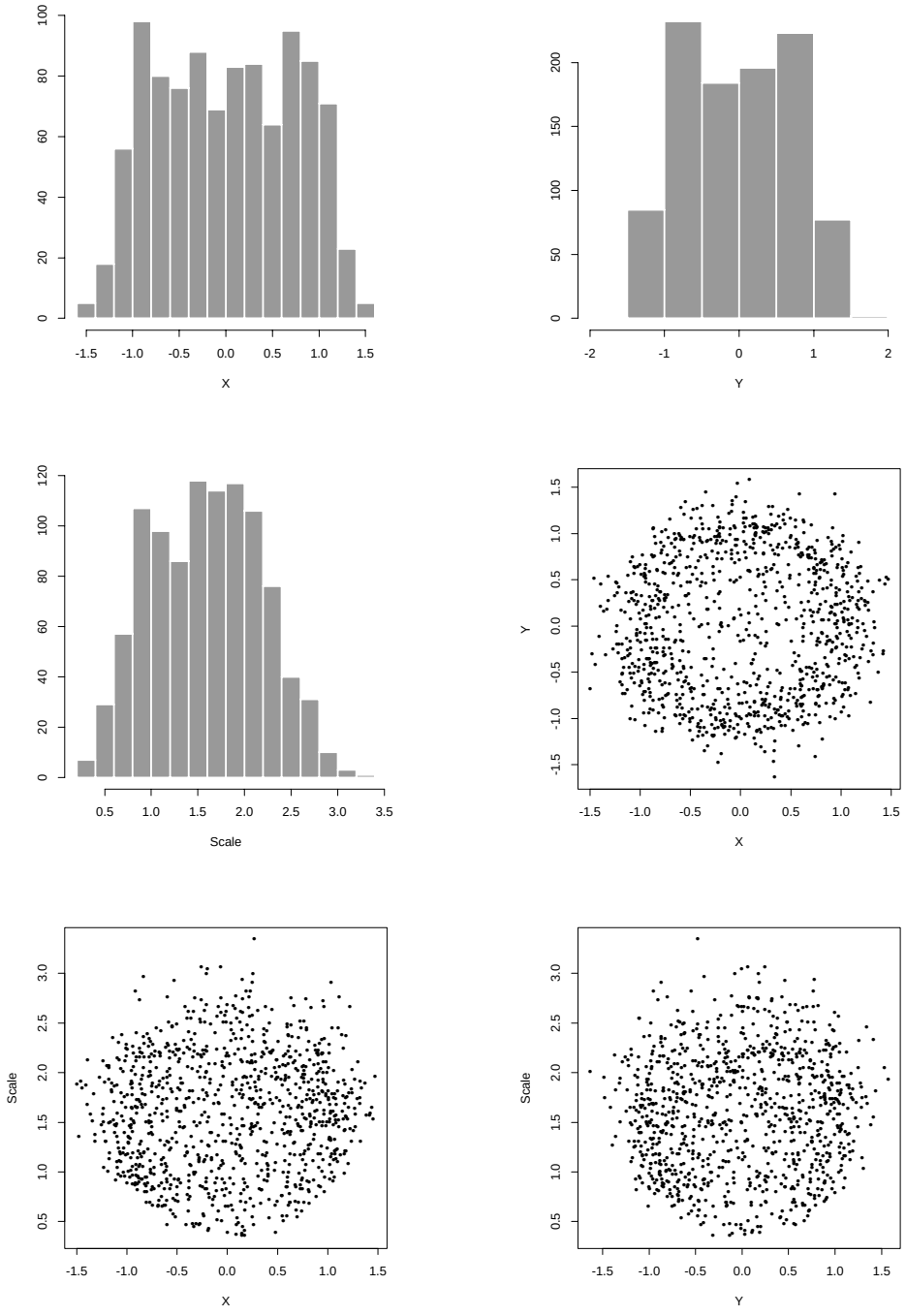


Figure 5.3: Potential $V^{(1)}$

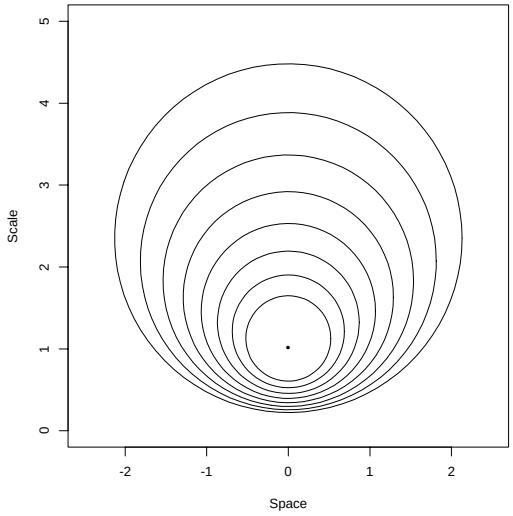


Figure 5.4: Points Equidistant from $(0, 1)$

specification of a multi-scale template description, (2) through the neighborhood structure on the template nodes, and (3) through the potential functions in the MRF model, which are functions involving scale space. Further, the specification given here is quite general and can be applied to continuous or discrete images of any dimensionality.

The variability at each node is influenced by the neighborhood structure of the template. A site with many neighbors is penalized more for moving than a site with few neighbors. This allows for the modeling of differing variability at the landmarks. The neighborhood structure in the template also allows the prior to handle obstructed views. If one branch of the template is blocked, the probability can still be relatively high due to contributions from other branches.

The prior developed in this chapter identifies likely locations for landmarks.

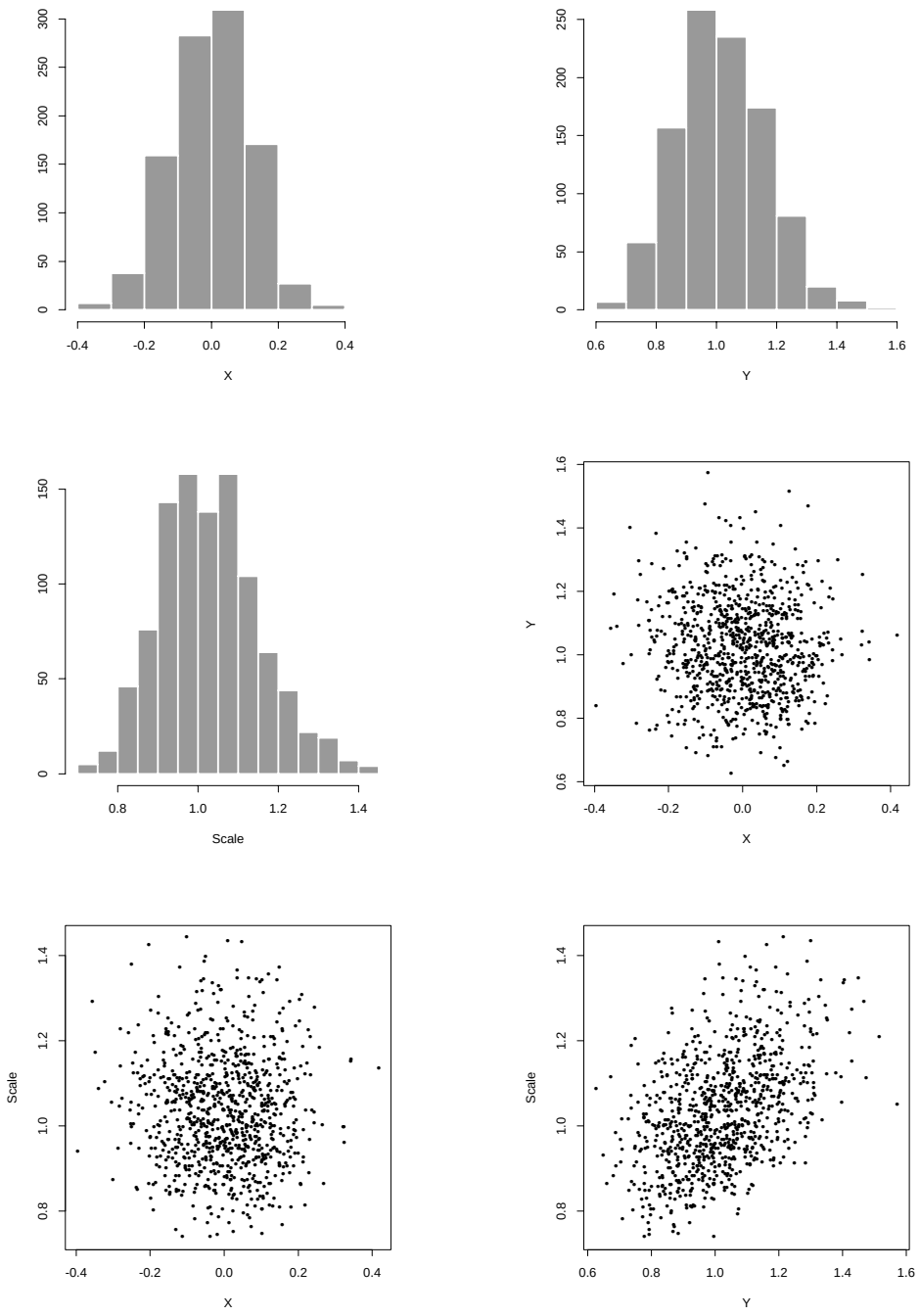


Figure 5.5: Potential $V^{(2)}$

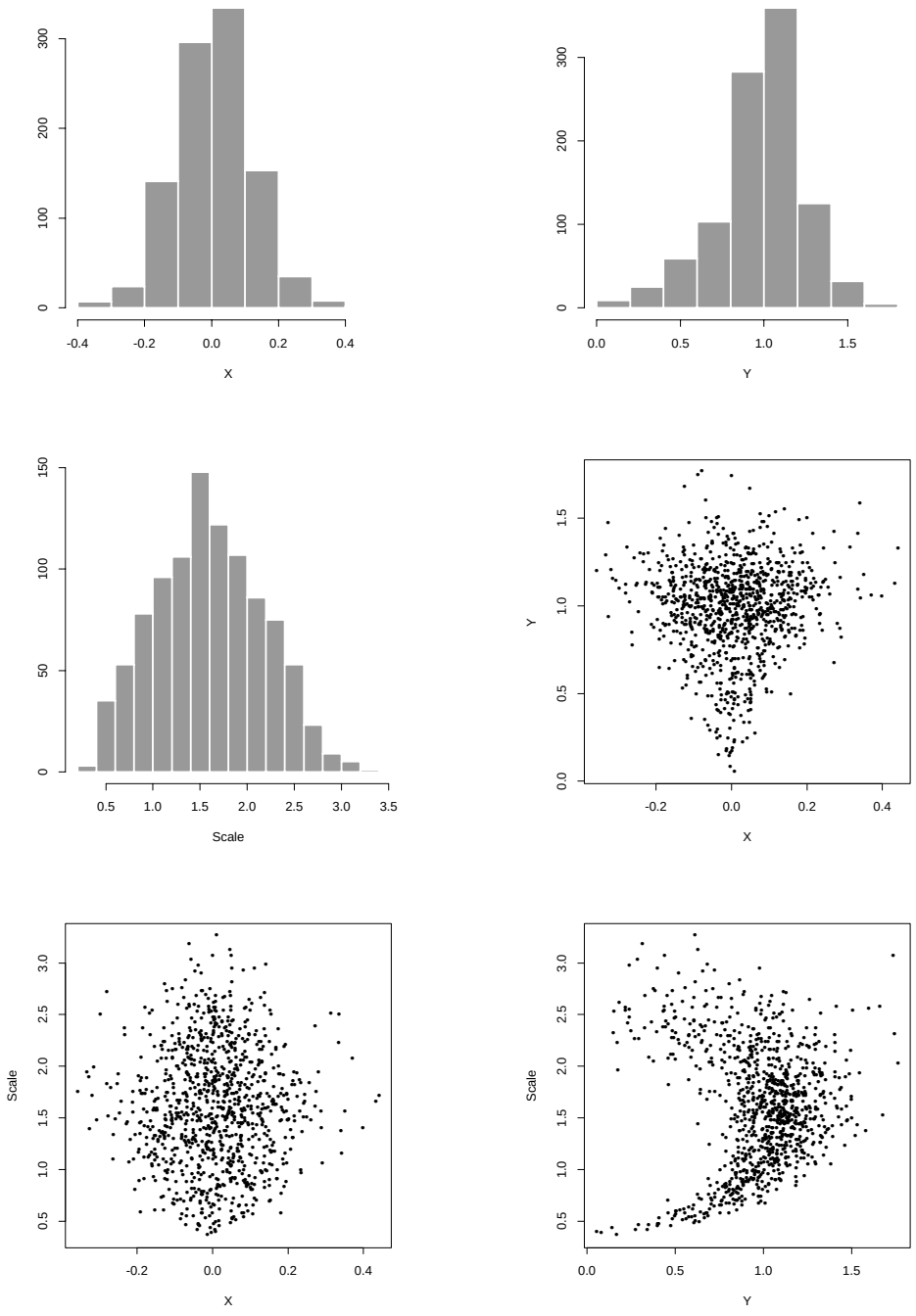


Figure 5.6: Potential $V^{(3)}$

However, the prior must interact with data through a likelihood function to determine where these locations are in a particular image. In the example in Chapter 6, I develop a template description for a particular class of images, specify two flexible classes of priors to capture information about the location of landmarks, and illustrate one method for developing a likelihood function for MR image data.

Chapter 6

Locating Ventricles in MR Brain Images

This chapter examines the problem of the automatic location of landmarks in an MR brain image. The primary focus is on locating and identifying landmarks within a ventricle. Figure 6.1 is an MR image of a brain, and the ventricle is the dark "butterfly" in the middle of the image. The landmarks that are identified include medial structures (cores) of both the brain and the ventricle, corners of the ventricle, and boundaries of the ventricle. The identification of boundaries and corners allows for the segmentation of the ventricle from the rest of the brain.

6.1 Developing the Likelihood

The development of the likelihood function, template description, and prior distribution proceeded iteratively in this example. Although conceptually the specification of the template description and prior precedes that of the likelihood, a discussion of the likelihood function will clarify some of the choices made for the prior and the template description.

Suppose that I have no prior information about the location of a particular

Figure 6.1: MR Brain Image

corner of a ventricle, and suppose that I observe a single MRI image. What function captures my posterior beliefs about the location of the corner? A reasonable first guess is the scaled corner and junction detector specified in (4.5). (The scaled detector is multiplied by σ^4.) Large positive or negative values of the corner detector specify locations that are likely to be corners. The sign varies depending on the sign of the isophote curvature at the corner. As calculated, the corner detector is inappropriate for use as a probability density, as it assumes both positive and negative values. However, for any particular corner, one knows whether its "cornerness" is positive or negative. If the particular corner that I am locating has negative cornerness, the function that captures my posterior beliefs is zero (or extremely small) where cornerness is positive and the absolute value of cornerness elsewhere.

Because my prior is uniform, and because the posterior distribution is proportional to the prior times the likelihood, this implies that my likelihood function is specified (up to a constant) by the same function as the posterior. Similar arguments can be made for using the scaled Laplacian for the "medialness" likelihood. (The scaled Laplacian is multiplied by σ^2.) A "boundariness" likelihood can be specified using the scaled, squared gradient magnitude, which is always positive. (The scaled gradient magnitude is multiplied by σ^2.) These likelihoods are translation, rotation, and zoom invariant. In this chapter, the joint posterior on all the landmarks is modeled as the product of the "features" observed at each landmark. The likelihoods can be generalized by introducing a parameter that raises these functions to a power; in the examples considered in this chapter, this parameter is specified rather than estimated.

Figure 6.2 shows the medialness, cornerness, and boundariness functions calculated on a vertical slice through the center of Figure 6.1 at scale 5.5. The likelihoods are known only up to a constant, so the vertical scales have been chosen for convenience. The positive areas of medialness correspond to the edge of the skull, the edge of the brain, and the middle of the ventricle. (Recall that the zero crossings of the Laplacian can be used for edge detection.) The positive spikes of cornerness correspond to the skull and the ends of the ventricle core. The boundariness spikes correspond to edges of the brain and ventricle.

Figure 6.3 shows medialness calculated at scales 5.5 (solid line), 6.5 (dotted line), and 7.5 (points and line). While medialness is well-localized in space, it changes slowly through scale. This implies that much of the information about the location of medial landmarks in scale comes from the prior. Cornerness and boundariness also change slowly through scale.

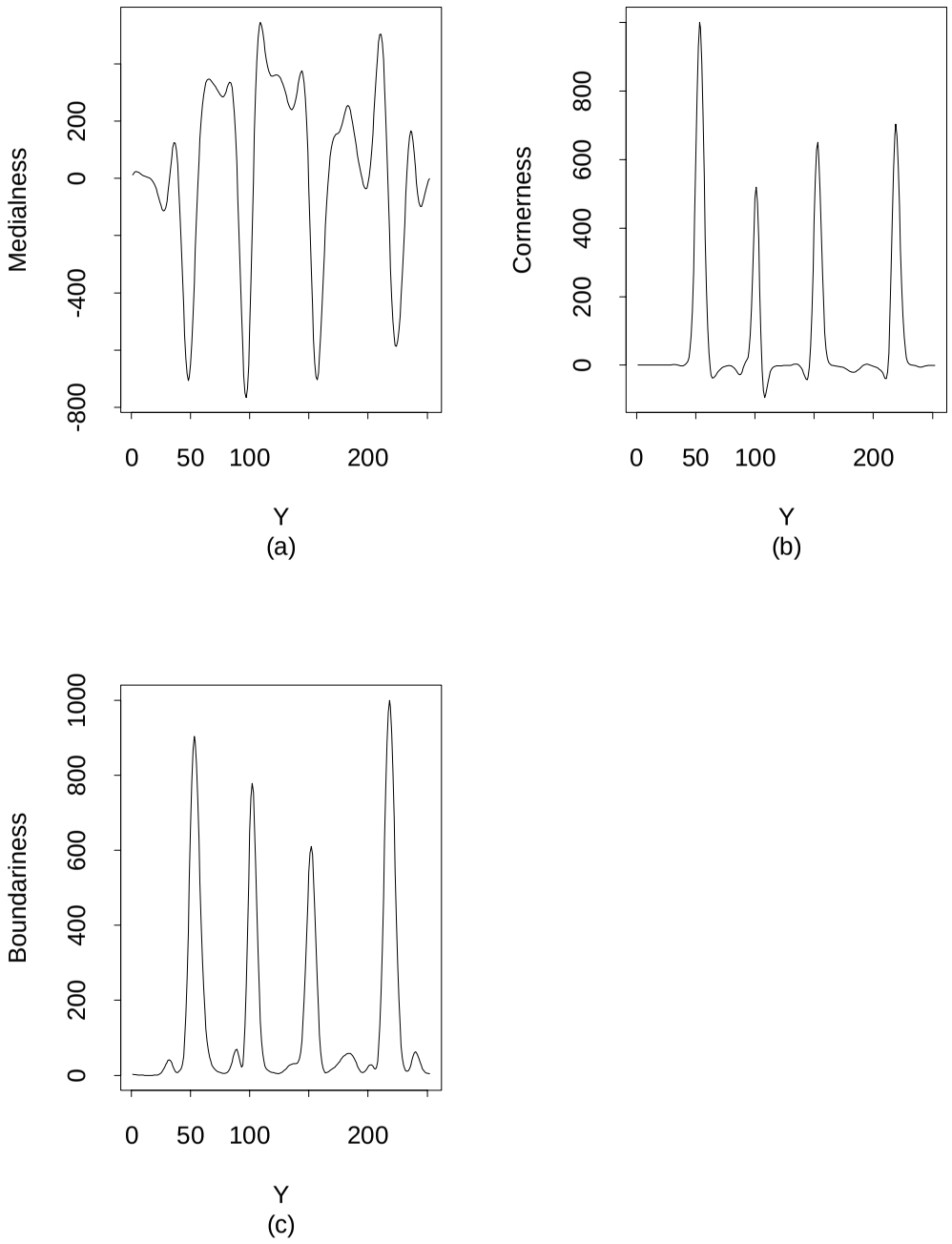


Figure 6.2: Likelihood Functions at Scale 5.5: (a) Medialness; (b) Cornerness; (c) Boundariness

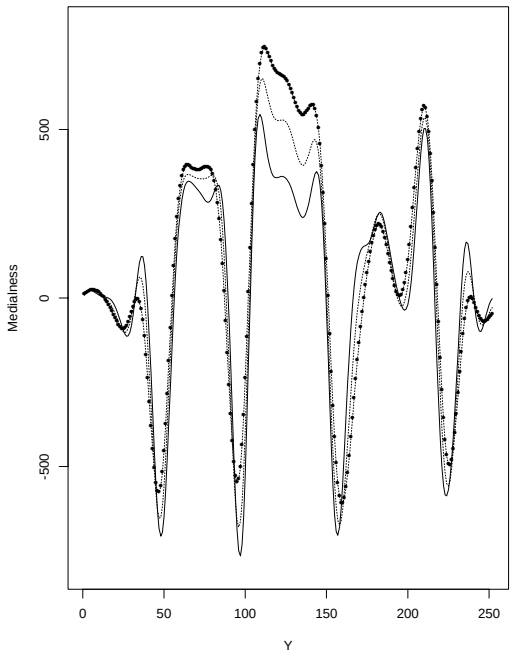


Figure 6.3: Medialness Likelihood

Figure 6.4(a) shows medialness calculated at the center of Figure 6.1 through scale. Notice that medialness peaks at approximately scale 13, which indicates that at this point, the ventricle is approximately 13 pixels wide. However, medialness begins to increase again after scale 21. The second peak, at approximately scale 55, indicates that the brain is approximately 55 pixels wide at this point. Notice that the second peak is higher than the first. This implies that some scale information must come from the prior: otherwise, the posterior on the medial landmarks will concentrate at higher scales with larger likelihood values. Similar phenomena are observed for boundariness. The squared gradient magnitude at the upper end of the ventricle core is plotted through scale in Figure 6.4(b). The small-scale peak corresponds to the edge of the ventricle; the large-scale peak corresponds the scale where this point is the most like the

edge of the (very) blurred brain.

6.2 Developing the Template

The template was developed from seven MR images. These images were segmented to identify ventricles, and their cores were calculated using the methods in Fritsch et al. (1995). The cores were translated and zoomed so they were positioned at the same point and had the same size (i.e., the same sum-of-squares difference from the mean). Eleven landmarks were identified: the ends of the central ventricle core, the ends of the four ventricle arm cores, and the largest-scale point on the central ventricle core. These landmarks were averaged to produce the initial outline of a template.

Two additional MR images were used to calculate the core of the brain. Three landmarks were identified: the ends of the core, and the largest-scale point on the core. The brain core is needed to help localize the ventricle core in scale. Because the likelihood function for medialness is much larger at large scales than at small, it is useful to locate the brain core and then "look down" through scale for the ventricle core. These points were added to the initial eleven points, and the resulting points were symmetrized slightly.

Although the landmarks that have been identified were found using cores, they are not necessarily medial points. I have observed that the endpoints (vertices) of cores occur at local extrema of the corner detector. Of the fourteen points already identified, twelve identify "corners" or core vertices. The other two are the largest-scale points along the brain and ventricle cores.

Four additional corner points were added outside the ventricle at the corners where the arms of the ventricle connect to the body of the ventricle. These

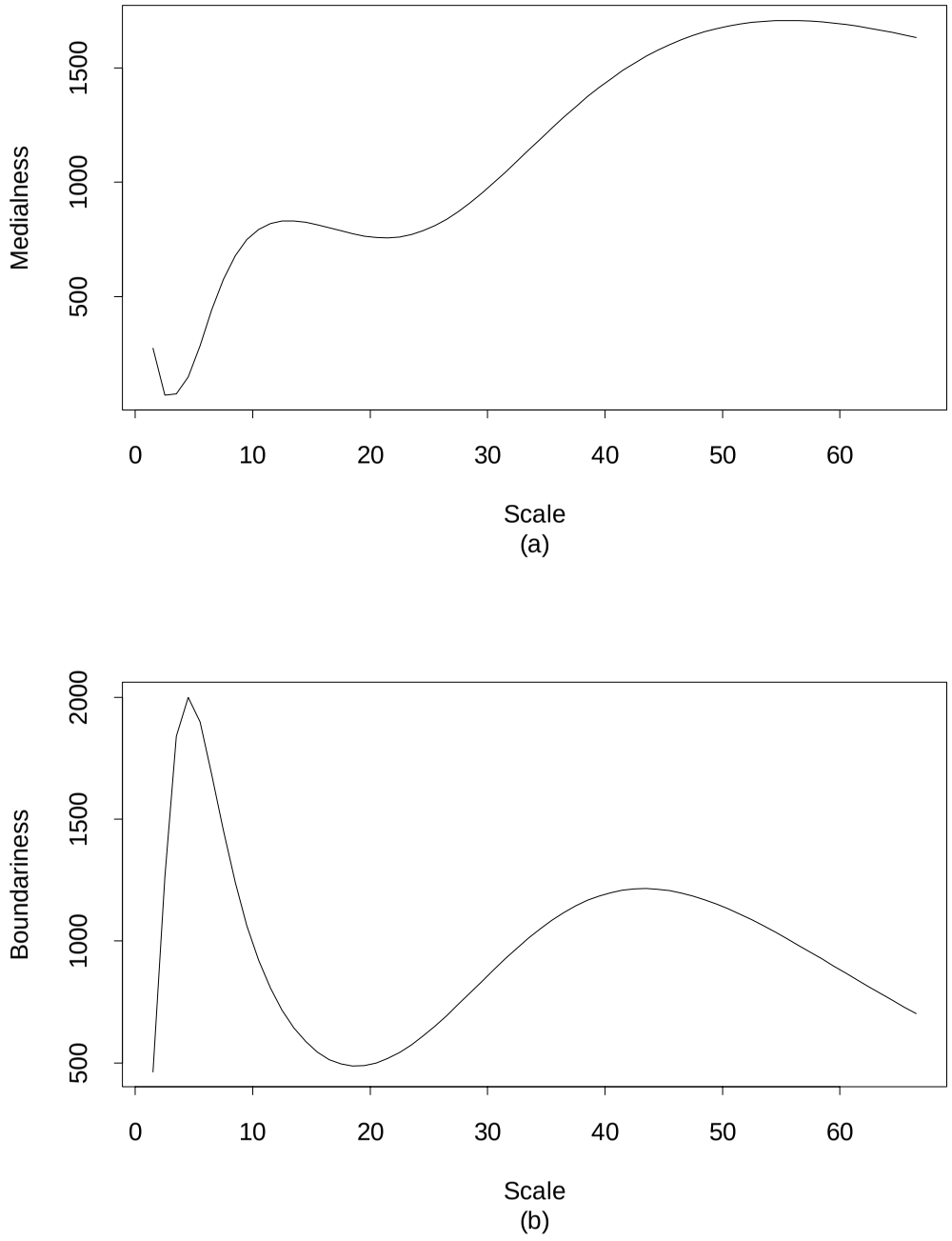


Figure 6.4: Medialness and Boundariness through Scale: (a) Medialness; (b) Cornerness

landmarks were selected by identifying the points on segmented ventricles and then re-centering, re-sizing, and re-averaging the template landmarks.

Seven medial points were added on each side of the largest-scale core point for the brain core; three were added on each side of the central ventricle core; three total were included between the endpoints of the ventricle arm cores. These points were chosen to be roughly equally spaced along their respective cores. Two boundary points, one on each side of the ventricle, were included for each medial point. These points were chosen by sketching a likely ventricle using the template points that had already been selected. Their scale was chosen by calculating the maximal boundariness response scale for the segmented ventricles. As discussed above, the boundary points have a large likelihood near the ventricle and help to locate the ventricle in scale.

The brain core is plotted in Figure 6.5(a), with the solid lines indicating neighbor relationships. The endpoints (vertices) are "corners" and other points are medial. The template points for the ventricle are plotted in Figure 6.5(b). The points indicated with an "x" are corners; boundaries and middles are clear from context. Figure 6.5(c) shows the neighbor relationships between the ventricle template points. The neighbor relationships between the brain core and the central ventricle core are shown in Figure 6.5(d). The cross-connections between the two cores allow scale information to be "shared" between the two cores.

6.3 Developing the Prior

As discussed above, the likelihood functions change slowly in scale. Consequently, much of the information about scale comes from the specification of

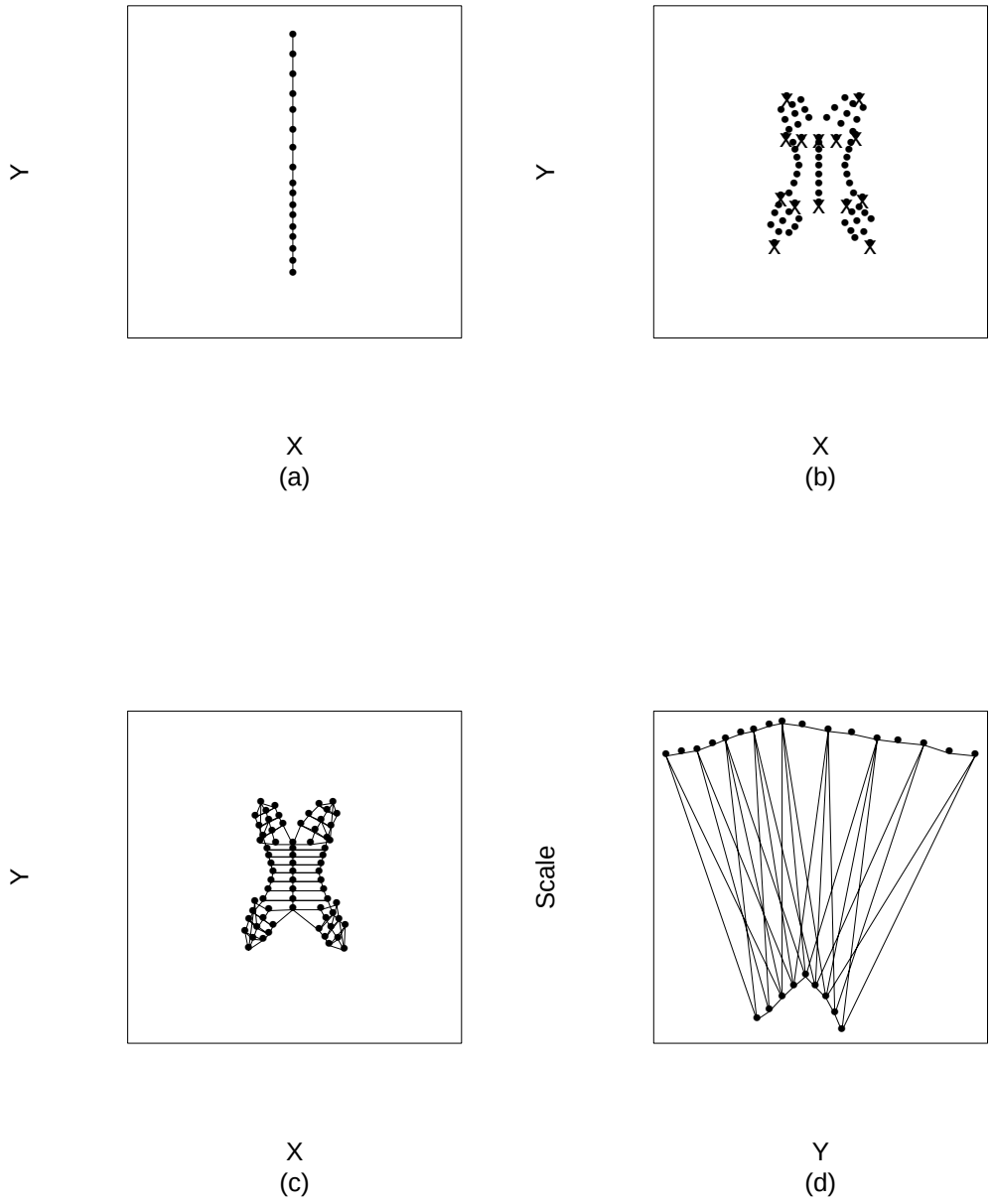


Figure 6.5: Template: (a) Brain Core; (b) Ventricle Landmarks; (c) Ventricle Neighbors; (d) Brain-Ventricle Neighbors

the template and the prior. The template developed above contains boundary landmarks, which help locate the ventricle in scale. This section proposes two classes of priors that can capture both shape and scale information.

The first class of priors uses the potential functions $V^{(2)}$ specified in (5.4). These potential functions penalize points that are not in the same spatial and scale relationships as points in the template. Including the brain core in the template provides large-scale landmarks that these potential functions can use to locate the smaller scale ventricle landmarks. While this prior is not rotation invariant, it is translation and zoom invariant.

One problem with using the $V^{(2)}$ potential is that it assigns a penalty if the arms of the ventricle do not intersect the center of the ventricle at the same angle as they do in the template. An extension that would improve this potential function would make the penalty depend on the angles that had been observed at other landmarks further up the hierarchy in the template description. This could likely be achieved by modifying the potential functions to use information from larger cliques.

The second class of priors uses the potential functions $V^{(1)}$ specified in (5.1) for two-element cliques and "inverse gamma" potentials for one-element cliques containing the scale coordinate of each landmark. The $V^{(1)}$ potential functions specify only the distances between landmarks. This solves the problem of assigning a penalty if the ventricle arms intersect the ventricle body at angles unlike those in the template; however, it does so at the cost of a great deal of scale information. Including the inverse gamma potentials adds back in some of that scale information. Notice that this also removes the effect of inversion with respect to the hyperellipsoid.

6.4 Computation

Samples from the posterior distribution of these models can be obtained using the Metropolis-Hastings algorithm. Candidate figures are generated by sampling the spatial and scale coordinates of one landmark at a time from a normal distribution centered at the current landmark with a standard deviation that is a constant times the scale of the current landmark. The normal distribution for scale is truncated at 0.5 pixels, which insures that the posterior distribution is proper.

The starting value for the algorithm is chosen using I (a function of the moment of inertia—see Section 5.2) and the centroid for the image. The template is translated so that it is centered at the centroid minus ten units in the y-coordinate and zoomed by $I/I^{(t)}$. The likelihood functions are pre-calculated on a lattice of points in scale space and interpolated as necessary using cubic splines. Parameter values in the priors and likelihood are chosen rather than estimated; Chapter 7 discusses future work in parameter estimation. The potential functions on two-element cliques are divided into two classes: those involving cliques containing the brain core, and those involving only ventricle nodes. One constant is chosen for each class of cliques so that an approximately equal penalty is assigned to the brain core and the ventricle nodes.

6.5 Results

In this section, the results of simulating from the posterior distribution for landmark locations in six images are displayed. Both types of priors and several values for the likelihood parameter are considered.

The first results were obtained without raising any of the likelihood func-

tions to a power. Figure 6.6(a) shows a sample from the posterior with $V^{(2)}$ potential functions; Figure 6.6(b) shows a sample from the posterior with $V^{(1)}$-inverse gamma potential functions. The posterior uncertainty in these samples does not adequately represent my posterior belief in the locations of the features: there is far too much variability. This suggests that the medialness, cornerness, and boundary functions require further calibration. Since their information content is apparently too weak on the scale employed, I tested higher powers of the likelihood.

To peak the posterior more, the likelihood was raised to both the fifth power and the tenth power. Figures 6.7 and 6.8 show four samples from the posteriors of each of two images with the $V^{(2)}$ potential functions and the likelihood raised to the fifth power. Figures 6.9 and 6.10 show four samples from the posteriors of each of two images with the $V^{(1)}$-inverse gamma potential functions and the likelihood raised to the fifth power. Figure 6.11 shows four posterior samples from an image with the $V^{(2)}$ potential function and the likelihood raised to the tenth power. Figure 6.12 shows four posterior samples from an image with the $V^{(1)}$-inverse gamma potential functions and the likelihood raised to the tenth power.

These images demonstrate several things. First, raising the likelihood to the fifth power has been successful in peaking the posterior distribution: the sampled images are much closer to the brain ventricles. Next, there is considerably less information in the data about the large-scale brain core than about the ventricle cores: this is especially evident in Figures 6.9 and 6.10, where the brain core does not stretch across the brain. Further, the prior has added information to the data: for example, in Figure 6.7, the boundary is well-located despite dark protrusions into the ventricle. Finding the ventricles

(a)

(b)

Figure 6.6: Posterior Samples, Original Likelihood: (a) $V^{(2)}$ Prior; (b) $V^{(1)}$-Inverse Gamma Prior

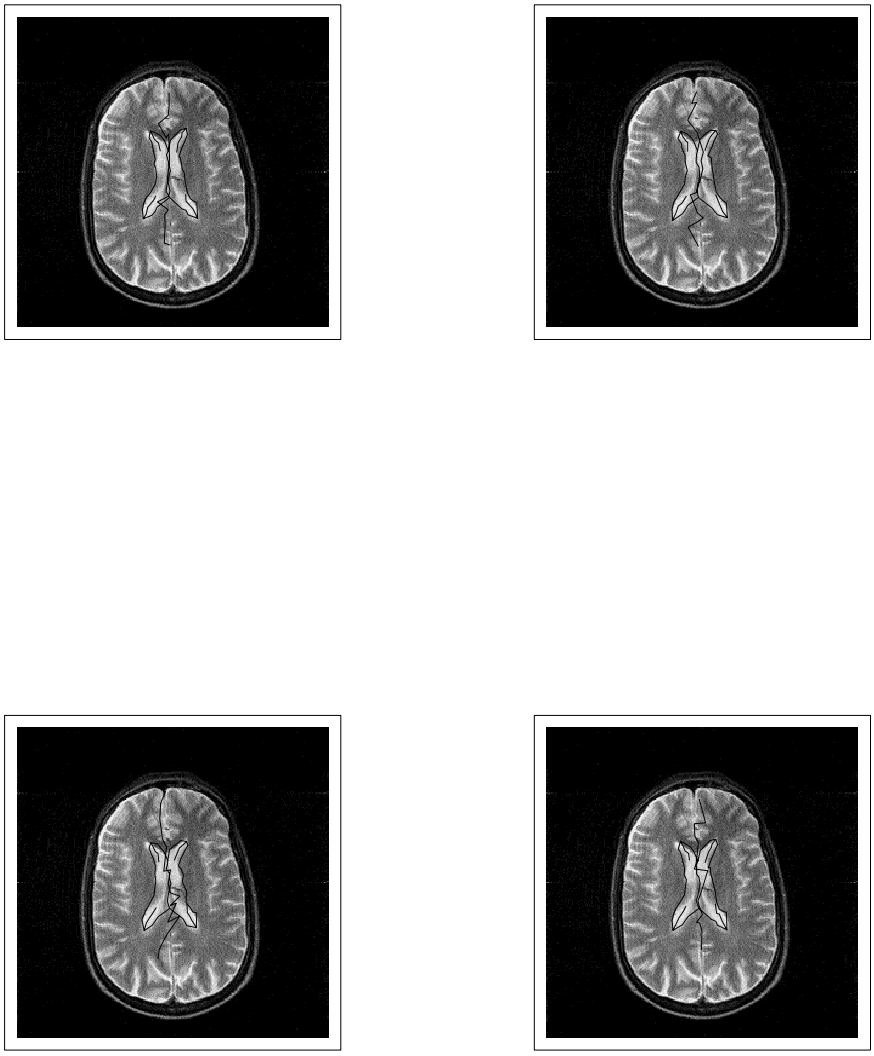


Figure 6.7: Posterior Samples, Fifth Power of Likelihood, $V^{(2)}$ Potential

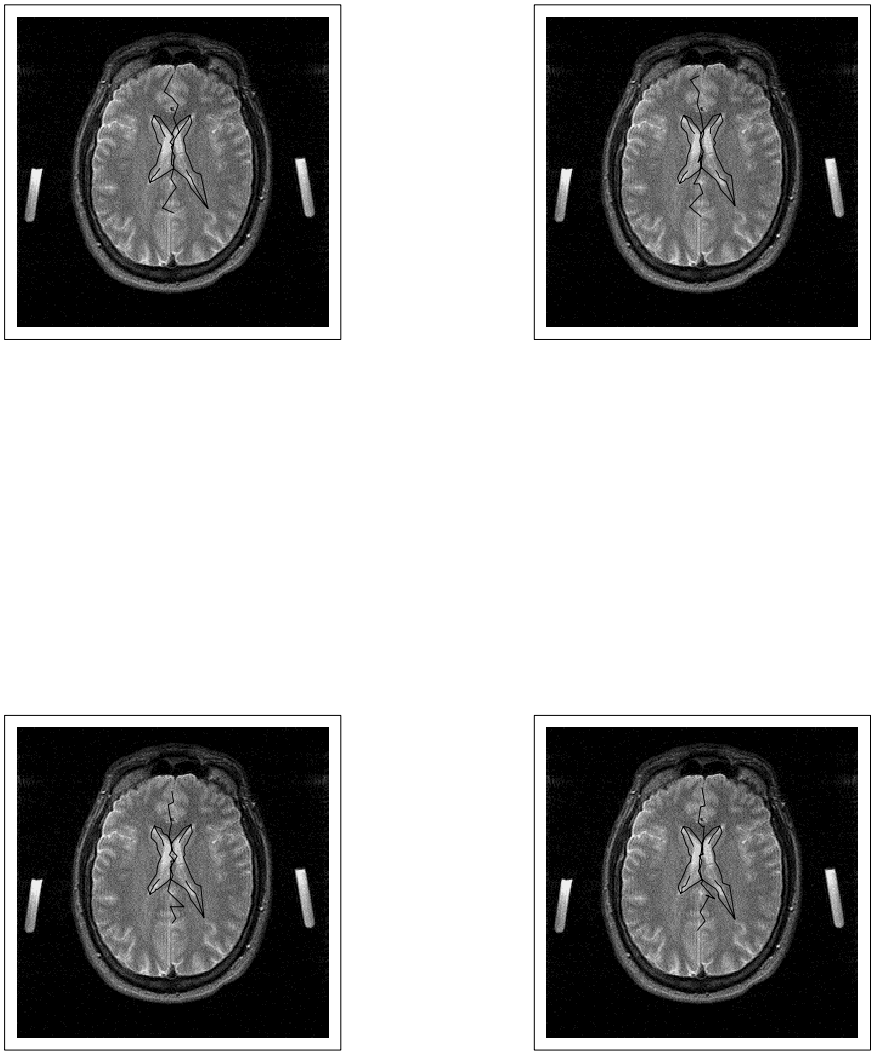


Figure 6.8: Posterior Samples, Fifth Power of Likelihood, $V^{(2)}$ Potential

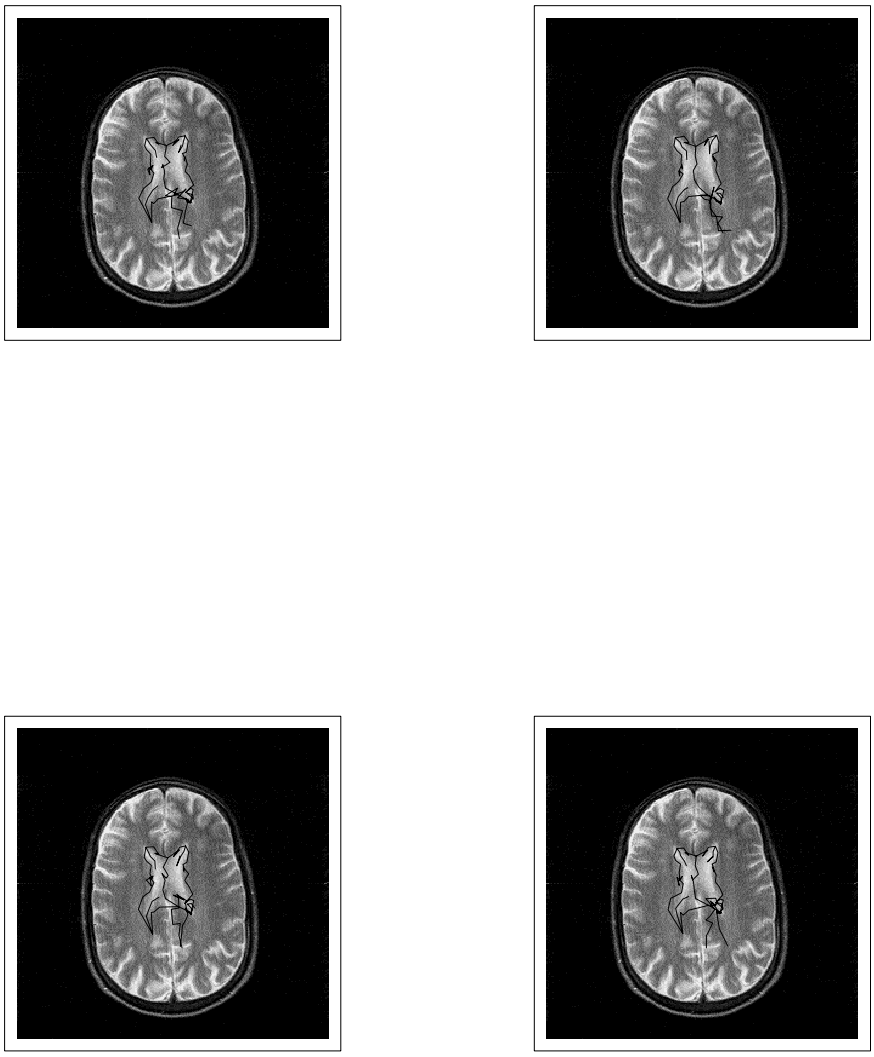


Figure 6.9: Posterior Samples, Fifth Power of Likelihood, $V^{(1)}$-Inverse Gamma Potentials

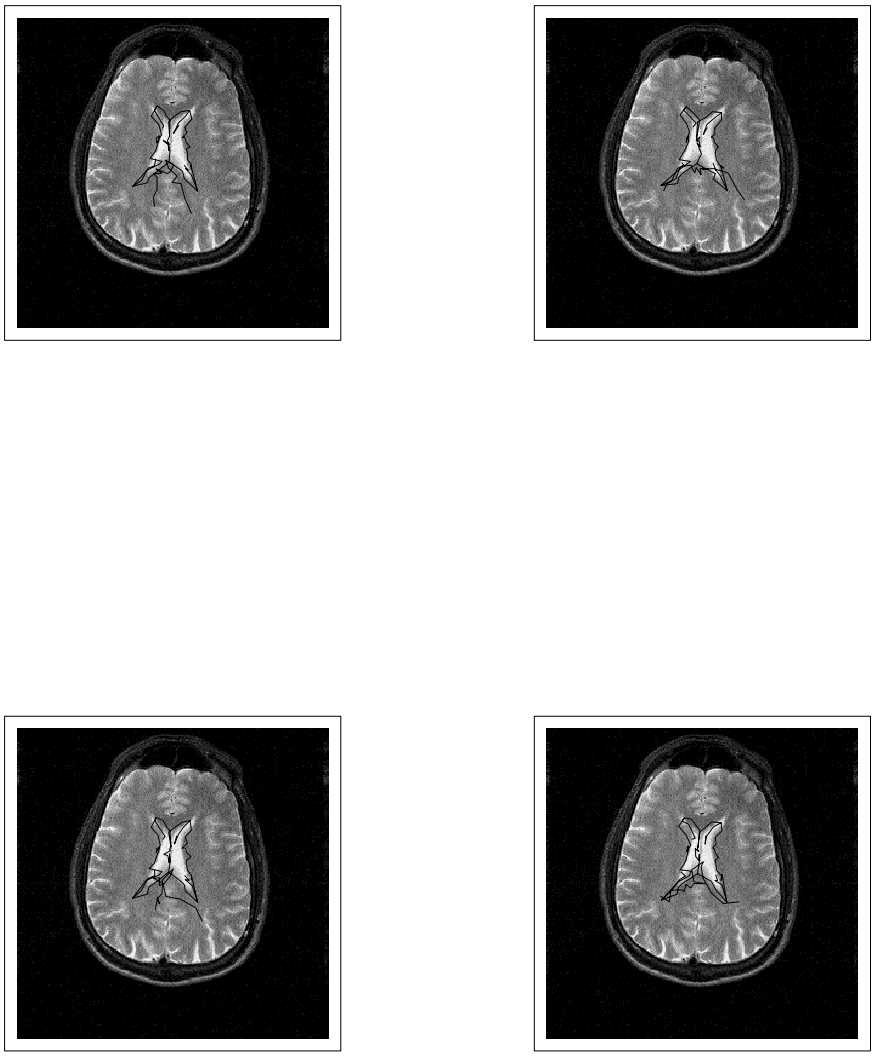


Figure 6.10: Posterior Samples, Fifth Power of Likelihood, $V^{(1)}$-Inverse Gamma Potentials

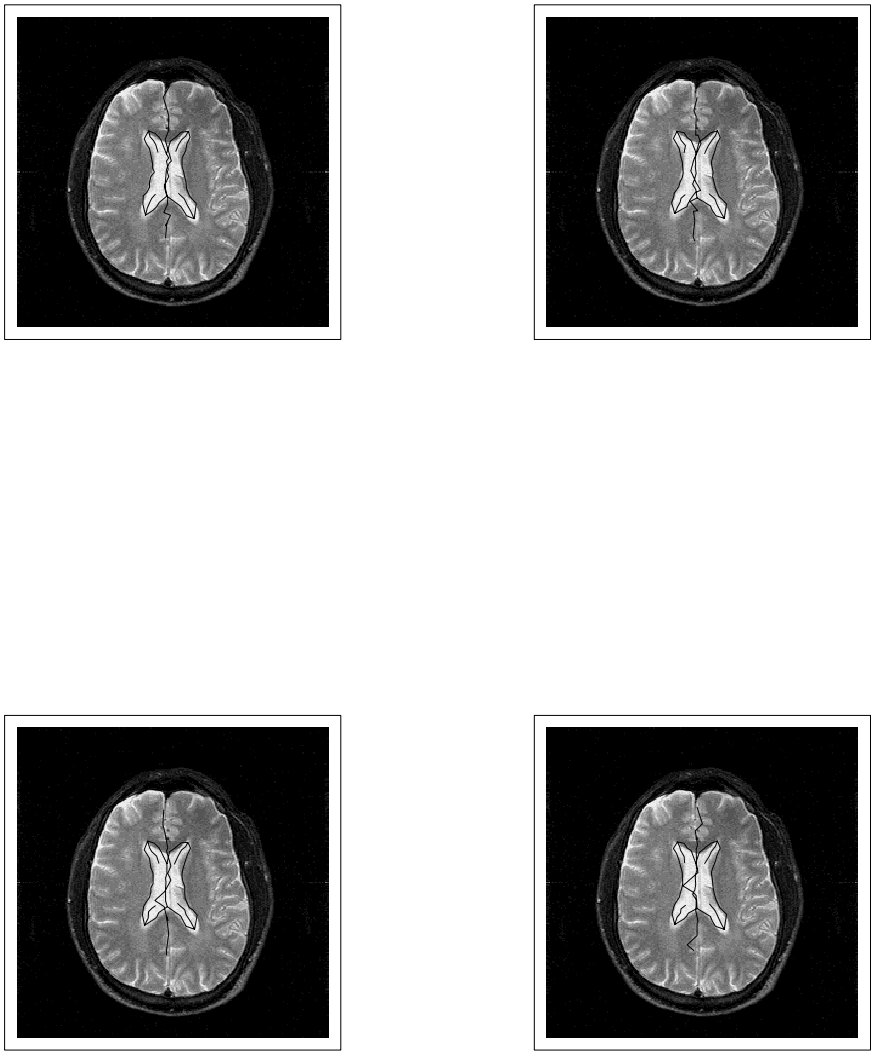


Figure 6.11: Posterior Samples, Tenth Power of Likelihood, $V^{(2)}$ Potential

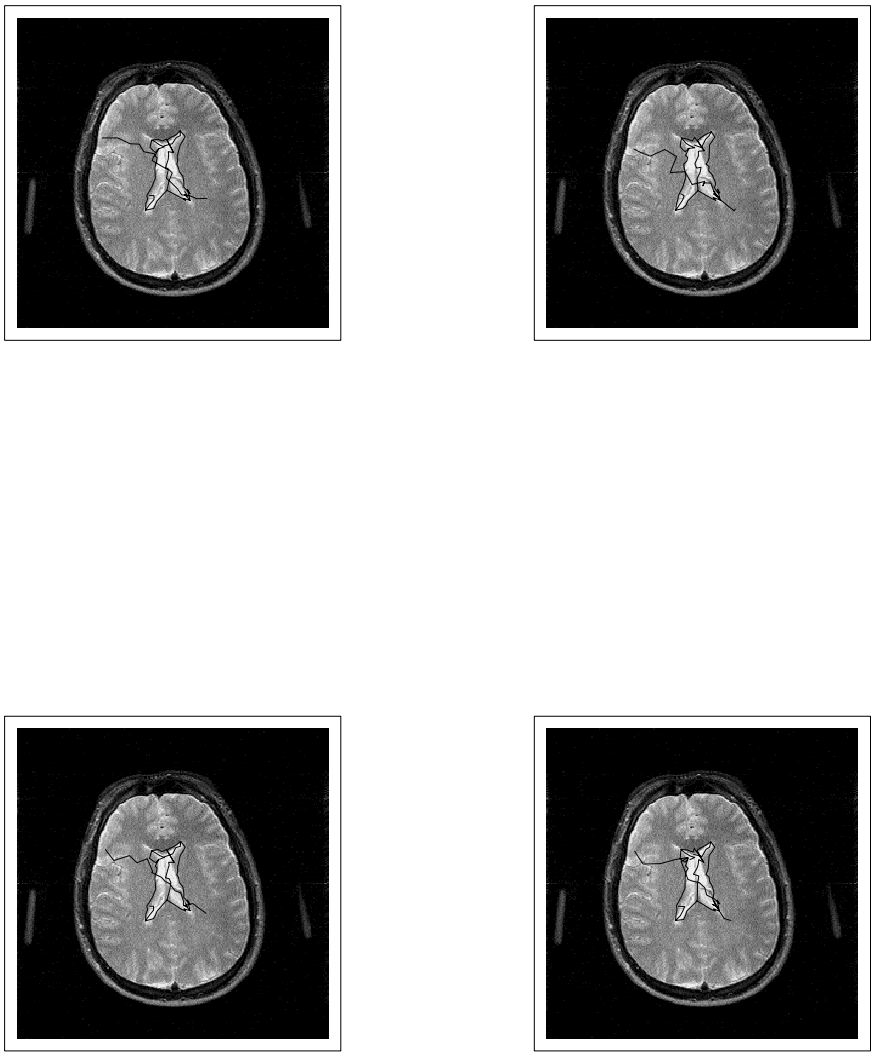


Figure 6.12: Posterior Samples, Tenth Power of Likelihood, $V^{(1)}$-Inverse Gamma Potential

in these images is not an extremely difficult task, and even techniques like intensity thresholding work reasonably well. However, thresholding would not be able to locate the boundaries of the images with dark protrusions as well as the model does. Finally, the $V^{(2)}$ potentials appear to be more successful than the $V^{(1)}$-inverse gamma potentials. There is no obvious bias toward the template in the samples and the brain core is better located.

Figure 6.13 and 6.14 show the images with the highest observed probability for the $V^{(2)}$ and the $V^{(1)}$-inverse gamma potentials respectively. The upper left-hand image in each set was sampled with the tenth power of the likelihood.

6.6 Summary

This chapter uses two flexible classes of template priors to automatically locate and segment ventricles from brain MR images. The optimal model probably lies somewhere between these two models: intuitively, the $V^{(2)}$ model provides too much direction in scale, while the $V^{(1)}$-Inverse Gamma model provides too little. However, the $V^{(2)}$ model performed much better in the simulations shown here. The methods proposed here are flexible and computationally tractable. The Metropolis-Hastings sampling algorithm provides not only an estimate of the mode, and consequently a probable segmentation, but also a posterior probability distribution for the location of cores and other landmarks of interest.

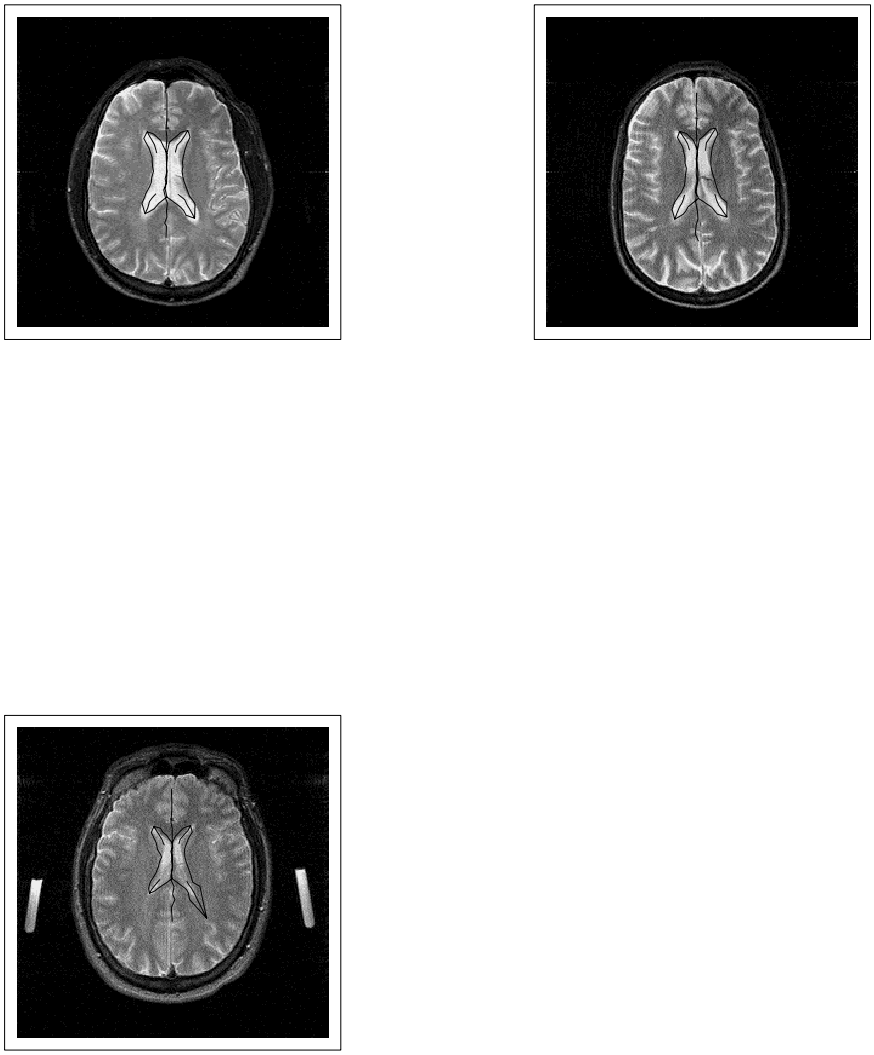


Figure 6.13: Maximum Observed Posterior Probability, $V^{(2)}$ Potential

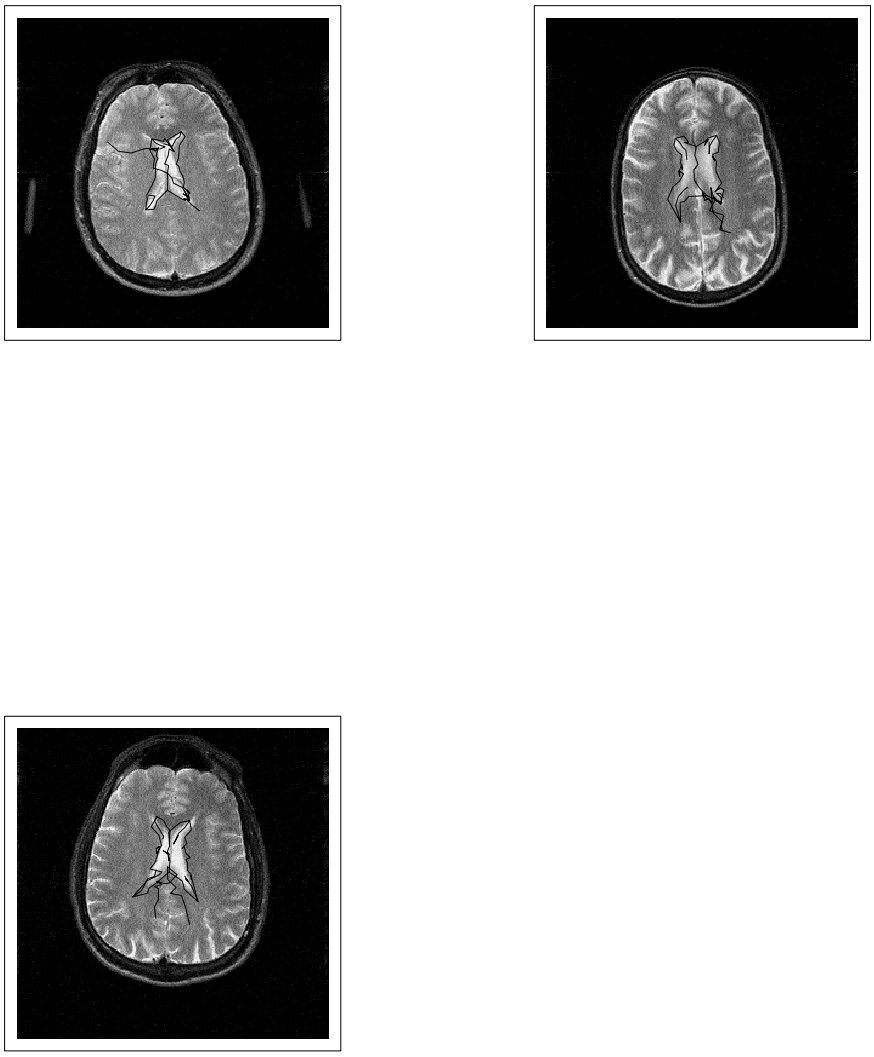


Figure 6.14: Maximum Observed Posterior Probability, $V^{(1)}$-Inverse Gamma Potentials

Chapter 7

Summary and Extensions

7.1 Important Results

The goal of my research is to develop flexible methods to incorporate prior knowledge from templates into algorithms for image analysis within a Bayesian framework. This dissertation has explored the value of scale-space methods for addressing this problem and has proposed two classes of prior models that capture template information.

The prior distributions I have developed assign probabilities to the locations of landmarks. Information about features enters the model through the likelihood function. The prior uses a MRF model with potential functions that are translation and zoom invariant; some of the proposed potentials are also rotation invariant. The sites of the MRF are anatomical structures in scale space, not pixel locations on a lattice. This is an important generalization of the MRF models that have been employed previously in image analysis.

One of the central ideas used in my "template priors" comes from computer vision research. Features in an image occur at a variety of scales, and to effectively model spatial and scale relationships, this variety of scales must be

modeled. Scale space is an image representation that handles image structure at all resolutions simultaneously and allows efficient calculation using features at multiple scales. By specifying a hyperbolic geometry on scale space, the priors can be chosen to be rotation, translation, and zoom invariant. These invariances allow the prior to favor configurations of landmarks like those in the template without regard for orientation, location, or size. Scale space also provides a natural framework for the specification of rotation, translation, and zoom invariant likelihood functions that capture ideas like "medialness" and "cornerness."

The priors on landmarks are especially easy to work with computationally due to their MRF structure. Hierarchical and geometric relationships between objects are specified using the neighborhood relationships between sites. MRF models can be simulated in a straightforward way using Markov chain Monte Carlo techniques. The results demonstrated in Chapter 6 are a strong first step toward the development of automatic segmentation methods.

The priors proposed in this dissertation provide effective and flexible methods for using template information to enhance image analyses. These methods can be applied to a variety of important real-world problems. For example, numerous PET data sets have been collected to identify areas of the brain activated by particular tasks. While the PET images show regions of activity, they do not identify anatomical structure well. By using high-resolution MR images to develop templates, the anatomical information from the MR images could be mapped to the functional areas identified by the PET images. This could allow reliable identification of anatomical areas of the brain that show high activation in response to particular stimuli without requiring extensive user interaction.

7.2 Extensions

7.2.1 Template Descriptions

There are several areas where this work can be extended. One area that needs exploration is the development of template descriptions. Questions remain about how to set up the neighborhood relationships among nodes. These connections must capture the shape of the template by specifying cliques that have interesting geometric relations. How does one choose which spatial and scale relationships are critical to modeling shapes that are like the template? Which cliques are most effective in providing the right balance of structure and flexibility?

Another issue in template description is the selection of landmarks. What features are stable across images? What features are reliable predictors of anatomical structure? How can landmarks be selected in three-dimensional images? What is the impact of adding and deleting landmarks from the template description? Can structural symmetries be exploited?

Methods for template description that generalize to multi-figure images are necessary to describe complete image scenes. What connections describe the relationships between figures in an image? What are the appropriate hierarchical relationships among figures?

7.2.2 Prior Parameter Estimation

The $V^{(1)}$, $V^{(2)}$, and $V^{(3)}$ potential functions have a parameter that must be specified. In the examples in Chapter 6, these parameter values were fixed, but assume instead that one wishes to estimate them. For simplicity, suppose that each potential function has the same constant k. Because the prior on land-

marks is improper, the importance sampling methods of Geyer and Thompson (1992) are not directly applicable. However, if one landmark is fixed, then the prior is proper (see Section 7.2.3) and their methods may provide useful insight into reasonable parameter values.

Even with a proper prior, the estimation of k is somewhat problematic, as the normalizing constant of the prior distribution depends on k. This normalizing constant is unknown. However, it is possible to estimate it using an importance sampling method (Geyer and Thompson 1992). This method uses the fact that "the ratio of the normalizing constants for any pair of distributions (with common support) can be expressed as an expectation with respect to one of the distributions" (Higdon et al. 1994). Consider two densities $\frac{1}{Z_0}h_0(x)$ and $\frac{1}{Z_1}h_1(x)$. Then

$$E_0(\frac{h_1(x)}{h_0(x)}) = \int \frac{h_1(x)}{h_0(x)} \frac{1}{Z_0} h_0(x)dx = \frac{Z_1}{Z_0} \int \frac{1}{Z_1} h_1(x)dx = \frac{Z_1}{Z_0}$$

Given N realizations from $\frac{1}{Z_0}h_0(x)$, the ratio can be estimated by

$$\frac{1}{N} \sum_{t=1}^{N} \frac{h_1(x^t)}{h_0(x^t)}$$

By pre-calculating the normalizing constants for several values of k, the prior parameter can be incorporated into the Metropolis-Hastings sampling method used to obtain posterior samples for the landmark locations. If information from several images is used, the prior can be calibrated to model the variability present in the range of human brains.

7.2.3 Priors for Shapes

The prior proposed for landmarks can be generalized into a prior for shapes. Shape is defined in Goodall (1991) as those characteristics of a figure that are unchanged when a figure is translated, rotated, or scaled—arguably, reflection can be included as well. The strategy is to condition the joint distribution of the random variables representing landmark locations on functions that remove the effects of translation, rotation, reflection, and scale.

For example, to remove the effect of translation from configurations of landmarks, one can require that the landmarks have the centroid of their spatial coordinates at the origin. All figures that differ only by a translation are mapped to the same point. Of course, there are any number of functions that accomplish the same goal (and this is true for rotation, reflection, and zooming as well).

To remove the effect of scaling, the squared Euclidean distance of the spatial coordinates from the spatial centroid can be set to a constant. (In the statistical theory of shape, this constant is chosen to be 1.) To remove the effect of rotation in two dimensions, it is sufficient to rotate the figure until a specific landmark, say, x_1, is on the x-axis.

In two dimensions, reflection has a forward mapping specified as (Golub and van Loan 1989)

$$A = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ \sin(\theta) & -\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

for $\theta \in [0, 2\pi)$. This corresponds to reflecting the vector (u, v) across the line defined by the span of $(\cos(\theta/2), \sin(\theta/2))$. Notice that reflection can also be

written as

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

which negates the y-coordinates (reflects about the x-axis) and then rotates. This implies that to remove the effects of reflection in two dimensions, rotate the figure so that a specific landmark, say, x_1 is on the x-axis and then negate the y-coordinates of all the landmarks. Reflection can also be written as a Householder matrix with the form $H = I - 2uu'$, where u is a unit vector.

These functions remove the effects of translation, rotation, and reflection, and uniform scaling. However, if this method is used with the potential function $V^{(1)}$ alone, there is an additional isometry of inversion with respect to a hyperellipsoid. This implies that the shape distribution generated above will be bimodal, with one mode at the template shape, and one at its hyperellipsoid inversion.

Figure 7.1(a) shows a template and a random sample from a prior for shapes using each of the potential functions. Figure 7.1(b) demonstrates what happens with the potential function $V^{(1)}$. Because there is no directionality, it is required only that each segment be the appropriate length, but nothing prevents the figure from bending, twisting, and folding. Potential function $V^{(3)}$ (Figure 7.1(d)) is less restrictive than potential function $V^{(2)}$ (Figure 7.1(c)), as it requires only that spatial directions be preserved, but not that scale direction be preserved. This is difficult to see because only the spatial dimensions of the samples are plotted.

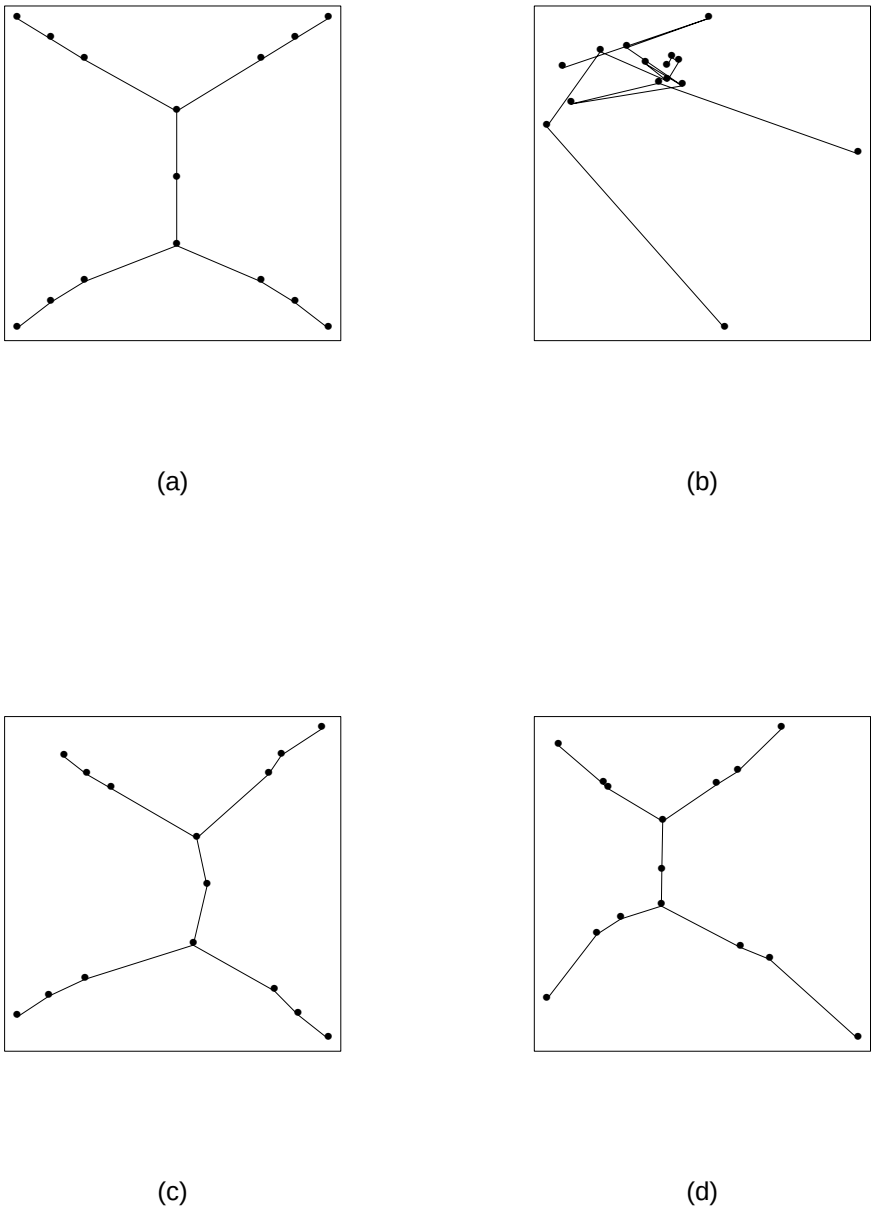


Figure 7.1: (a) Template; (b) Random Sample with Potential $V^{(1)}$; (c) Random Sample with Potential $V^{(2)}$; (d) Random Sample with Potential $V^{(3)}$

7.2.4 Likelihood Functions

Once the template description is specified, the appropriate likelihood model is problematic. The likelihood function used in Chapter 6 has intuitive appeal, but how well does it model experimental results? Can the likelihood be calibrated so that it more closely resembles the way human subjects perceive medialness, boundariness, and cornerness? To address these issues, I have obtained the experimental data discussed in Burbeck and Pizer (1994). In the experiment that they conducted, the stimuli were rectangles of varying widths with the horizontal edges replaced by sinusoids of varying frequency. A dot was placed in one of eight horizontal locations within the rectangle, and the observers were asked to judge whether the dot was to the left or right of the "perceived local center." Each observer made at least 30 observations for each experimental condition. Near the core of the stimulus, about half of the responses should be "left" and about half should be "right." Plotting the proportion of one of the responses against position gives an empirical function whose derivative can be compared against the actual medialness function for the stimulus. Burbeck and Pizer (1994) suggest the use of a probit model to estimate the response function.

The model used in Chapter 6 specified the likelihood function "backward." Typically, the likelihood function is found by specifying the distribution of the data given the parameters; in this case, the likelihood was found by specifying the distribution of the parameters given the data. To specify the distribution of the data given the parameters, it is necessary to specify a distribution over the space of images given the location of landmarks. It is intuitively simpler to specify the probable location of landmarks given an image.

In chapter 6, the likelihood function is a product of the features observed at each landmark. Other functions of the observed features may be more appropriate; the function used in Chapter 6 has the flavor of assuming that the features observed at each landmark are independent. The independence assumption may be reasonable if the landmarks are sufficiently far apart, where distance, as usual, depends on the scale of measurement. For points close together, however, the features are clearly not independent. In future work, it will be important to consider how to capture the dependencies among features in the likelihood function. The methods developed in Blom et al. (1993) provide preliminary insight into these issues by modeling the variability introduced into features by adding noise to an image. Further insight can be gained by considering the variability introduced by random affine transformations of an image.

Bibliography

Amit, Y., U. Grenander, and M. Piccioni. 1991. Structural image restoration through deformable templates. *Journal of the American Statistical Association* 86 (June): 376–387.

Amit, Y. and A. Kong. 1994. Graphical templates for image matching. Technical Report No. 373, Department of Statistics, University of Chicago.

Ballard, D. and C. Brown. 1982. *Computer Vision.* Englewood Cliffs, NJ: Prentice-Hall.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* 36, no. 2: 192–236.

Besag, J. 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B* 48, no. 3: 259–302.

Besag, J., J. York, and A. Mollié. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, no. 1: 1–59.

Blom, J., B. ter Haar Romeny, A. Bel, and J. Koenderink. 1993. Spatial derivatives and the propagation of noise in Gaussian scale space. *Journal of Visual Communication and Image Representation* 4 (March): 1–13.

Blum, H. and R. Nagel. 1978. Shape description using weighted symmetric axis features. *Pattern Recognition* 10, no. 3: 167–180.

Bookstein, F. 1986. Size and shape spaces for landmark data in two dimensions. *Statistical Science* 1 (May): 181–242.

Bookstein, F. 1991. *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge: Cambridge University Press.

Burbeck, C. and S. Pizer. 1994. Object representation by cores. Tehnical Report TR94-160, Department of Computer Science, University of North Carolina, Chapel Hill.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (November): 679–698.

Coggins, J. and C. Huang. 1993. Defining optimal feature sets for segmentation by statistical pattern recognition. In *Mathematical Methods in Medical Imaging II, Proceedings of the International Society for Optical Engineering Vol. 2035 held in San Diego, CA, 15-16 July 1993*, edited by D. Wilson and J. Wilson, 80–88. Bellingham, WA: SPIE.

Cootes, T., A. Hill, C. Taylor, and J. Haslam. 1993. The use of active shape models for locating structures in medical images. In *Information Processing and Medical Imaging: Proceedings of the 13th International Conference held in Flagstaff, AZ, 14-18 June 1993*, edited by H. Barrett and A. Gmitro, 33–47. Lecture Notes in Computer Science, No. 687. Berlin: Springer-Verlag.

Eberly, D. 1994. Geometric Methods for Analysis of Ridges in N-Dimensional Images. Ph.D. dissertation, Department of Computer Science, University of North Carolina, Chapel Hill.

Florack, L. 1993. The syntactical structure of scalar images. Ph.D. dissertation, Faculteit Geneeskunde, University of Utrecht, Utrecht, Netherlands.

Fritsch, D. 1993. Registration of radiotherapy images using multiscale medial descriptions of image structure. Ph.D. dissertation, Department of Biomedical Engineering, University of North Carolina, Chapel Hill.

Fritsch, D., D. Eberly, S. Pizer, and M. McAuliffe. 1995. Stimulated cores and their applications in medical imaging. Preprint.

Geman, D. 1988. Random fields and inverse problems in imaging. In *École d'Été de Probabilités de Saint-Flour XVIII*, edited by P. Hennequin, 117–193. Lecture Notes in Mathematics, No. 1427. Berlin: Springer-Verlag.

Geman, S. and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (November): 721–741.

Geman, D. and B. Gidas. 1991. Spatial statistics and digital image analysis. In *Image Analysis and Computer Vision* by the National Research Council. Washington, DC: National Research Council, 9–36.

Geyer, C. and E. Thompson. 1992. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B* 54, no. 3: 657–699.

Gindi, G., M. Lee, A. Rangarajan, and I. Zubal. 1991. Bayesian reconstruction of functional images using registered anatomical images as priors. In *Information Processing and Medical Imaging: Proceedings of the 12th International Conference held in Wye, UK, 7-12 July 1991*, edited by A. Colchester and D. Hawkes, 121–131. Lecture Notes in Computer Science, No. 511. Berlin: Springer-Verlag.

Gindi, G., A. Rangarajan, M. Lee, P. Hong, and I. Zubal. 1993. Bayesian reconstruction for emission tomography via deterministic annealing. In *Information Processing and Medical Imaging: Proceedings of the 13th International Conference held in Flagstaff, AZ, 14-18 June 1993*, edited by H. Barrett and A. Gmitro, 322–338. Lecture Notes in Computer Science, No. 687. Berlin: Springer-Verlag.

Golub, G. and C. van Loan. 1989. *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press.

Gonzalez, R. and R. Woods. 1992. *Digital Image Processing*. Reading, MA: Addison-Wesley.

Goodall, C. 1991. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B* 53, no. 2 : 285–339.

Grenander, U. and K. Manbeck. 1993. A stochastic shape and color model for defect detection in potatoes. *Journal of Computational and Graphical Statistics* 2 (June): 131–151.

Grenander, U. and M. Miller. 1994. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society, Series B* 56, no. 4: 549–603.

Hanson, K. 1993. Lumen reconstruction from two angiograms. Preprint.

Haring, S., M. Viergever, and J. Kok. 1993. A multiscale approach to image segmentation using Kohonen networks. In *Information Processing and Medical Imaging: Proceedings of the 13th International Conference held in Flagstaff, AZ, 14-18 June 1993*, edited by H. Barrett and A. Gmitro, 212–224. Lecture Notes in Computer Science, No. 687. Berlin: Springer-Verlag.

Higdon, D., V. Johnson, T. Turkington, and J. Bowsher. 1994. Fully Bayesian reconstruction of positron emission tomography data. Preprint.

Hoffman, E., P. Cutler, W. Digby, and J. Mazziotta. 1990. 3-D phantom to simulate cerebral blood flow and metabolic images for PET. *IEEE Transactions on Nuclear Science* 37 (April): 616–620.

Jackway, P. 1992. Morphological scale-space. In *Proceedings of the 11th IAPR International Conference on Pattern Recognition held in the Hague, the Netherlands, August 30-September 3, 1992*, Volume 3, Los Alamitos, CA: IEEE Computer Society Press, 252–255.

Jackway, P. 1993. Multiscale image processing: A review and some recent developments. *Journal of Electrical and Electronics Engineering* 13 (June) 88–98.

Johnson, V., W. Wong, X. Hu, and C. Chen. 1991. Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (May): 413–425.

Johnson, V., J. Bowsher, R. Jaszczak, and T. Turkington. 1993. Analysis and reconstruction of medical images using prior information. To appear in *Case Studies in Bayesian Statistics*. New York: Springer-Verlag.

Johnson, V. 1994. A model for segmentation and analysis of noisy images. *Journal of the American Statistical Association* 89 (March): 230–241.

Karr, A. 1991. Statistical models and methods in image analysis: a survey. In *Statistical Inference in Stochastic Processes*, edited by N. Prabhu and I. Vasawa, 1–34. New York: Marcel Dekker.

Kendall, D. 1984. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society* 16 (March): 81–121.

Kendall, D. 1989. A survey of the statistical theory of shape. *Statistical Science* 4 (May): 87–120.

Koenderink, J. 1984. The structure of images. *Biological Cybernetics* 50 (August): 363–370.

Leahy, R. and X. Yan. 1991. Incorporation of anatomical MR data for improved functional imaging with PET. In *Information Processing and Medical Imaging: Proceedings of the 12th International Conference held in Wye, UK, 7-12 July 1991*, edited by A. Colchester and D. Hawkes, 105–120. Lecture Notes in Computer Science, No. 511. Berlin: Springer-Verlag.

Lele, S. 1989. Some comments on coordinate free and scale invariant methods in morphometrics. Mimeo series #2006, Department of Statistics, University of North Carolina at Chapel Hill.

Lifshitz, L. and S. Pizer. 1990. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (June): 529–541.

Lindeberg, T. 1990. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (March): 234–254.

Lindeberg, T. 1994. *Scale-Space Theory in Computer Vision*. Boston: Kluwer Academic Publishers.

Llacer, J., B. ter Haar Romeny, L. Florack, and M. Viergever. 1993. Representation of medical images by visual response functions. *IEEE Engineering in Medicine and Biology* 12 (March): 40–47.

Llacer, J., B. ter Haar Romeny, and M. Viergever. 1992. The use of geometric prior information in Bayesian tomographic image reconstruction: a preliminary report. In *Mathematical Methods in Medical Imaging, Proceedings of the International Society for Optical Engineering Vol. 1768 held in San Diego, CA, 23-24 July 1992*, edited by D. Wilson and J. Wilson, 82–96. Bellingham, WA: SPIE.

Marr, D. and E. Hildreth. 1980. Theory of edge detection. *Proceedings of the Royal Society of London, Series B* 207 (29 February): 187–217.

Morse, B., S. Pizer, and A. Liu. 1993. Multiscale medial analysis of medical images. In *Information Processing and Medical Imaging: Proceedings of the 13th International Conference held in Flagstaff, AZ, 14-18 June 1993*, edited by H. Barrett and A. Gmitro, 112–131. Lecture Notes in Computer Science, No. 687. Berlin: Springer-Verlag.

Phillips, D. and A. Smith. 1994. Bayesian faces via hierarchical template modeling. *Journal of the American Statistical Association* 89 (December): 1151–1163.

Pizer, S., D. Eberly, and D. Fritsch. 1994. Objects, boundaries, and medical image analysis. Preprint.

Ripley, B. 1988. *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.

Ripley, B. 1991. The use of spatial models as image priors. In *Spatial Statistics and Imaging*, edited by A. Possolo, 309–340. IMS Lecture Notes, Monograph Series, Vol. 20. Hayward, CA: Institute of Mathematical Statistics.

Ripley, B. and A. Sutherland. 1990. Finding spiral structures in images of galaxies. *Philosophical Transactions of the Royal Society of London, Series A* 322 (September): 477–485.

Rosenfeld, A., ed. 1984. *Multiresolution Image Processing and Analysis.* Springer Series in Information Sciences, vol. 12, New York: Springer-Verlag.

Short, T. H. 1993. An algorithm for the detection and measurement of rail surface defects. *Journal of the American Statistical Association* 88 (June): 436–440.

Small, C. 1988. Techniques of shape analysis on sets of points. *International Statistical Review* 56 (December): 243–258.

Staib, L. and J. Duncan. 1992. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (November): 1061–1075.

Stoyan, D., W. Kendall, and J. Mecke. 1987. *Stochastic Geometry and its Applications.* Chichester, UK: John Wiley & Sons.

ter Haar Romeny, B., L. Florack, J. Koenderink, and M. Viergever. 1991. Scale-space: its natural operators and differential invariants. In *Information Processing and Medical Imaging: Proceedings of the 12th International Conference held in Wye, UK, 7-12 July 1991*, edited by A. Colchester and D. Hawkes, 239–255. Lecture Notes in Computer Science, No. 511. Berlin: Springer-Verlag.

van den Boomgaard, R. 1992. The morphological equivalent of the Gauss convolution. *Nieuw Archief Voor Wiskunde* 10 (November): 219–236.

van den Elsen, P., E. Pol, and M. Viergever. 1993. Medical image matching—A review with classification. *IEEE Engineering in Medicine and Biology* 12 (March): 26–39.

Whitaker, R. and S. Pizer, 1993. A multi-scale approach to nonuniform diffusion. *CVGIP: Image Understanding* 57 (January) 99-110.

Witkin, A. 1983. Scale-space filtering. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence, held in Karlsruhe, Germany, August 8-12, 1983*, edited by A. Bundy, 1019–1022. Los Altos, CA: W. Kaufmann.

Wolberg, G. 1990. *Digital Image Warping.* Los Alamitos, CA: IEEE Computer Society Press.

Yuille, A. and T. Poggio. 1986. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (January): 15–25.

Biography

Alyson Gabbard Wilson was born in Cincinnati, Ohio, on May 4, 1967. She received her B.A. in Mathematical Sciences from Rice University in Houston, Texas, in May 1989, graduating summa cum laude, Phi Beta Kappa. She received her M.S. in Statistics from Carnegie Mellon University in Pittsburgh, Pennsylvania, in August 1990.

Her graduate studies were funded by a National Science Foundation Graduate Research Fellowship, a James B. Duke Fellowship in Statistics, and a Clare Boothe Luce Fellowship for Women in Science. Ms. Wilson was a finalist for the Gertrude Cox Scholarship, sponsored by the American Statistical Association's Committee on Women in Statistics and the Caucus for Women in Statistics. She is an Associate Editor for the International Society for Bayesian Analysis newsletter and has served as a reviewer for the *Journal of the American Statistical Association.* She is a member of the American Statistical Association, the International Society for Bayesian Analysis, and the Association for Women in Science.

Ms. Wilson has authored or co-authored the following articles: Remarks on the Gibbs sampler and its implementation on a parallel machine (1992), Comparison of breathing comfort during weaning with two ventilatory modes

(1994), Cognitive factors affecting subjective probability assessment (1994), Using features to model prior structural information (1994), Priors on scale-space templates (1994), Models for shape deformation (1995).