Chapter 5

Empirical analysis

5.1 Introduction

Chapters 3 and 4 present multiscale, geometry-limited diffusion as a means of image segmentation and feature or structure extraction [1]. These techniques follow from a strategy of using the local homogeneity of geometric structure to segment images. After following these arguments and seeing the examples presented in this dissertation, the question remains, "How good is it?"

How does one answer such a question? Previous chapters pursue two different approaches for studying the behavior of the variable-conductance diffusion processes. The first approach is to examine the mathematics that describe these processes. The mathematics help to understand the local behavior under specific circumstances, such as the enhancement of edges (local maxima in gradient) in anisotropic diffusion. The mathematics also indicate some very general properties that apply to a wide range of circumstances; the geometric invariance associated with geometry-limited diffusion is an example. However, the various simplifications and assumptions that underlie these analyses can limit the usefulness of the results. In the case of edges enhancement, for instance, one could ask, "What happens at places *near* edges? Is the behavior desirable?"

Some of the doubts that stem from the simplicity of the mathematical models are addressed by looking at the way various processes behave when they are applied to "real-life" examples. The strategy is simple: *Try it out and see if it works.* This mode of research is a staple of conventional research in computer vision and image processing. First, construct a mathematical model (complete with the appropriate simplifications), and from that model derive an algorithm. Second, try the algorithm on some examples to see if it works. Indeed, I adopt this strategy throughout most of this dissertation.

One of the shortcomings of studying specific examples is that those examples often undergo only a qualitative or subjective analysis. The results *look* better, but *are* they better? Even in the presence of a quantitative comparison between results (presumably, using some kind of metric), one can still question the generality of such results. "It 'worked' for this image, but will it work for others?"

The purpose of this chapter is to conduct an objective analysis of the effectiveness of the processes described in Chapts. 2, 3, and 4 for image segmentation and structure extraction (Hereafter all of the diffusion processes, including Gaussian blurring, are called generically "processes"). First, I will propose a new experimental framework for conducting objective analyses of image processing and computer vision algorithms. The strategy is to treat image processing as a statistical problem

[1] I hereafter refer to what are commonly called "visual features" as "structures". I reserve the word "feature" to mean one element of a feature vector or a single value of multi-valued image as in Chapt. 4

and to use the statistics associated with groups of similar images to make some general statements about the behavior of a process. The focus is on the *aggregation* of image information. The detailed behavior of a process at specific places in images is lost in order to obtain *general indications* of the overall behavior of a process. Most importantly, the goal of this research is to avoid looking at lots of images and making subjective comparisons. Images are reduced to numbers.

The framework that I will propose offers strategy for constructing experiments in order to understand image processing and computer vision algorithms. It has several basic components:

- A definition of "truth" that derives from a test image.
- A stochastic model of noise which modifies the test image.
- An experimental design that requires the simulation of sets noisy images.
- A series of metrics that quantify the behavior of an algorithm with respect to various noise processes.

After laying out the basic framework, I will use it to conduct a series of experiments. These experiments will study the behavior of the geometry-limited diffusion process over a range of different test images, noise processes, and parameter values. The goal is twofold. First, I want to gather some quantitative evidence to support the observation that variable-conductance diffusion is better than Gaussian blurring. Second, I want to use the quantitative analysis to explore the parameter space of these processes and gain a better understanding of the limitations of each process. This analysis will focus primarily on developing evidence for (or against) the following assertion.

Assertion 1 *The geometry-limited diffusion method described here offers a* reliable *and* accurate *means of measuring the geometric structure of images.*

The remainder of this chapter is organized as follows. I will motivate the development of this experimental framework and define some of the terms of Assertion 1. That discussion will be quite general. It will start with a definition of noise and then propose Monte Carlo simulation as a means of studying the noise sensitivity of a process.

Given this general description of the experimental framework, the challenge will be to construct metrics that quantify relevant image properties. Instead of a single, all-encompassing metric, I will quantify a number of different aspects of images using different metrics. The metrics operate by making comparisons within sets of noisy or distorted images and by comparing those images to a goal. Each of the metrics has some drawbacks and limitations that will become apparent as results are presented. My position throughout this chapter is that these metrics constitute evidence, as opposed to certain proof, of certain kinds of general behavior. I will then define several different types of noise that are relevant to a wide range of computer vision problems. Each type of noise is described as a "process" that is used to stochastically generate sets or ensembles of images that share the same noise characteristics.

The experimental framework will be used first to study the effects of Gaussian blurring in the presence of uncorrelated noise on a single test image. I will then examine the anisotropic and multiscale diffusion processes under the same circumstances. Successive sections will study the effects of different test images and different types of noise in order to make more general statements about the relative merits of the different diffusion processes. Although I will describe the experimental framework in somewhat general terms, its application in this chapter is very specific. In particular, I will not examine the diffusion of multi-valued images, and I will discuss only the zero-order diffusion processes. The metrics, also, will be quite specific and will focus primarily on the detection of edges or regions of homogeneous luminance. I intend to offer these specific quantitative results as evidence that the entire family of diffusion processes are reliable and accurate.

5.2 Philosophy

For a sufficiently well-defined task and a given image, there is a *correct answer*. Defining this correct answer for any particular image could be quite difficult. However, most of the results in this chapter come from images in which the correct answer is arguably straightforward.

In this analysis I am particularly concerned with the behavior of different methods in the presence of noise. *Noise* is a family of image variations that are not relevant to a particular visual task, i.e. variations of an image that have the same correct answer.

Assumption 1 *Given an image and a visual task relating to that image there is a population of images that vary in irrelevant ways, i.e. a set of images that are measurably different in their luminance functions but have the same correct answer for that task.*

Definition 2 *Variations in the image which do do not affect the correct answer of a particular task are called* noise*.*

For example, imagine that a person (a visual system) is given a photograph (an image) and asked whether or not there is cat in the picture (task). There are a great many ways that the pigments (luminance) associated with different points on that photograph could vary. Some of these variations affect the correct answer, e.g. the answer could go from "cat" to "no cat". Any image variation which does not affect whether or not there is a cat in the picture (according to the correct answer) is called noise. If the cat-or-no-cat decision of a person is affected by such variations, then the person is *sensitive* to noise. The "fuzziness" of such relationships is also an issue. One could argue that such decisions are not discrete and that I should discuss the likelihood that there is a cat in the picture. Given such a graded task, the notion of noise would also be graded in order to capture the *degree* to which different variations in color affect this probability. Indeed, the analysis in this chapter quantifies degrees or levels of sensitivity and accuracy.

A *reliable* processing method should produce similar results for different images that vary only in their noise content. An *accurate* process should produce sufficiently different results for images that vary in their relevant content.

In general, the relevance of different kinds of variations in images depends on the image, the task, and the assumptions of the visual system performing that task. These notions can be rather broad, and they do not lend themselves to precise definition. For the experiments of this chapter, the relatively narrow goals of the geometry-limited diffusion process simplify the problem. The diffusion process described here does not aspire to anything so ambitious as finding cats in photographs. Instead, it describes a means of making reliable differential measurements and image segmentations in the presence of irrelevant, small-scale structure. Instead of trying to analyze noise in the most general sense, I deal with some simple notions of noise. These noise processes are reasonable approximations to the kinds of irrelevant variations that exist in a number of visual problems associated with medical image analysis.

The strategy (depicted in Fig. 5.1) for studying the geometry-limited diffusion process is to identify certain types of noise to which we would like an algorithm to be insensitive. Describe these noise processes statistically so that there is a population of noisy images, $\mathcal{P}$, which is characterized by a given degree of variation as well as an expected value image, maximum likelihood, etc. Draw from that population (using some pseudo-random process) a sample $\mathcal{S}$, which is an ensemble of N images. Treat each image from this sample separately using the geometry-limited diffusion process (or any process that is of interest), $\mathcal{G}$. Study the statistics of the resulting sample, which is a collection of processed images $\mathcal{G}(S)$. Measure the effects of various parameters by examining the behavior of a set of *metrics* that are constructed to capture interesting properties in the processed sample. In this chapter the metrics are constructed to quantify the ideas of *accuracy* and *reliability* which are expressed in Assertion 1. This analysis is a Monte Carlo approach, which has as its input

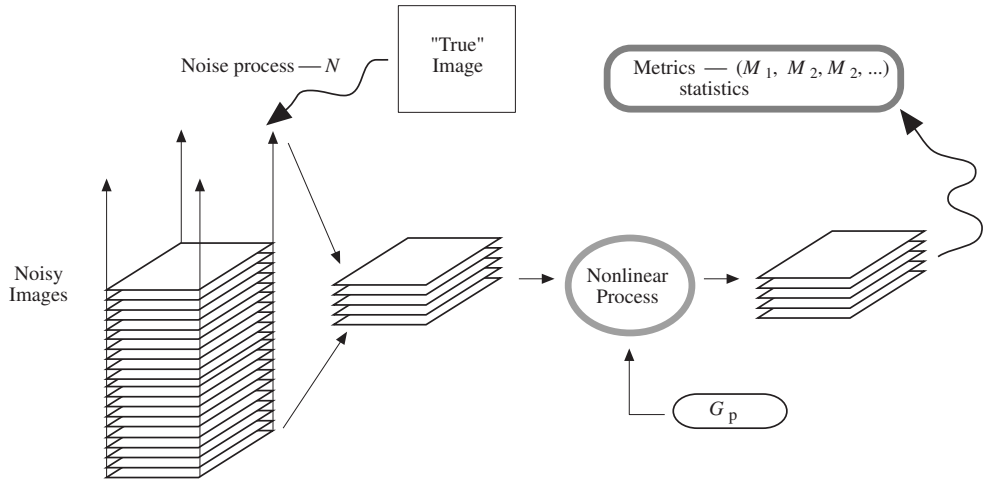


Figure 5.1: Samples from a population of noisy images are processed separately so that accuracy and sensitivity can be measured based on the statistics from the resulting images.

stochastically generated images. Each image is drawn from a population of images which exhibits the noise properties that are under investigation.

The Monte Carlo analysis provides a means of quantifying the differences in processes without relying on the subjective impressions that typically result from qualitative examination of processed images. The goal of this empirical analysis is to create an objective evaluation of a process $\mathcal{G}$. This process might have a set of parameters p that control its behavior and determine the set of output images $\mathcal{G}(S)$. The process and the associated parameters are then denoted $\mathcal{G}_p$. In geometry-limited diffusion the conductance parameter k and the starting scale s_0 are two such parameters. Using the Monte Carlo analysis, a researcher would analyze the behavior of $\mathcal{G}_p$ with respect to the parameter space by conducting a series of experiments with simulated noise data. The results of these experiments is a set of numbers that vary with different noise processes and process parameters. The key to obtaining useful results is constructing the metrics so that they capture relevant information.

5.3 Metrics

The metrics for analyzing a process should reflect the task for which the process is to be used. Suppose, for instance, that the geometry-limited diffusion were used as a preprocessing step in an algorithm that measures the width of some particular object in an image. Then the metrics should indicate the accuracy and reliability of the width measurements that result from the diffused images. Instead of a specific task, this dissertation addresses the more general problem of segmentation. Thus, I propose several very simple metrics that reflect the typical issues that a faced by a number of segmentation techniques. These metrics are neither unique nor optimal. They are, however, indicative of the kinds of measurements that are relevant to image segmentation, and they serve to give some insights into the behavior of the geometry-limited diffusion process.

The metrics quantify two distinct aspects of the processed sample $\mathcal{G}_p(S)$. The first is the reliability, i.e. variability within this ensemble. If the sample S consists of images which differ only in their noise content and the process $\mathcal{G}_p$ is not sensitive to noise (if it is reliable), the processed sample should consist of images which all "look alike" in some respect. The second aspect of these metrics is the overall quality of the processed sample. Does the sample capture the relevant structure of the underlying population? Or does it introduce some consistent bias in the results? As an extreme example, consider a process which creates a field of zeros for any input image and any choice of

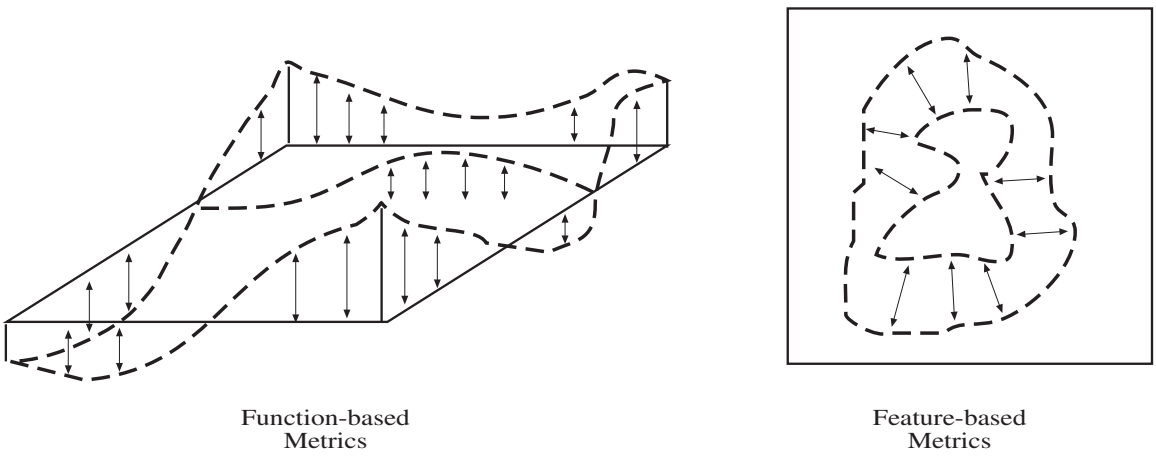


Figure 5.2: The two categories of metrics: Function-based metrics compare functions over the image space point-by-point, and structure-based metrics compare sets of structures.

parameters. Such a process is very consistent (reliable?) but certainly not very accurate (for any kind of interesting visual task). In this analysis I use a number of different metrics which give different information about the behavior of a process $\mathcal{G}_p$. Some of these metrics capture reliability or accuracy separately while others combine these concepts into a single value.

These concepts of reliability and accuracy become concrete definitions by their application to specific *image representations*. The strategy is to represent an image as a set of numbers that are relevant to the task at hand. In this way metrics can become conventional statistical measurements (such as the mean and variance) of the numbers that are used to represent the information in an image. For example, a closed boundary, B, in two dimensions can be constructed by a linear combinations of basis functions, that is,

$$B = a_1 g_1 + a_2 g_2 + \ldots,$$

where the functions $g_1, g_2, \ldots$ form the *basis*. Suppose the segmentation problem were simplified to the problem of finding the shape of a light object on a dark background. Then the coefficients, $a_1, a_2, \ldots$, could serve as a description of the boundary of that object. The individual elements of this description could lend themselves to a statistical analysis. Each image w_i in the ensemble $\mathcal{G}(S)$ has an image description, $a_1^{(i)}, a_2^{(i)}, \ldots$, and each individual element of the description has, among other measures, a mean and variance that reflect the statistical properties of the ensemble. Thus, the variance, which measures the deviations of the array, $\vec{a}$ across the ensemble of images, could be used to quantify the sensitivity of $\mathcal{G}$. The deviation of a central measure of the ensemble, such as the mean, from the goal, could indicate the accuracy of the process $\mathcal{G}$.

The choice of image description is a critical factor in the development of useful metrics. The metrics in this chapter fall into two basic categories that are distinguished by the particular image description that is used (Fig. 5.2). The first is the category of *function-based* metrics. Function-based metrics treat one or more functions that are defined over the image space, and they entail calculations *point by point* to obtain a metric value at each point $\vec{x}$ in the image. Thus, the image representation is the set of pixel values. The pointwise calculations are averaged (with an integral) over the entire image space I in order to produce a single value. Function-based metrics quantify the *differences among functions defined over the image space*.

The second category of metrics I call *structure-based* metrics. Structure-based metrics quantify differences in sets of image points (subsets of the image space) which represent interesting *visual*

structures such as edges or creases. In two-dimensional images, visual structures are points or curves in the plane. Structure-based metrics quantify *the difference between two or more sets of points or curves in the plane.*

5.3.1 Function-based metrics

The motivation for using function-based metrics is as follows. Some images are relatively easy to segment if the purpose of the segmentation is precisely defined. For instance, if I wish to segment an image based on brightness or luminance, then an image that has patchwise-constant luminance patches (and no noise) is easy to segment. Because the output of a variable-conductance diffusion process is an image or set of images, one measure of the usefulness of that image is to compare it point-by-point (pixel-by-pixel) to a goal. This is reasonable if the goal should has two properties. First it should be easy to segment—piecewise constant (in whatever property) for instance. Second, the "easy segmentation" of the goal should be the correct or desired segmentation.

The function-based metrics in this chapter are meant as rough indicators of how well processed images lend themselves to certain kinds of segmentation paradigms. The strategy of comparing processed images point-by-point to a well-defined goal is relevant to a number of common segmentation strategies. One example is *thresholding*. In some applications, interesting segments are found by choosing all of those points in an image that have values greater than T (or lie between two values, T_1 and T_2). Such methods can be useful, but they often fail in the presence of noise. In order to quantify the effectiveness of thresholding, compare processed images point by point to another image which contains segments that have values greater than T in all of the "right" places.

Another related class of segmentation techniques are those developed in the context of statistical pattern recognition. Typically, these methods compare the positions of pixels in a feature space and classify these pixels based on their proximity to the nearest cluster. Each cluster represents one of a number of predefined classes. In this case, the function-based goal should contain homogeneous patches with feature values that are the "centers" of the appropriate clusters. The function-based metrics indicate how tightly the positions of pixels in the processed image cluster around these centers.

I use three different function-based metrics for the experiments in this chapter. Each of these metrics involves some averaging over the processed samples $\mathcal{G}_p(S)$ as well as the image space I. Thus, the computation of these metrics requires two different kinds of averages: averages over sets or ensembles of images (such averages map sets of images to a single image), and averages of pixel values over the image space (they map one image to a single number). I denote an average or expected value using angle brackets, $\langle\cdot\rangle$. The set over which the average is taken is denoted as a subscript. The average image from an ensemble $\mathcal{G}(S)$ is denoted

$$\langle w \rangle_{w \in \mathcal{G}(S)} = \frac{1}{|\mathcal{G}(S)|} \sum_{w \in \mathcal{G}(S)} w,$$

where $w : I \mapsto \mathbb{R}$ is an image, and the absolute value $|\cdot|$ of a set is the *cardinality* or number of elements in that set. The average value of an image w is denoted,

$$\langle w(\vec{x}) \rangle_{\vec{x} \in I} = \frac{1}{|I|} \int_I w(\vec{x}) \mathrm{d}\vec{x}.$$

In discrete images the integral becomes a sum, and $|I|$ is the number of pixels in the image.

I use this notation to define the three function-based metrics. The first is the variance, $\mathcal{M}_{\text{var}}$; it

captures the degree of variation within an image set. For a processed processed sample $\mathcal{G}(S)$,

$$\mathcal{M}_{\text{var}} = \left\langle \langle w(\vec{x})^2 \rangle_{w \in \mathcal{G}(S)} - (\langle w(\vec{x}) \rangle_{w \in \mathcal{G}(S)})^2 \right\rangle_{\vec{x} \in I}. \tag{5.1}$$

This measure is the sample variance of w at each image point averaged over the entire image. The function w could be the image intensity, any one of a collection of geometric features, or some function that is computed from the derivatives of those features (the dissimilarity, for instance). I use this metric as an indicator of reliability (lower values mean greater reliability). The reasoning is that if a process is reliable in the presence of noise, it should produce consistent results across a noisy population, and the variance of the processes sample should be low.

The variance metric $\mathcal{M}_{\text{var}}$ might give some idea of how reliable is a process in the presence of noise, but it says nothing of accuracy. To measure accuracy, define another metric:

$$\mathcal{M}_{\text{mean}} = \left\langle \left(\langle w(\vec{x}) \rangle_{w \in \mathcal{G}(S)} - w_g(\vec{x}) \right)^2 \right\rangle_{\vec{x} \in I}^{\frac{1}{2}}. \tag{5.2}$$

The image w_g is the "goal" or "truth", which is defined as an easily segmentable image which produces the "correct" segmentation. This metric can be used to compare measurements other than luminance. In such a circumstance the "goal" must be represented using those same features.

In order to better understand these metrics, consider the following example. A sample of images S consists of N images that are derived from a single image by the addition of random noise of mean zero. The metric $\mathcal{M}_{\text{mean}}$ applied directly to S is zero, which is perfect. In this case the processed sample is $\mathcal{G}(S) \equiv S$, and I call such a process the *identity process*. In the case of zero-mean, additive random noise, the metric $\mathcal{M}_{\text{mean}}$ represents a kind of *bias* in the the process $\mathcal{G}$. That is, it represents the systematic error in the process sample $\mathcal{G}(s)$. For zero-mean random noise, the identity process has no bias. If we considered only the metric $\mathcal{M}_{\text{mean}}$, we would be led to believe that the identity process performs quite well, but it does not. The set of noisy images thats result from the identity process has a particularly high value of $\mathcal{M}_{\text{var}}$. In the case of additive noise, the identity process introduces no bias and offers a processed sample variance equal to that of the noise. The relationship between accuracy and reliability will prove to be important; an analysis of Gaussian blurring will show that it improves the variability within the processed ensemble at the cost of introducing a bias.

This example indicates that both accuracy and reliability are important when considering the overall performance of a process. At times juggling both numbers, $\mathcal{M}_{\text{var}}$ and $\mathcal{M}_{\text{mean}}$, can be cumbersome. This is especially true when studying the behavior of a process with respect to some multidimensional parameter space. Once again, the focus is on consolidating information. Thus, I use a metric which measures the average squared distance between images (rms distance):

$$\mathcal{M}_{\text{rms}} = \left\langle \langle (w(\vec{x}) - w_g(\vec{x}))^2 \rangle_{w \in \mathcal{G}(S)} \right\rangle_{\vec{x} \in I}^{\frac{1}{2}}. \tag{5.3}$$

This metric quantifies the expected differences from the goal without distinguishing between differences that result from variability in the processed sample or systematic biases. This metric is also called *rms error*.

Normalizing function-based metrics

The above metrics are more useful if they can be compared among various test images and noise processes. In addition, it would be valuable to be able to put different metrics on the same graph in order to analyze or compare them. To achieve this I employ the principles of *dimensionless* measures from Sect. 4.5.2. These metrics are expressed in a normalized form using the following normalization

conventions.

$$\mathcal{M}^{\mathrm{n}}_{\mathrm{var}} = \frac{\mathcal{M}_{\mathrm{var}}}{\sigma_N^2} \tag{5.4}$$

$$\mathcal{M}^{\mathrm{n}}_{\mathrm{mean}} = \frac{\mathcal{M}_{\mathrm{mean}}}{\langle w_g^2(\vec{x})\rangle^{\frac{1}{2}}_{\vec{x}\in I}} \tag{5.5}$$

$$\mathcal{M}^{\mathrm{n}}_{\mathrm{rms}} = \frac{\mathcal{M}_{\mathrm{rms}}}{\langle w_g^2(\vec{x})\rangle^{\frac{1}{2}}_{\vec{x}\in I}} \tag{5.6}$$

For these normalizations, σ_N^2 is the pixel-by-pixel variance of the noise, and the expression $\langle w_g^2(\vec{x})\rangle^{\frac{1}{2}}_{\vec{x}\in I}$ is the l_2 norm of the test image (truth). Unless otherwise noted, these metrics are always normalized in this fashion, and the superscript "n" is not used.

5.3.2 Structure-based metrics

The strategy of comparing images point by point has some limitations. If the goal of the geometry-limited diffusion process is to provide good segmentations or facilitate the extraction of interesting visual structures from an image, the process could accomplish this without reproducing or restoring the underlying luminance function (or some other dense feature map). Two images that are very different in terms of their values at every point could have visual structures (edges, corners, etc.) at exactly the same locations.

This observation motivates a category of metrics that compare *sets of structures* in images. Creating a metric which compares sets of structures poses some challenges. One strategy is to try to determine the distance between individual structures. Given this, the challenge is how to "pair" or "match" structures from one image to the next. For instance, suppose that one is given the two sets of *corners* associated with two different images A and B. The distance between those two sets is sensitive to the particular convention that is used for associating one corner from image A to another corner in image B.

There are a variety of options for deciding this issue, but here I propose a very simple metric that is based on the idea that *one would like to have just the right number of structures in all of the right places.* This gives rise to the notion that some measurable structures can be "wrong" or "false" because they do not conform to the goal. Other structures are "correct" or "true", because they *do* coincide with structures in the goal. This remainder of this section proceeds by developing some precise definitions and then using these definitions to construct the appropriate metrics.

A *structure set* F is a subset of the image space, $F \subset I$. Image-space positions that are members of a specific structure set have an associated *visual structure*. Thus, we can talk about the set of "edges" associated with a particular image w. A *detector* is a mapping from an image function w to a structure set F. More generally, a detector could also be a mapping from a multi-valued image $\vec{W}$ to a structure set. A detector generates a set of image points from any given image or images. For example, an *edge detector* might require a single input image and return a list of points in the image space that are designated as "edges" or "edgels".

It is useful to represent such structure sets as images (mappings from the image domain to some range) which have at each point a boolean value that indicates whether or not that point is designated as a structure. For an image w and a detector D, the *structure image* h is

$$h : I \mapsto \{0, 1\},$$

where

$$h_{(D,w)}(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in D(w) \\ 0 & \text{otherwise} \end{cases} .$$

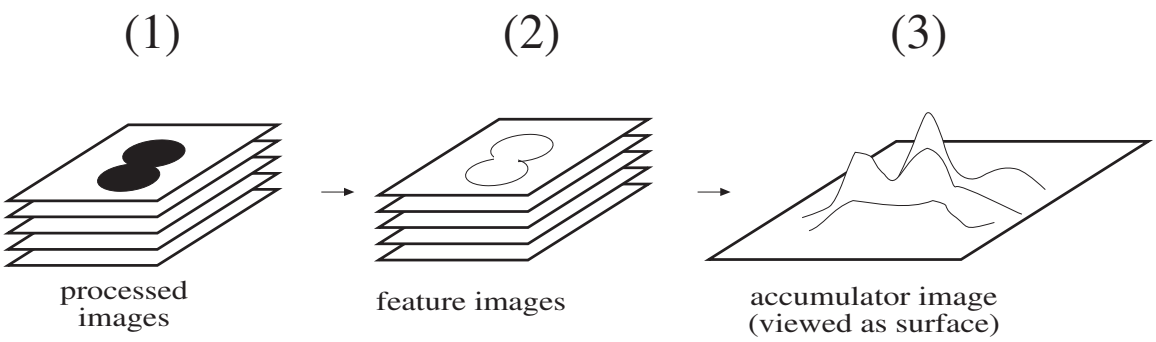


Figure 5.3: The accumulator image is formed by 1) processing sampled images, 2) constructing a structure image (using a detector) for each sample image, and 3) summing the structure images.

If D is an edge detector, then $h_{(D,w)}(\vec{x})$ would, in conventional computer vision terminology, be called the "edge map" of w.

The strategy for creating a structure-based metric is to evaluate a process by comparing the structure sets from processed noisy images $\mathcal{G}_p(S)$ with the structure set that results from applying the structure detector to the goal image, w_g. The purpose is to understand the statistical properties of such structure sets over the entire sample of images S. On the average how accurate, i.e., close to the goal, are the structure sets that result from the process $\mathcal{G}$? Also, how reliable are such structures sets? In order to answer questions this analysis *accumulates* structure images over the sample. For a detector D and a set of images S, define a structure accumulator $A_{(D,S)}$ as

$$A_{(D,S)}(\vec{x}) = \langle h_{(D,w)}(\vec{x}) \rangle_{w \in \mathcal{G}(S)}. \tag{5.7}$$

The accumulator image looks like a histogram over the image space. The value at each point indicates the average number of times a visual structure occurs at that point (Lin, Pizer and Johnson 1991).

In continuous images the accumulator as defined in (5.7) is probably not so useful, because points in the accumulator image would have values of $h(\vec{x}) > \frac{1}{N}$ only where the curves associated with the structure sets of different samples happen to intersect one another. In this case blurring this "spatial histogram" would account for the finite sample size; blurring the accumulator from a set of continuous images would generate a continuous distribution. In discrete images, which are the subject of study in this analysis, the individual pixels act like the "bins" that are associated with a finite-resolution histogram. The value at each bin indicates the likelihood that a structure would pass through that pixel for any one image in the population of noisy images.

The challenge is to define a metric (or metrics) that compares the accumulator to the goal in a way that captures the overall *usefulness* of the structure images produced by the process $\mathcal{G}$. There is a great number of issues to be considered. For example, a *good* set of structures should have the same *shape* as the goal. Spurious (or extra) structures are undesirable as are *missing* structures. A missing structure is probably less desirable than a structure that is present but not in precisely the "correct" location in the image space. The precise definition as well as the relative importance of each of these different aspects of structure-set quality depends on the application.

For instance, some applications might require *connectivity* of segment boundaries and a preservation of the overall shapes of those boundaries, while other applications could be particularly sensitive to the presence of false boundaries. The notion of boundary shape is particularly complex, and the problem of finding useful shape representations (and associated shape metrics) is an area of active research in computer vision and pattern recognition. Rather than dealing directly with that issue, I construct metrics that do not include an explicit shape representation. The practical implications

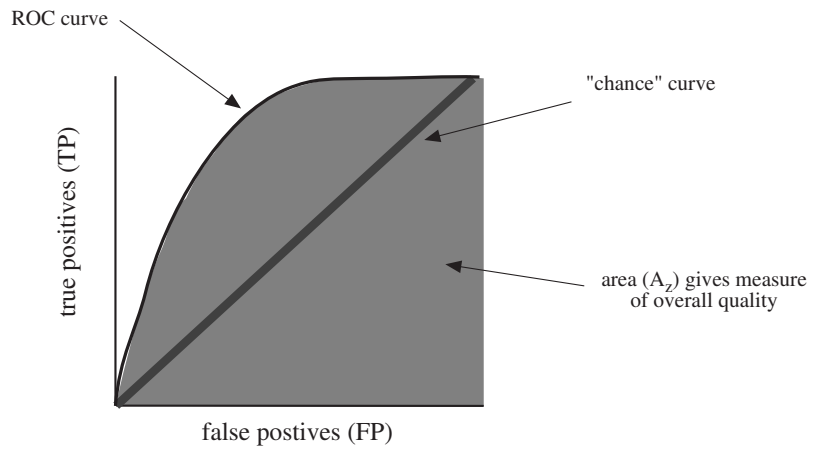


Figure 5.4: The graph of the expected number of true and false positives for a range of decision strategies is an ROC curve. The area under the curve, A_z, gives an overall measure of quality.

of this choice will become apparent in later sections.

For the analysis in this chapter, I simplify the problem somewhat by considering two important properties of structure sets. The first property is the extent to which structures in the goal are present in the accumulator. An ideal accumulator image should have values of 1.0 at *all* image space positions that are part of the goal. The second property is the extent to which structures that are present in the accumulator are also present in the goal. An ideal accumulator image should have values *only* at those positions which are part of the goal—the accumulator should not have spurious structures. These two properties are reminiscent of Receiver Operator Characteristic (ROC) analysis, which is often used to evaluate the properties of discriminants within statistical structure spaces in the presence of noisy structures (Macmillan 1991).

ROC analysis recognizes the relationship between the expected number of true and false estimates. For a given decision strategy one expects a certain number of true-positive (TP) and false-positive estimates (FP). Over a range of decision strategies one expects different values of TP and FP. A plot of TP versus FP is an *ROC curve* (Fig. 5.4). Typically these plots include a *chance curve*. The chance curve indicates the expected number of TP and FP estimates if one were to guess at the correct estimate without using any information about the particular sample. A perfect decision strategy would produce the largest possible number of TPs and *no* FPs. The overall effectiveness of various strategies can be compared by calculating the integral *under* the ROC curve, A_z. Better decision strategies have larger values of A_z.

With some generalization, this analysis can apply to the evaluation of accumulator images. Consider two structure sets: f_e, the estimate, and f_g, which is the goal. Evaluate the estimate f_e by deciding whether or not each member of this set (recall that a structure set is a subset of the image space) is in the goal f_g. Each point in f_e is either TP or FP depending on whether or not it is also in f_g. If each point in the structure set is equally important, it is reasonable to sum the number of TPs and FPs over the entire set to a obtain a single value. I define the number of true positives as

$$T = \int_{\{\vec{y} | (\vec{y} \in f_e \cap f_g)\}} \mathrm{d}\vec{x}, \tag{5.8}$$

and the number of false positives as

$$F = \int_{\{\vec{y} | (\vec{y} \in f_e \cap \bar{f_g})\}} \mathrm{d}\vec{x}, \tag{5.9}$$

where $\bar{f}_g$ indicates the complement of the image set f_g.

In discrete images the integrals are sums over all of the the positions of pixels that belong to the appropriate sets. One could construct an ROC curve by examining the values of T and F over a range of parameters that are relevant to the process $\mathcal{G}$. Later sections, for instance, will show that Gaussian blurring can reduce F at the cost of reducing T. Rather than decide an "optimal scale", the ROC analysis quantifies the range of T and F that Gaussian blurring has to offer. This analysis assumes that the cost of reducing spurious structures is the loss of useful structure information. In this way, *the relative rate at which useful information is lost* is the underlies the measure of process performance. The area under the curve quantifies the tradeoff between T and F, and gives an overall metric of performance.

As they stand, these definitions are unsatisfactory for comparing structure sets because they include no notion of tolerance. The accumulator image, for instance, is not a set of discrete image points but an image consisting of values from 0 to 1. Furthermore, the definition (5.8) counts a point as true-positive only if it coincides *exactly* with a point in the goal—there are no near misses. To address these problems, I use *fuzzy sets*. Fuzzy sets are graded membership functions that indicate *degree of membership* in a particular set. For images, a fuzzy set can be specified as a mapping from the image space to the interval $[0, 1]$.

I apply the notion of fuzzy sets to ROC analysis of structures by constructing fuzzy structure sets and generalizing the definitions of T and F to accommodate these sets. The accumulator, by definition, indicates the likelihood of membership, so it is treated as a fuzzy structure set estimate, f'_e. I construct a fuzzy goal, f'_g, from a the structure-set goal, f_g, by introducing a *tolerance* on the spatial extent of the structure-set. A tolerance provides a means of constructing a fuzzy set from a conventional structure set by extending the membership function on the basis of local proximity. I construct a membership function which drops from 1 to 0 monotonically with the distance to the nearest goal point. In these experiments, tolerances are constructed with a Gaussian falloff from the nearest structure point. If $\mathbb{D}(f_g)$ is the distance transform (Serra 1982) for the structure-set goal, then the fuzzy structure-set goal is

$$f'_g(\vec{x}) = \exp\left(-\frac{\mathbb{D}(f_g)^2}{2\sigma_t^2}\right), \tag{5.10}$$

where σ_t is the tolerance associated with the structure-set goal f_g (see Fig. 5.5). I choose a tolerance $\sigma_t = 2$ (in pixel units) for all of the analyses in this chapter. The sensitivity of the results to σ_t is examined in Sect. 5.7.2.

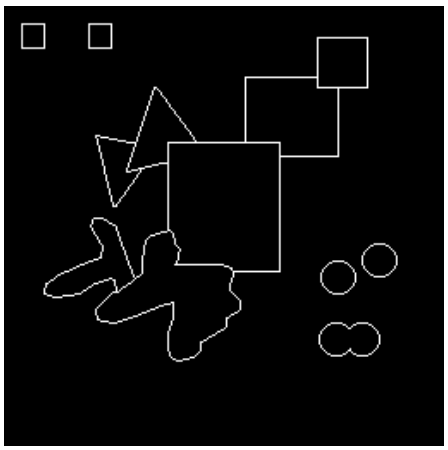
The definitions of (5.8) and (5.9) have fuzzy-set equivalent definitions. The intersections are replaced by multiplications, the complement of the goal, $\bar{f}'_g$, becomes $1 - f'_g$, and the integrals take place over the entire image space I. If f'_e and f'_g are the membership functions associated with the estimate and the goal respectively, then

$$T = \int_I f'_e(\vec{x}) f'_g(\vec{x}) \mathrm{d}\vec{x} \text{ and} \tag{5.11}$$

$$F = \int_I f'_e(\vec{x})(1 - f'_g(\vec{x})) \mathrm{d}\vec{x}. \tag{5.12}$$

The estimate f'_e is the accumulator image $A_{D,\mathcal{G}(S)}$, the average estimate from a detector D and a process $\mathcal{G}$ operating on a sample S.

The strategy for a structure-based analysis is to construct an accumulator image and then an ROC curve over a range of process parameters such as scale or conductance. The graphs of these ROC curves are used to analyze and compare different processes. Typically, the overall quality of an ROC curve is computed as the integral under the ROC curve, A_z. Instead of A_z, I use as the



A structure set shown as a binary image

A fuzzy structure with a tolerance

Figure 5.5: A fuzzy-structure-set goal (b) is constructed from the structure image (a) by using a Gaussian (standard deviation 2.0 pixel units) which falls off as distance to the nearest structure point.

metric $\mathcal{M}_{\text{ROC}}$ the complement of A_z, which the difference of a curve from the integral of a *perfect* process (the total area represented by the integration limits as in Fig. 5.6). I use the integral above the curve instead of A_z, because $\mathcal{M}_{\text{ROC}}$ is independent of the limits of integration, as long as those limits include all of the TP values that are less than perfect. The ROC metric, $\mathcal{M}_{\text{ROC}}$, serves as an overall estimate of the behavior of a process over that particular range of parameters. I calculate the integral using a linear interpolation between measured data points, where each point corresponds to a single T and F from (5.11) and (5.12) respectively. For a particular diffusion process, multiple values of T and F are obtained by varying the scale or time of that process.

Although this metric does not include an *explicit* representation of the shape, it does incorporate an *implicit* description of the shape of a structure set, and it does provide a very simple measure that quantifies small differences in shape. The integral of Eq. (5.11) is an inner product of two images, and it resembles a *template matching* approach to comparing shapes. The goal f'_g serves as a kind of fuzzy template, and the number total number of true positives T is a measure of the "fit" of the accumulator image to the template. This template approach to shape comparison is primitive relative to state-of-the-art research in the area of shape (Morse, Pizer and Liu 1993, Kimia and Zucker 1992, Brechbühler et al. 1992, Kimia, Tannenbaum and Zucker 1992, Pizer et al. 1991, Blum and Nagel 1978, Zahn and Roskies 1972). Thus, the ROC metric fails to capture a number of important shape properties such as rotational and translational invariance, part and sub-part relationships, and continuity of regions and boundaries. The consequences of these shortcomings are discussed in later sections as they become apparent in the experimental results. Despite these shortcomings, the T and F metrics are capable of quantifying small *distortions* in the shapes of region boundaries as long as these boundaries have not undergone gross changes such as translation or rotation. In this limited sense, the structure-based metrics do offer some insight into the effects of various processes on the shapes of region boundaries or structure sets.

5.3.3 The zero-order structure detector - edges

The definition of the structure-based metric $\mathcal{M}_{\text{ROC}}$ requires a structure detector D. In the case where diffusion is applied to the intensity the gradient magnitude serves as the dissimilarity ($D \equiv |\nabla f|$),

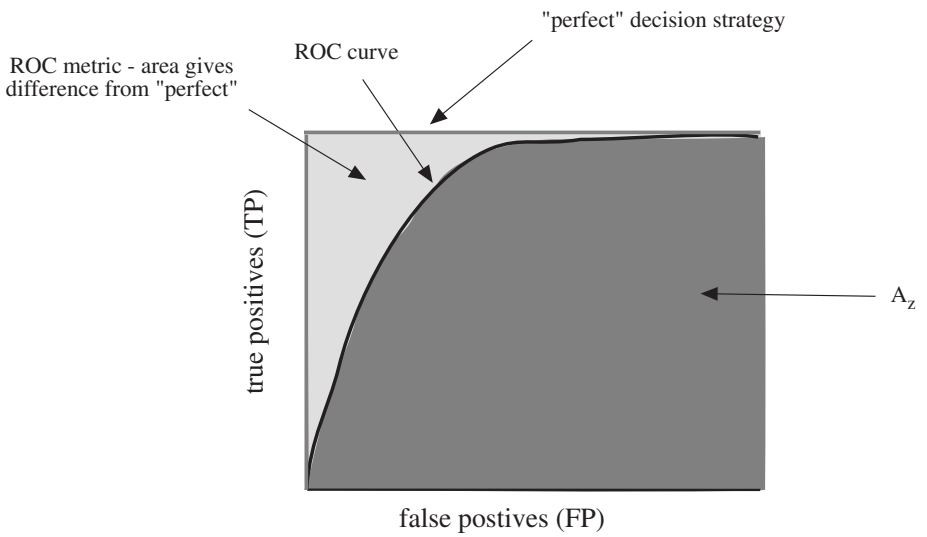


Figure 5.6: The graph of the expected number of true and false positives for a range of decision strategies is an ROC curve. The area under the curve, A_z, gives an overall measure of quality.

and the relevant structures are places of high gradient magnitude—edges. For this analysis I choose a detector that is often called the "Canny edge" (Canny 1983, Canny 1987) operator. This operator finds places that are singularities in the gradient of the image. These singularities are defined as local maxima of gradient magnitude in the gradient direction, i.e., where $f_{ww} = 0$ and $f_{www} < 0$. Additionally, I include the minimal gradient requirement $f_w > \Gamma$ in order to reduce the number of spurious edges in very flat regions. This edge-detection scheme entails two free parameters, the scale s_{edge} at which the derivatives are made and the gradient threshold Γ. In order to maintain a consistent basis for comparison, I choose the same scale and threshold for all of the experiments. The measurement scale s_{edge} is 1.0 (measured in pixel units), and the threshold Γ is 1.0 times the average gradient magnitude of the luminance image, i.e. $\Gamma = \langle |\nabla f(\vec{x})|^2 \rangle_{\vec{x} \in I}^{\frac{1}{2}}$. Ideally, this threshold should be chosen to produce the best results for the particular task, but here I have chose a single, reasonable standard for comparing a variety of different processes in a consistent manner.

5.3.4 Truth

Several of the metrics described above require a "goal" which is also the "correct" answer for a segmentation task. In the case of the function-based metrics, $\mathcal{M}_{\text{mean}}$ and $\mathcal{M}_{\text{rms}}$, the goal is an image or set of images. For the structure-based metrics; T, F, and $\mathcal{M}_{\text{rms}}$; the goal is a set of "correct" structures.

The use of a single correct answer raises several important issues. First, for any given input image there might not be a single correct answer. Indeed, the assumption that drives multiscale and hierarchical segmentations is that the size of desirable structures or segments could depend on the task. The absence of such a multiscale specification is a clear limitation of the metrics that I have defined above. I attempt to offset this limitation by choosing test images that have well-defined structures with few ambiguities. Such images may present alternative segmentations, but I make comparisons using only the "smallest-scale" structures in these images. Results in later sections will show that introducing more complex test images complicates the issue of using a single segmentation.

The second issue is *deciding* or *calculating* that single correct answer. Once again, I rely on simple test images to resolve this issue. For much of the analysis I use piecewise homogeneous images which are composed of patches that are generally bigger than the noise (discussed in Sect.

5.5). For these images the function-based goal is the image itself, while the structure-based goal is the structure detector applied directly to that image. In later sections I will introduce images that are not piecewise homogeneous. Such images do not present obvious correct answers, and using the small-scale structure of the test image as truth produces inconclusive experimental results. For this kind of test image I will define a correct answer *subjectively* and analyze the performance of various processes with respect to this subjectively-defined answer.

5.4 The test images

The Monte Carlo analysis strategy requires the use of a test image from which the image ensemble image is derived. The experiments in this chapter examine three different test images which are shown in Fig. 5.7. The *shape* image is chosen because it has "obvious" region boundaries that are indicated by sharp, isolated changes in luminance. It contains regions of various sizes and shapes as well as different levels of contrast. The *blocks* image has the same easy-to-segment property, but it contains only one relatively large object. This image is useful in studying the properties of correlated noise. The *head* image is used because it *does not* contain regions of constant luminance and it is relevant to the general problem of segmenting medical images.

5.5 Noise

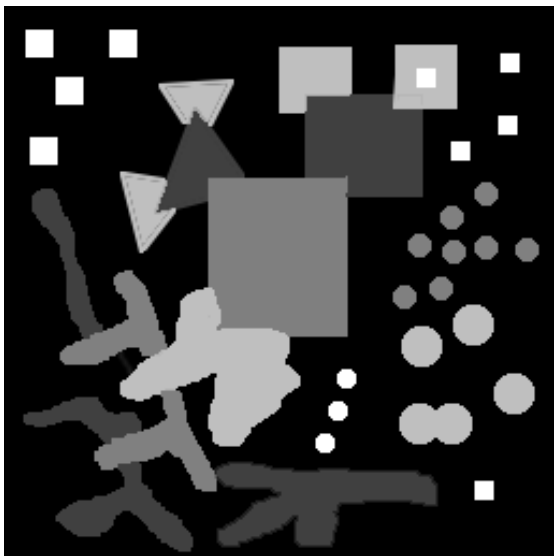
The capability of making stable, precise measurements in the presence of noise is of great practical importance in medical imaging as well as in natural vision. One widely-studied class of noise is additive Gaussian white noise, which is often simulated in the discrete domain by the adding pixelwise independent intensity values that are drawn from a Gaussian distribution (normal distribution) . Uncorrelated additive noise, however, describes only a fraction of the kinds of image variations which qualify as noise according to Definition 2. The analysis in this chapter extends the definition of noise to include various kinds of additive correlated noise as well as a kind of nonlinear noise process which acts on the boundaries of greyscale objects. Four different types of noise are used in this analysis. Figure 5.8 shows examples of these noise types applied to the MRI head image from Fig. 1.1.

1. Gaussian white noise (GWN) $[\sigma]$—additive, pixelwise-uncorrelated noise in which each intensity value is drawn from a normal distribution with standard deviation σ. For the experiments in this chapter, Gaussian noise is produced independently at each pixel using a pseudo-random number generator.

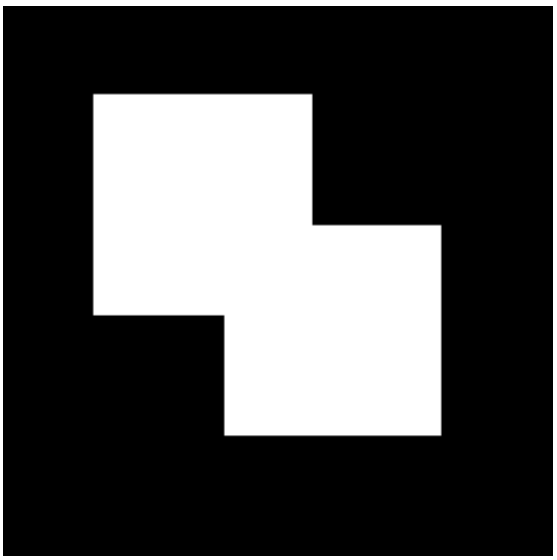
2. Correlated Gaussian noise (CGN) $[\sigma, s]$—additive noise produced by blurring GWN with a Gaussian kernel of size s, where the standard deviation of the resulting intensities is σ. This type of noise is used to produce patchwise and boundary noise.

3. Patchwise noise (PN) $[\sigma, p]$—additive noise that is formed in randomly generated patches that have a size of approximately p and intensities with a standard deviation σ. These patches are produced by applying the anisotropic diffusion process to CGN noise. The resulting noisy images are nearly piecewise constant. The parameter p is an indicator of the approximate average patch size, and $p = 2s$. The parameter s is the scale of the CGN that is used as the initial conditions of the anisotropic diffusion process that creates the patchwise behavior.

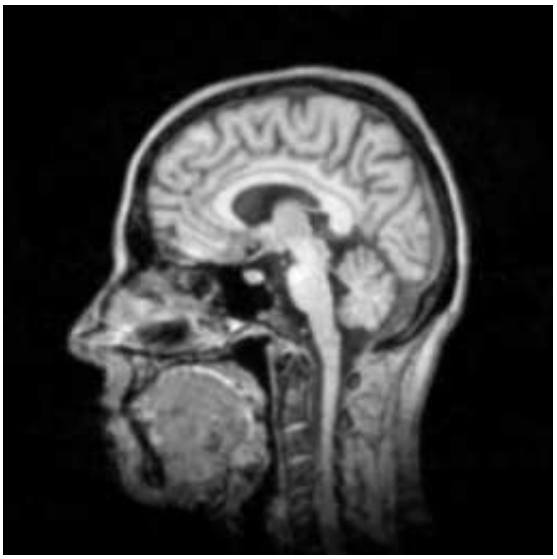
4. Boundary noise (BN) $[s, t]$—a greyscale morphological propagation operator, as described by van den Boomgaard (1992), is used with random coefficients in order to perturb boundaries



shape



blocks



head

Figure 5.7: The three test images are the (a) *shape* image, (b) *blocks* image, and (c) *head* image.

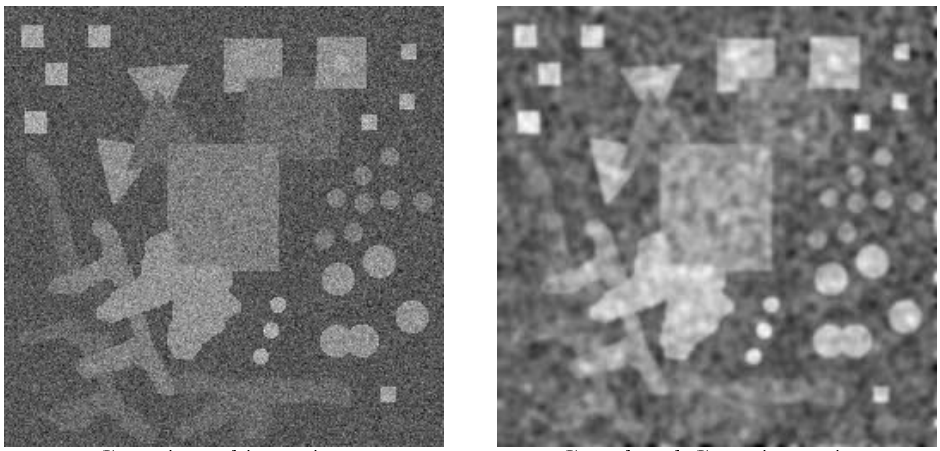


Figure 5.8: Example of the different kinds of noise that are used in the empirical analysis. From the top left: (a) GWN with a standard deviation of 50% of the average contrast in the image, (b) CGN with a standard deviation of 50% and scale of 2.0 (in pixel units), (c) PN with a deviation of 25% and a patch size of $p = 2.0$, and (d) boundary noise with a scale of 2.0 and an amplitude τ of 20.0.

of greyscale objects. The equation for a noisy image $f(\vec{x}, \tau)$ is

$$\frac{\partial f}{\partial t} = c(\vec{x})|\nabla f|, \tag{5.13}$$

where the test image is used as the initial conditions. The space-varying coefficient c is randomly generated CGN of size s, and the process runs for a time τ (the solution at time τ is used). The parameter s controls the size of the perturbations along the curve ("frequency") while τ controls the degree or extent of the perturbations ("amplitude")

The noise in the examples that follow is given as a percentage of the average intensity difference between the various patches in the image. For instance, if an image contains a single white object of intensity 1000 on a black background with intensity 0, a 50% noise level consists of noise with an intensity variance of 500.

As it stands, this specification of noise is ambiguous. The scale or degree of correlation is another important aspect of noise that is not captured by the variation in intensity. Even in the case of pixelwise uncorrelated noise, the *effective size* of the pixels has an impact on the scale-space behavior of the noise. In the very least one needs to know the number of pixels in the image. For all of the experiments in this chapter, the images are 256×256 pixels. In the case of patch correlated noise (PN), the size of the patches is the relevant measure of noise correlation. In later sections I give this size (p) in terms of the pixel spacing—pixels are separated by a distance of one unit.

The underlying issue regarding correlated noise is the relative *size* of the noise patches (or size of noise structures) compared with the size of the interesting image structures (the signal). One option for characterizing noise is to explicitly specify the relative size of interesting and uninteresting structures in addition to the relative intensity characteristics of these structures. Alternatively, one could quantify the signal-to-noise ratio of the image *through* scale-space as a mean of describing both intensity variation and spatial coherence of noise in an image. For these experiments I use the number of pixels in the image combined with the size, in pixel units, of the noise patches as indicators of the degree of noise correlation.

5.6 Overview of results

The following sections will examine Gaussian blurring, anisotropic diffusion, and multiscale diffusion under variety of process parameters, noise conditions, and test images. Table 5.1 shows the cases which are presented explicitly in this chapter. The columns are the three diffusion processes, and the rows are the types of noise. Each entry contains the names of test images that are examined and the number of the section in which the results appear.

5.7 Gaussian blurring

This section uses the Monte Carlo analysis strategy (depicted in Fig 5.1) to analyze the "performance" of Gaussian blurring (denoted $\mathcal{G}_s^{\text{iso}}$). Gaussian blurring is interesting for two reasons. First, the analysis of Gaussian blurring can help to understand the behavior of the Monte Carlo method and the metrics that I have developed in earlier sections. Second, geometry-limited diffusion offers an alternative scale space. As such, we hope that it compares favorably with isotropic or Gaussian scale over a range of scales, process parameters, and noise parameters. Gaussian blurring is used throughout this chapter as a benchmark for evaluating the effects of the geometry-limited diffusion process. These results, and all of the results in this chapter, are computed from image ensembles S that contain 10 images.

	Gaussian blurring		Anisotropic diffusion		Multiscale diffusion	
Uncorrelated Gaussian	(5.7)	*shape*	(5.8)	*shape*	(5.9)	*shape*
	(5.10)	*head*	(5.10)	*head*	(5.10)	*head*
	(5.10.1)	*shape*†	(5.10.1)	*shape*†	(5.10.1)	*shape*†
	(5.10.1)	*head*†	(5.10.1)	*head*†	(5.10.1)	*head*†
Patchwise Correlated	(5.11)	*shape*			(5.11)	*shape*
	(5.11)	*blocks*			(5.11)	*blocks*
Boundary	(5.12)	*shape*			(5.12)	*shape*

Table 5.1: The section numbers of the various combinations of diffusion processes, noise processes, and test images where the † indicates ROC analysis with *subjectively-defined* boundaries

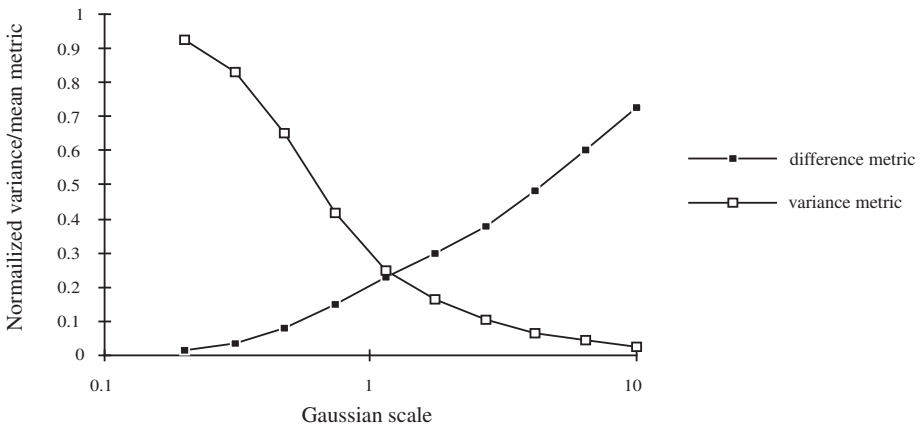


Figure 5.9: The metrics $\mathcal{M}_{\text{mean}}$ and $\mathcal{M}_{\text{var}}$ as a function of Gaussian scale show the tradeoff between accuracy and reliability for 50% GWN.

5.7.1 Function-based metrics

Figures 5.9 and 5.10 show the results for this analysis for uncorrelated noise (GWN) using the *shape* image from Fig. 5.7a. These graphs illustrate several important aspects of this analysis. First, 5.10 shows the explicit tradeoff between accuracy ($\mathcal{M}_{\text{mean}}$) and reliability ($\mathcal{M}_{\text{var}}$). *In the presence of uncorrelated noise, progressively larger isotropic kernels give improved reliability at the expense of decreased accuracy.* Second, with regard to the rms metric ($\mathcal{M}_{\text{rms}}$) there is, for each level of noise, an "optimal" Gaussian scale which is indicated by the lowest point on the $\mathcal{M}_{\text{rms}}$ curve. This optimal scale increases with increasing noise, and with greater noise the best value (over all scales) for these metrics becomes progressively worse (larger). Hereafter the *value* of the rms metric at this optimal scale (scale at which $\mathcal{M}_{\text{rms}}$ is lowest) is denoted $\mathcal{M}^{O}_{\text{rms}}$.

The rms error metric, $\mathcal{M}_{\text{rms}}$, when used in this fashion, measures the "quality" of a restoration. Because the goal image is the original *shape* image before the addition of noise, $\mathcal{M}_{\text{rms}}$ is effectively an evaluation of the ability of Gaussian blurring to reproduce an image that has been contaminated by uncorrelated noise. The issue of restoration quality, however, is very complex, and rms error is not necessarily a good measure of this quality. The rms error does not necessarily indicate the best *looking* image; it is meant to indicate the potential for pixelwise classification using some statistical technique. However, our understanding of this metric could be improved by comparing the rms error with our own subjective evaluation. Thus, I briefly revisit the subjective mode of computer vision research and ask, "Does the optimal Gaussian-blurred restoration of an image look most like the

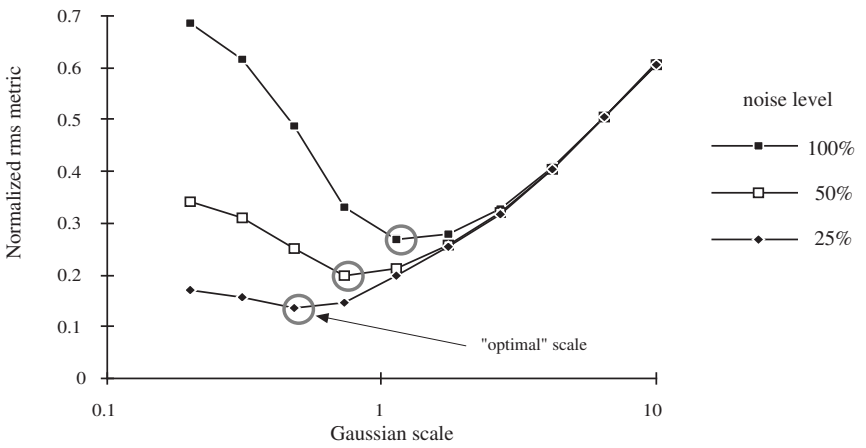


Figure 5.10: The metric $\mathcal{M}_{\text{rms}}$ as a function of Gaussian scale shows that there is an optimal scale for minimizing the rms error which includes both the bias $\mathcal{M}_{\text{mean}}$ and the variance $\mathcal{M}_{\text{var}}$.

original image?" Figure 5.11 shows the *shape* image with 100% GWN and five different levels of Gaussian blurring. A systematic answer to the question of which image "looks" the best is a matter for psychophysical study and is beyond the scope of this dissertation.

5.7.2 Structure-based metrics

Figure 5.12 shows the measures T and F as a function of scale for 50% GWN. All of these results are normalized by the number of true positives associated with the identity process so that the results from different images and noise levels can be compared.

Consider the relationship between true and false positives as a function of scale. From a very limited point of view, Gaussian scale can be seen as a means of reducing the number of false-positive edges at the expense of reducing the number of true-positive edges. The two quantities do not fall off equally—at small scales there are great reductions in the number of false edges with little impact on the number of true edges. This is consistent with the results of the function-based metrics; at small scales increasing Gaussian scale reduces unwanted information more quickly than it undermines the signal. However, the false positives are *not* monotonically decreasing with increasing scale (in Fig. 5.12 the graph "turns up" at about $s = 2.5$). At large scales the number of false positives increases with increasing scale. This is because large Gaussian kernels can distort the shapes of large-scale objects so that the boundaries no longer lie within the tolerance (σ_t) that is built in to the fuzzy structure-set goal f'_g.

The relationship between TP and FP is more apparent in the ROC graphs shown in Fig. 5.13. These plots show true versus false positives (over a range of scales) for three different levels of noise. The "elbows" on the left hand side of the graph indicate the trend of increasing FPs from the distortion of object boundaries. When computing integrals of these curves, the *maximum* value of T is used for each value of F.

The behavior of these elbows depends on the choice of spatial tolerance, σ_t, used in the formation of the fuzzy structure-set goal. Figure 5.14 shows the behavior of T and F over a range of scales for three different tolerance values. Larger tolerance measures allow for more distortion in object boundaries from Gaussian scale and therefore provide decreasing FPs for a larger range of scales. Larger tolerances should provide a better tradeoff between TPs and FPs, especially with the application of large Gaussian kernels.

The elbows are indicative of the limitations of this ROC analysis. The notions of false and true edge locations (however fuzzy) capture only a fraction of the issues that are relevant to image seg-

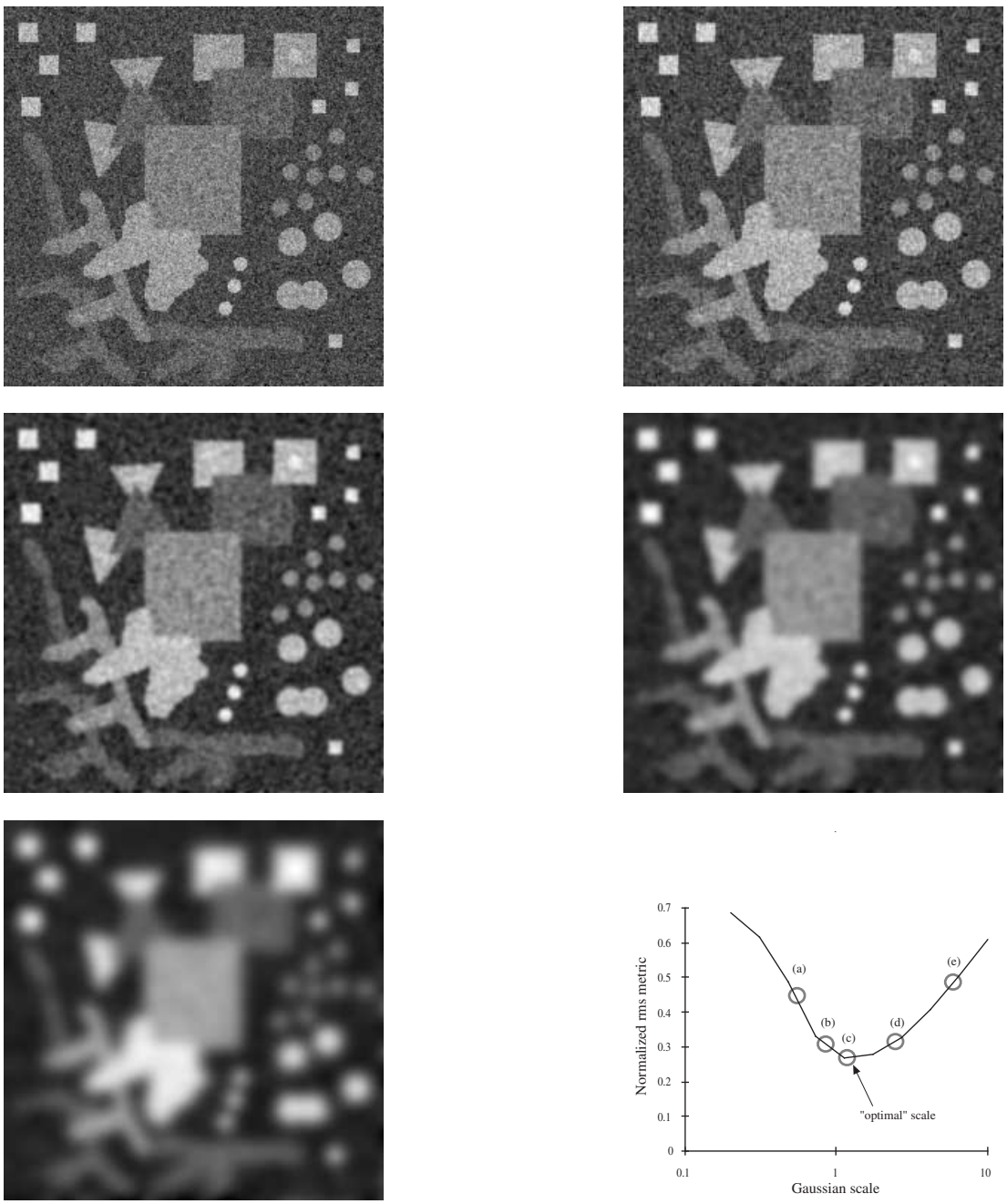


Figure 5.11: The *shape* image with 100% GWN and five different levels of Gaussian blurring. The standard deviations of the Gaussian kernels are (a) 0.45, (b) 0.70, (c) 1.25, (d) 2.5, and (e) 5.0. The value 1.25 has the lowest rms error. The rms errors are indicated on the line graph in (f).

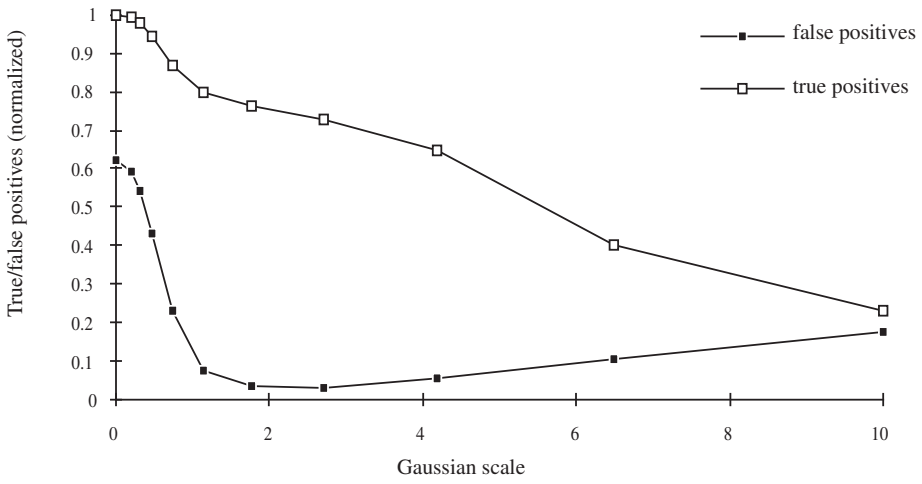


Figure 5.12: True-positive T and false-positive F edges as a function of scale for 50% GWN.

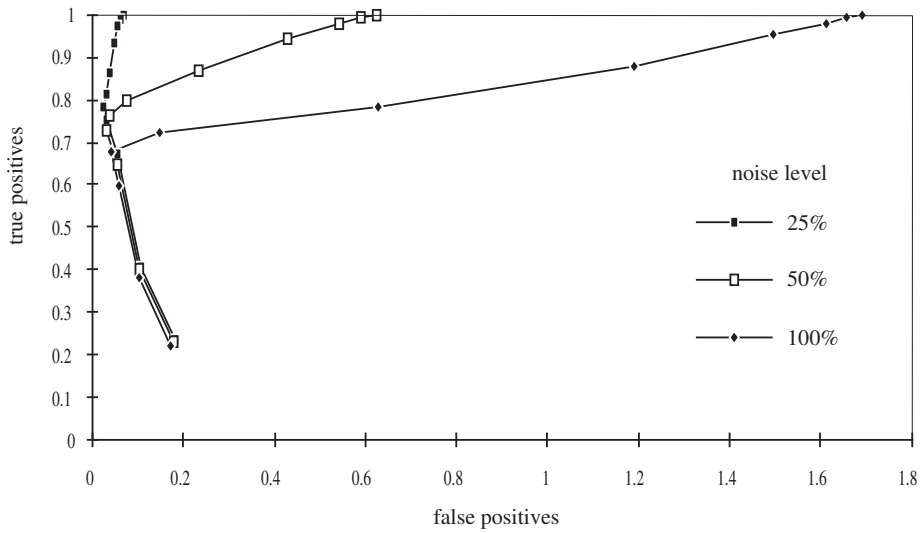


Figure 5.13: The ROC graphs plot T versus F for three different levels of noise. The integrals below the curves are A_z and the integrals above the curves are $\mathcal{M}_{\mathrm{ROC}}$.

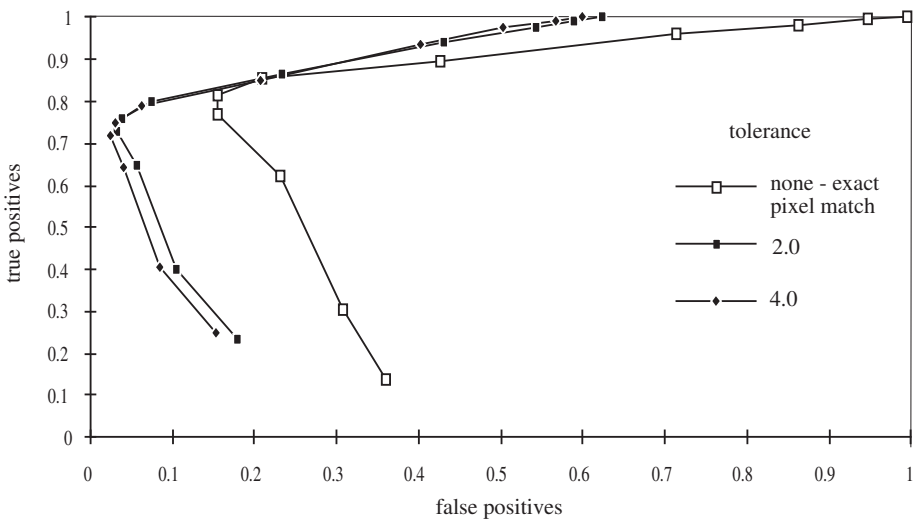


Figure 5.14: The ROC graphs plot T versus F for three different tolerance levels in the presence of 50% GWN.

mentation. The ROC curves do not, for example, capture the global shape characteristics associated with sets of structures. Imagine that the structure-set goal is a set of edges which represent a circle in the middle of an image (Fig. 5.15). Consider two structure-set estimates. The first is an identical circle which is offset to the left by two the positions of pixels and the second is a set of disjoint edges *all of which* lie less than three pixel units from the positions of boundaries in the goal. We would expect these two estimates to have similar values of T and P, yet the first example is (arguably) a better estimate of the original circle. Paradoxes of this kind plague many different approaches to comparing structure sets. A "perfect" metric for comparing structure sets should account for global shape properties. Producing such metrics, however, is a fundamental challenge in computer vision and pattern recognition. Developing more comprehensive metrics is an area of future research and is beyond the scope of this dissertation.

5.8 Anisotropic diffusion

This section applies the same evaluation methods to anisotropic diffusion. From (2.9) the anisotropic diffusion equation is

$$\nabla \cdot g(|\nabla f|)\nabla f = \frac{\partial f}{\partial t}. \tag{5.14}$$

Recall that this process, denoted $\mathcal{G}_{t,k}^{\text{aniso}}$, introduces *two* free parameters: the diffusion time t and the conductance parameter k, which is expressed as a multiple of the root-mean-squared gradient magnitude of f. The experimental framework of Sects. 5.2-5.7 provides an opportunity to study the performance of $\mathcal{G}_{t,k}^{\text{aniso}}$ and its sensitivity to the choice of parameters. Additionally, it provides a means of evaluating its effectiveness relative to isotropic Gaussian blurring.

5.8.1 Comparisons with isotropic scale: function-based metrics

Figure 5.16 shows the behavior of function-based metrics for the *shape* image with the introduction of 25% GWN. For this graph I chose a value of $k = 0.8$ in the anisotropic diffusion, which gives particularly good results (i.e., lowest rms error from among the range of k that is studied). The

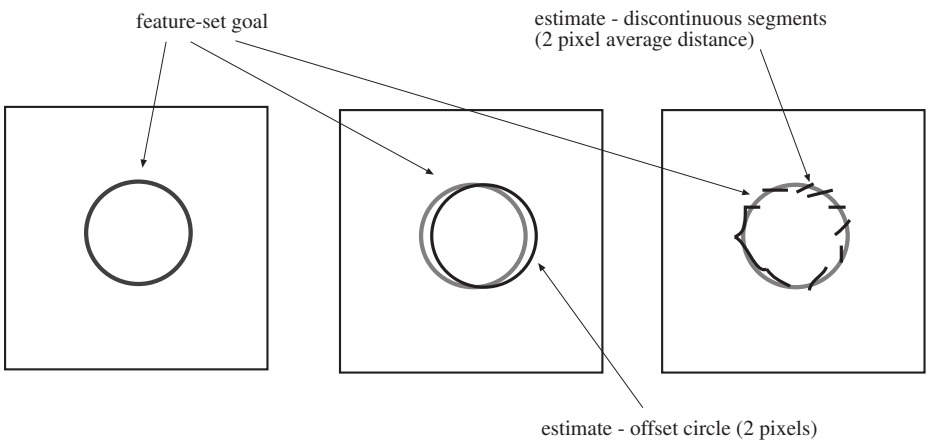


Figure 5.15: The concept of comparing true and false positives with a goal or template does not capture global shape properties (such as continuity). Left, a goal with a tolerance takes the shape of a circle; center, an offset circle shows a low number of true positives; right, a set of discontinuous edges gives a comparable response.

horizontal axis represents both s, the scale parameter for Gaussian blurring, and t, the time parameter for anisotropic diffusion. The parameters s and t are generally not commensurate, but these parameters are combined using the relationship $s = \alpha(2t)^{\frac{1}{2}}$ (as described in Sect. 3.4) with a choice of $\alpha = 1.0$. Recall that this relationship follows from the role that each of these parameters play in the diffusion equation. I choose $\alpha = 1.0$ solely for the purposes of visualizing and comparing the results of these two types of processes.

Increasing t and s have qualitatively similar effects: greater reliability (decreasing $\mathcal{M}_{\text{var}}$) at the expense of decreased accuracy (increasing $\mathcal{M}_{\text{mean}}$). However, the anisotropic diffusion process gives consistently lower variance for a given value of accuracy (alternatively, better accuracy for a given value of variance). This advantage is born out by the consistently lower values of $\mathcal{M}_{\text{rms}}$ that result from the anisotropic diffusion process. The values of $\mathcal{M}^{O}_{\text{rms}}$ (optimal scale, time, and conductance within the ranges that are studied) from each process show that anisotropic diffusion offers a rms error of 0.05, while the best result from Gaussian blurring is 0.15. *For the shape image, in the presence of 25% uncorrelated noise, the anisotropic diffusion process offers approximately three times better performance than Gaussian blurring* (1/3 of the rms error).

5.8.2 The conductance parameter

The performance of the anisotropic diffusion process depends on one's choice of parameters. Figure 5.17 shows surface plots of $\mathcal{M}_{\text{rms}}$ as a function of both t and k. These plots reveal several important properties of $\mathcal{G}^{\text{aniso}}_{t,k}$. First, the best choices of k, those choices that produce the lowest values of $\mathcal{M}_{rms}$, depend on the level of noise in the image. These graphs are quite flat near optimal values of k (approximately 0.8 for 25%, 1.2 for 50%, and 1.6 for 100%) which indicates that there is some latitude in choosing k. This observation is of great practical importance, because it suggests that the exact choice of k is not critical to the performance of the anisotropic diffusion process; values in the range $k = 0.6$–1.2 provide comparable results. However, for very low values of k, values well below the optimum, the process performs poorly, and it does considerable worse than isotropic (Gaussian) blurring. This is evident for noise levels of 50% and 100% and choices of $k = 0.6$

These same graphs indicate that under these circumstances (additive, uncorrelated noise on the shape image) relatively small values of t ($t \approx 1.0$) provide the best results. Figure 5.17 shows the same measurements for three different values of GWN. The choices of t that provide the lowest $\mathcal{M}_{\text{rms}}$

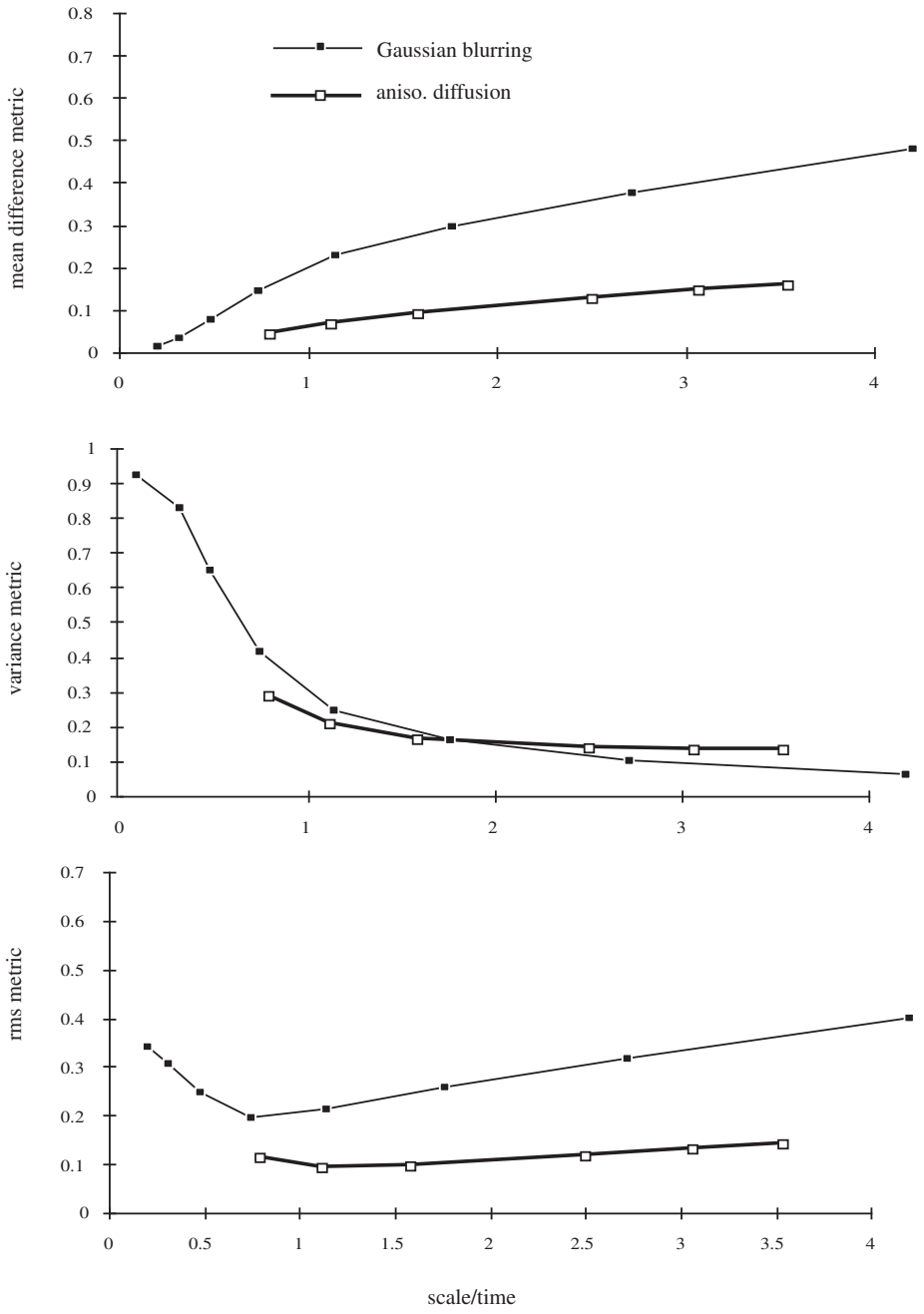


Figure 5.16: WIth 50% uncorrelated noise, graphs of the function-based metrics $\mathcal{M}_{\text{var}}$, $\mathcal{M}_{\text{mean}}$, and $\mathcal{M}_{\text{rms}}$ versus s or t show that $\mathcal{G}_{t,k}^{\text{aniso}}$ and $\mathcal{G}_{s}^{\text{iso}}$ have similar effects on accuracy and reliability through scale or time. However, $\mathcal{G}_{t,k}^{\text{aniso}}$ shows consistently better accuracy for a given level of reliability.

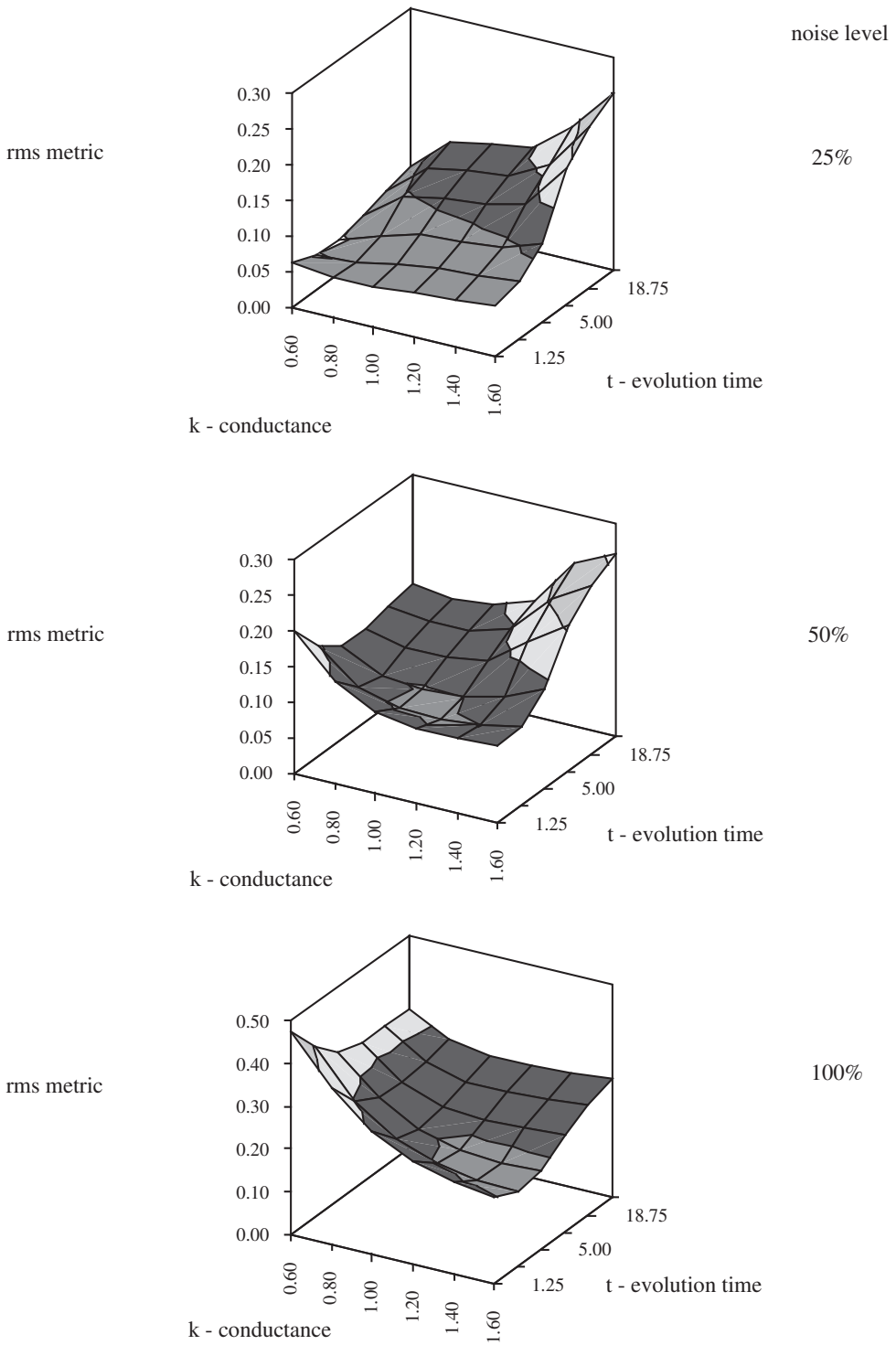


Figure 5.17: The performance of the anisotropic diffusion process $\mathcal{G}^{\text{aniso}}_{t,k}$. Surface plots give the values of metric $\mathcal{M}_{\text{rms}}$ as a function of t and k for three levels of noise; 25%, 50%, and 100% GWN.

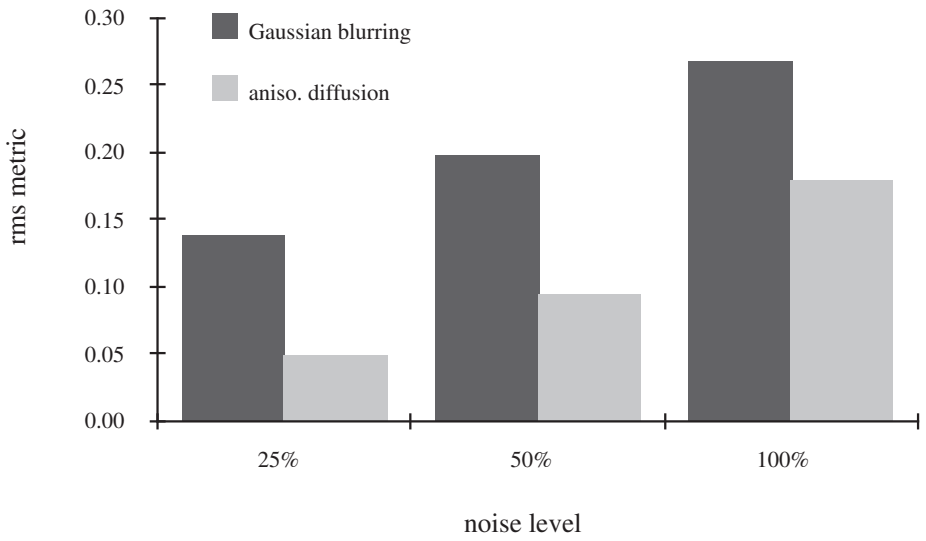


Figure 5.18: The optimal choices of parameters show that $\mathcal{M}_{\text{rms}}\ \mathcal{G}_{t,k}^{\text{aniso}}$ achieves a lower rms error than $\mathcal{G}_s^{\text{iso}}$, but the relative magnitude of this advantage advantage diminishes at greater levels of noise.

increase with progressively more noise. Greater levels of noise require longer processing times. This is analogous to the behavior of Gaussian scale processing, which requires larger kernels (larger s) for greater levels of noise. For the best choices of k, these plots are relatively flat in the t direction. This behavior is also shown in Fig. 5.16, which compares $\mathcal{G}_s^{\text{iso}}$ and $\mathcal{G}_{k,t}^{\text{aniso}}$. In anisotropic diffusion the rms error does not "take off" at larger scales as it does with Gaussian blurring. This stability over time is a result of the nonconducting properties of boundaries. Once the small luminance fluctuations have blurred away, the remaining boundaries have low conductance, and they decay at a very slow rate. Thus, in practical applications there is also a great deal of latitude in choosing the total evolution time t.

5.8.3 Signal to noise

With greater levels of noise both processes $\mathcal{G}_s^{\text{iso}}$ and $\mathcal{G}_{t,k}^{\text{aniso}}$ are less effective at producing accurate, reliable results. *Although the anisotropic diffusion process produces consistently better results, this advantage diminishes with increasing noise levels.* Figure 5.18 shows $\mathcal{M}_{\text{rms}}^{O}$ for both processes at four different levels of noise. For low values of noise the optimal rms error of $\mathcal{G}_{t,k}^{\text{aniso}}$ is one third of the error associated with Gaussian scale ($\mathcal{G}_s^{\text{iso}}$ is 70% better), but for the largest values of noise the difference is less than 50%. This result is consistent with the notion that large values of noise undermine the ability of the anisotropic diffusion process to "control" the flow of intensity. The process $\mathcal{G}_{t,k}^{\text{aniso}}$ makes use of derivative measurements in order to constrain the diffusion process. As derivatives become less reliable (more noise), the process cannot make effective use of derivative measurements. Under such circumstances, the best performance $\mathcal{G}_{t,k}^{\text{aniso}}$ results from a choice of parameters that constrain the flow of intensity much less than in conditions of little noise.

The relationship between isotropic and anisotropic diffusion becomes clearer if we view Gaussian blurring is a special case of anisotropic diffusion. Because

$$\lim_{k\to\infty} \exp\left(\frac{|\nabla f|^2}{2k^2}\right) = 1,$$

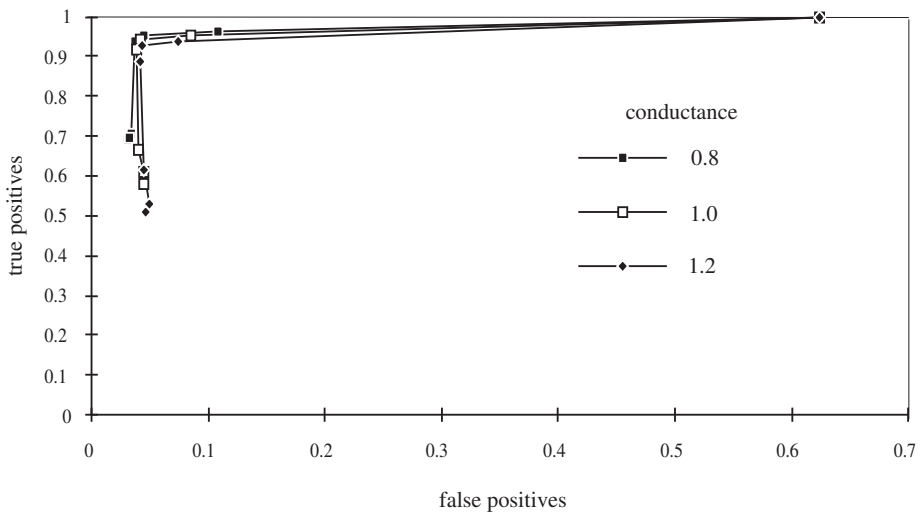


Figure 5.19: The ROC curves with 50% GWN for different values of k show that the ROC analysis is relatively insensitive to conductance.

Gaussian blurring is the endpoint of the anisotropic diffusion processes defined by the limit

$$\lim_{k \to \infty} \left[\frac{\partial f}{\partial t} = \nabla \cdot g(|\nabla f|) \nabla f \right].$$

As k increases, $\mathcal{G}_{t,k}^{\text{aniso}}$ behaves more like Gaussian scale (isotropic diffusion). Recall from Sect. 3.3.2 that the conductance parameter k acts as a kind of threshold on gradients in the image. For greater levels of noise, greater k values are required in order to obtain the optimal distinction between signal and noise. This increased threshold is reflected explicitly in the conductance parameter k and implicitly in the larger gradient magnitudes of the noisier images. Recall that I have expressed k as a multiple of the root-mean-squared gradient (over the entire image), so for noisier images the effective conductance parameter is larger even if k were to remain constant. *Thus, greater levels of noise require greater values of the conductance parameter, and the anisotropic diffusion process behaves more like Gaussian blurring.*

5.8.4 ROC Analysis: structure-based metrics

The analysis of the structure-based metrics tells a similar story. Figure 5.19 shows the ROC curves for $\mathcal{G}_{t,k}^{\text{aniso}}$ with three different levels of k. ROC curves with different conductance parameters and noise levels confirm the observation that the behavior of the process is not highly sensitive to the choice of the conductance parameter. Figure 5.20 shows the difference between the ROC curves for Gaussian scale and anisotropic diffusion (for a single choice of conductance parameter, $k = 0.8$). *Anisotropic diffusion offers more true positives for a given level of false-positive edges.* This is particularly evident at very low values of FP, where the anisotropic diffusion process is able to "blur away" false edges with little or no impact on true edges. Compared with Gaussian blurring, the curves for anisotropic diffusion have flatter "tops" (upper middle and left on the graphs) and therefore drop down more dramatically in the low range of FPs. These observations suggest that when the process runs for short time periods, it reduces the small-scale noise, but for larger times the underlying large-scale structure begins to deteriorate. With increasing scale, Gaussian blurring shows a consistent undermining of large-scale structures at relatively constant rates, until it reaches the point where the curve turns back on itself and the number of FPs begin to increase with larger

scale. This fact leads to lower values of $\mathcal{M}_{\text{ROC}}$ for anisotropic diffusion (as shown in Fig. 5.21). Anisotropic diffusion offers a consistently better compromise between too many spurious edges and the right number (and placement) of true edges.

5.9 Multiscale diffusion

The function and structure-based metrics from Sect. 5.3 show that the anisotropic diffusion process offers an improvement over Gaussian blurring. The examples in Chapt. 3 indicate, qualitatively, that the multiscale diffusion process is a further improvement. Does the quantitative analysis bear this out?

Recall that the multiscale diffusion process introduces a scale parameter $s(t)$ into the conductance term.

$$\nabla \cdot g(|\nabla_{s(t)} f|)\nabla f = \frac{\partial f}{\partial t}, \tag{5.15}$$

Furthermore, I have proposed in Sect. 3.4 that among the reasonable options for $s(t)$ is that of defining a "tradeoff" between isotropic and anisotropic scale. The simple notion that a *unit* of isotropic scale should be worth a fixed number of *units* of anisotropic scale results in the following relationship:

$$s(t) = \sqrt{s_0^2 - 2\alpha t} \tag{5.16}$$

The constant α indicates how many units of anisotropic scale it takes to make one unit of isotropic scale. The multiscale diffusion process, $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ has three free parameters: the starting scale s_0, the trade-off constant α, and the conductance parameter k. The effects of these parameters are studied together and separately in order to evaluate the effectiveness of this process.

5.9.1 Effects of the scale trade-off

Experiments with the multi-valued diffusion process $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ show that in the presence of uncorrelated white noise, the behavior of $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ is similar to $\mathcal{G}^{\text{aniso}}_{t,k}$ with respect to the conductance parameter k. A range of k values, centered around $k = 1.0$ provide comparable results. Thus, for the experiments in this section, I fix the value of k at 1.0 in order to study the effects of the other parameters.

Figure 5.22 shows the values of the metric $\mathcal{M}_{\text{rms}}$ over a range of values for the starting scale s_0 and tradeoff constant α. These surfaces show that in the presence of uncorrelated noise $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$, gives the best results for relatively small starting scales, $s_0 = 0.75$–1.5, and small values of α, $\alpha = 0.125$. The *difference* in performance among low values of α ($\alpha = 0.125$ and 0.25) is not great—the graphs become progressively more flat for lower values of α. Recall that α affects how many time steps are required for the process to complete, and the number of time steps is inversely proportional to α. The practical consequences (in terms of computation time) of very small values of α are overwhelming. Multiscale diffusion processes with values of $\alpha < 0.25$ require prohibitively large times to complete, so I choose a compromise between optimal performance and processing time and use values of $\alpha = 0.25$ throughout the remainder of this analysis.

Referring to Fig. 5.22, notice that for higher levels of noise (lowermost graph), greater values of s_0 are required in order to achieve optimal results. This is a consistent trend throughout the analysis in this chapter; greater levels of noise require larger scales and/or longer processing times.

5.9.2 Comparisons with anisotropic and isotropic scale

The results for multiscale diffusion process are mixed. For uncorrelated noise in the *shape* image, the function-based metrics show that $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ compares favorably with Gaussian blurring but is not significantly different from anisotropic diffusion. Analysis of structure-based metrics and other test

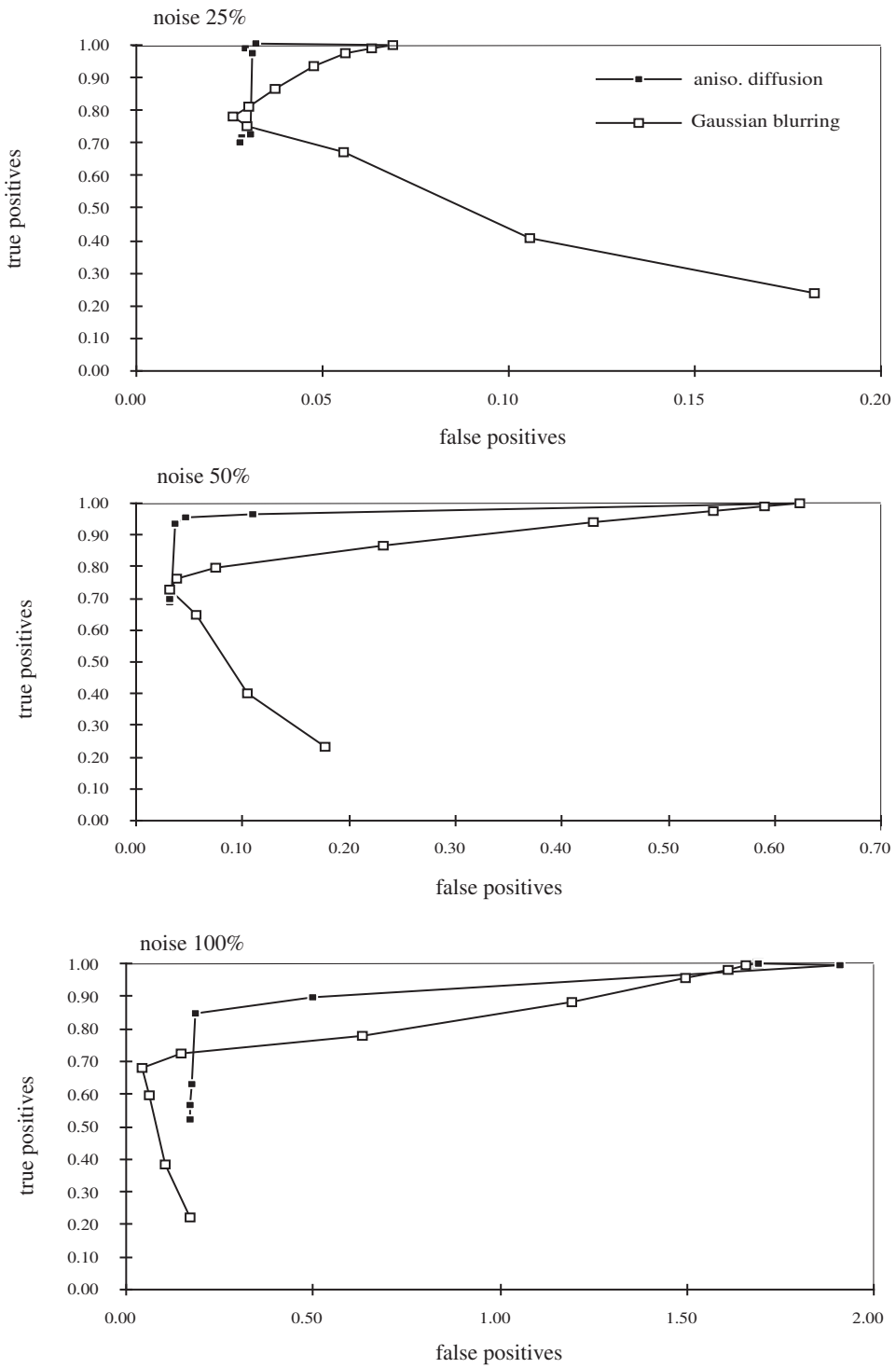


Figure 5.20: The ROC curves for three different levels of GWN, show that anisotropic diffusion compares favorably with Gaussian blurring.

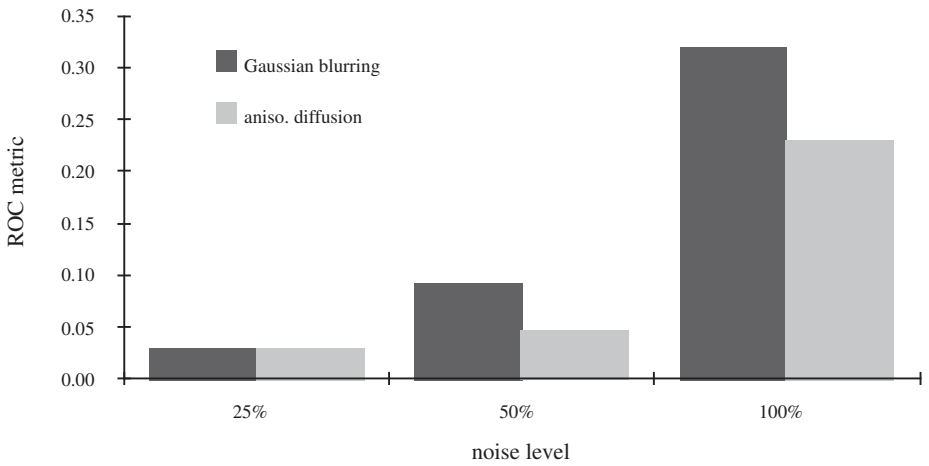


Figure 5.21: For the metric $\mathcal{M}_{\mathrm{ROC}}$ and the *shape* image, comparisons between Gaussian blurring and anisotropic diffusion shows that $\mathcal{G}_{t,k}^{\mathrm{aniso}}$ shows a consistently better compromise between too many spurious edges and the right number (and placement) of true edges.

images shows that $\mathcal{G}_{s_0,k,\alpha}^{\mathrm{multi}}$ does offer a significant advantage over $\mathcal{G}_{t,k}^{\mathrm{aniso}}$. Figure 5.23 shows the smallest (chosen from within the parameter ranges that are evaluated) values of $\mathcal{M}_{\mathrm{rms}}$ for three different levels of noise. *These results indicate that for the* shape *image (piecewise flat) in the presence of uncorrelated noise the multiscale diffusion process offers little or no advantage over* $\mathcal{G}_{k,t}^{\mathrm{aniso}}$ *in rms error.*

The ROC analysis indicates an advantage of multiscale over anisotropic diffusion. Figure 5.24 shows the ROC metric $\mathcal{M}_{\mathrm{ROC}}$ for all Gaussian blurring, anisotropic diffusion, and multiscale diffusion. Multiscale diffusion is particularly effective (relative to anisotropic diffusion) at greater levels of noise. This is consistent with the subjective observation of Sect. 3.1 that the anisotropic diffusion process "breaks down" at high levels of noise. The introduction of scale into the conductance parameter allows the multiscale process to achieve lower levels of false positives for a given level of true positive edge measurements.

5.10 Dependence on the test image

Results in previous sections apply to the image that I have called the *shape* image (Fig. 5.7a). How relevant are these results for other test images? Figure 5.25 s shows the rms error results for both the *shape* and *head* images. For the function-based metric $\mathcal{M}_{\mathrm{rms}}$, neither the anisotropic nor the multiscale diffusion process is measurably better than Gaussian blurring. *According to the function-based metrics, anisotropic diffusion applied to the* head *image is no more effective than Gaussian blurring.* The ROC analysis (not presented here) confirms this observation.

This result discloses a fundamental property of the edge-affected diffusion processes. Given that the $\mathcal{M}_{\mathrm{rms}}$ metric measures the effectiveness of a process $\mathcal{G}$ in *restoring a luminance image* from a noisy image, *edge-affected diffusion as a mean of restoring intensity images is most effective where the goal or truth is piecewise constant.* This is evident in the examples of $\mathcal{G}_{s_0,k,\alpha}^{\mathrm{multi}}$ and $\mathcal{G}_{k,t}^{\mathrm{aniso}}$ throughout much of this dissertation, which have a flat, "cartoon-like" appearance. This behavior is confirmed by the observation of Nordstrom (1990) and Snyder et al. (1992). They show that anisotropic diffusion is mathematically equivalent to a gradient-descent minimization of energy functionals that seek piecewise homogeneous approximations to images. The advantage of anisotropic diffusion (over Gaussian blurring) is lost when the goal is significantly different from piecewise constant.

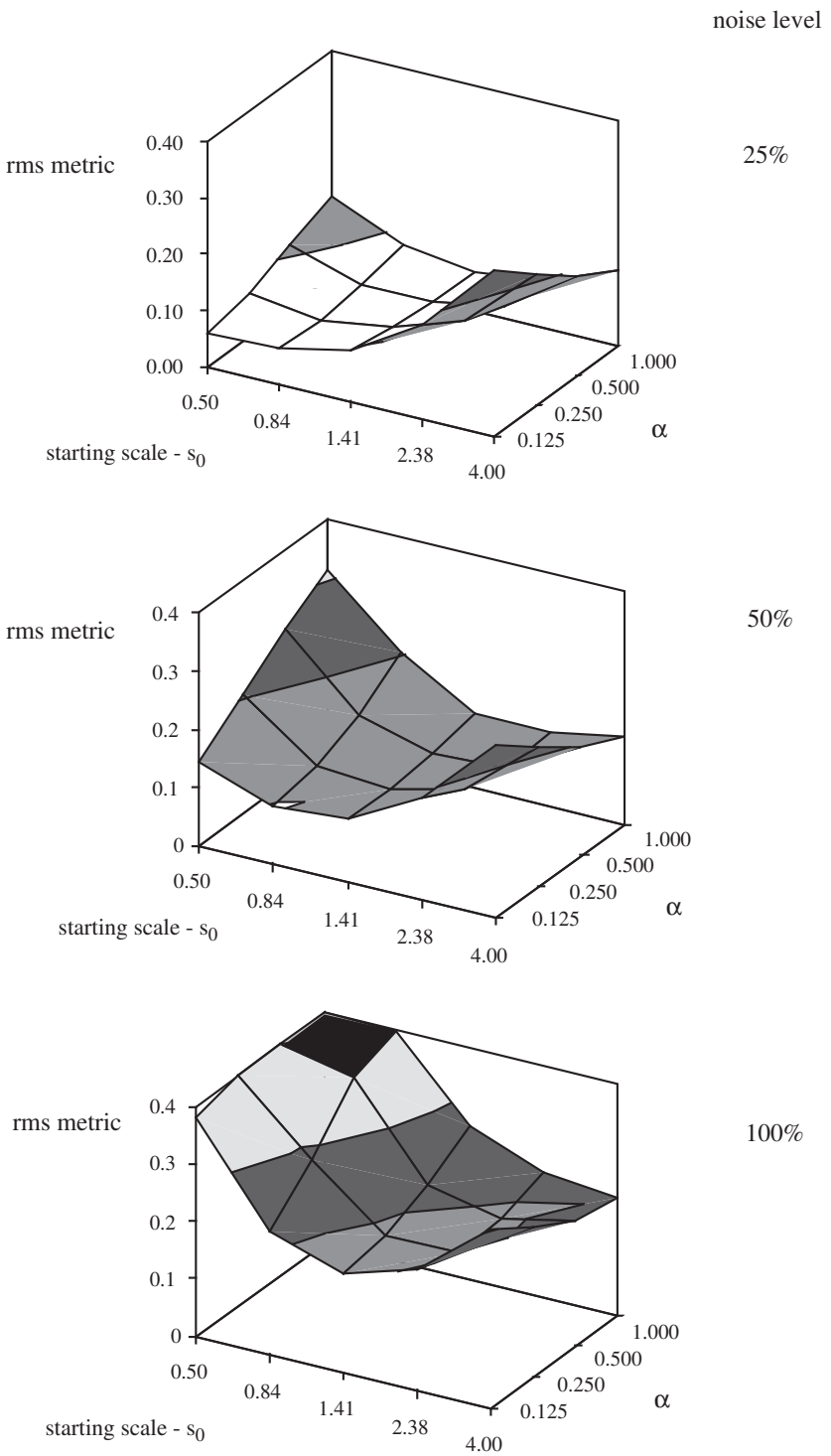


Figure 5.22: The performance of the multiscale diffusion process $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$. Surface plots give the values of metrics $\mathcal{M}_{\text{rms}}$ as a function of s_0 and α for three levels of noise; 25% (top), 50% (middle), and 100% (bottom).

Figure 5.23: Comparisons of $\mathcal{M}^{O}_{\text{rms}}$ over a range of noise levels for $\mathcal{G}^{iso}_{s}$, $\mathcal{G}^{aniso}_{k,t}$, and $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$.

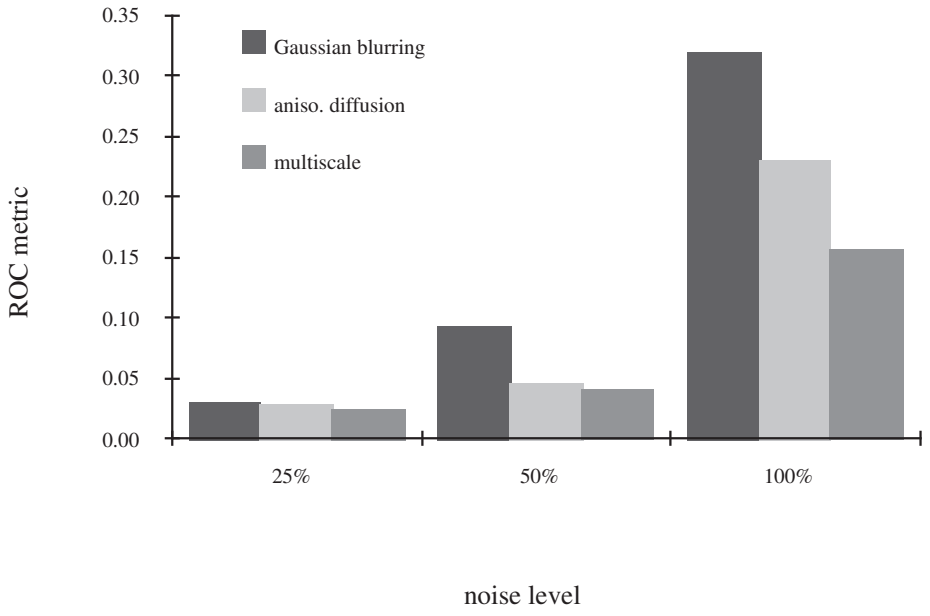


Figure 5.24: Comparisons of $\mathcal{M}^{O}_{\text{ROC}}$ over a range of noise levels for $\mathcal{G}^{iso}_{s}$, $\mathcal{G}^{aniso}_{k,t}$, and $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$.

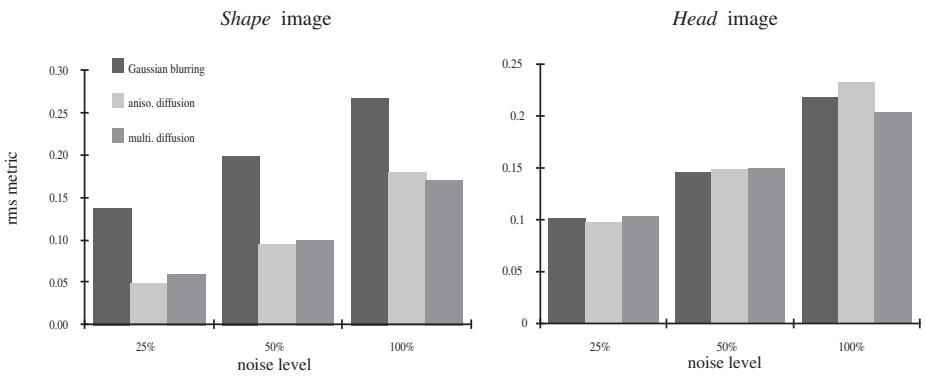


Figure 5.25: The function-based metric $\mathcal{M}_{\text{rms}}$ computed on the *head* image indicates that the performance advantages that $\mathcal{G}^{\text{aniso}}_{k,t}$ and $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ showed in the *shape* image are lost in the *head* image.

5.10.1 Subjective structure goal

The underlying difficulty with applying the Monte Carlo analysis to the *head* image is the assumption of "truth". The previous experiments have assumed that the "truth" lies in the small-scale (pixel-scale) structure of the test image. The *head* image contains many complex structures (including some uncorrelated noise from the image process) of various sizes, and many of these structures could be considered noise at one level or another. The assumption that a useful process should reproduce the small-scale structure of this image is dubious. This assumption contradicts the obvious fact that this same image serves as the motivation for developing robust segmentation algorithms.

I have proposed a segmentation strategy which *first* seeks a piecewise constant approximation to an image and *then* performs a discrete segmentation. The question remains: Does this strategy produce "useful" segmentations for images like the *head* image?

Instead of using the test image as truth, I develop the notion of a subjective goal. In the event that the "true" segmentation is not immediately discernible from the test image, one can subjectively define a set of image structures to serve as the goal. For this analysis I use a set of hand-drawn boundaries that indicate coherent structures. Figure 5.26 shows the subjective contours that I use with the *shape* and *head* images. These structures are chosen because they are visually interesting, and because they are of similar size (subjectively).

The boundaries of Fig. 5.26 are not complete because I have left out several obvious structures. In the ROC analysis the missing boundaries should add a constant factor to the number of false positives (those detected edges that are associated with the "missing" edges in the subjective goal). This constant factor should not undermine the comparisons of one process to another. In these examples, I assume that subjective boundaries represent approximately 50% of the objects that are present in the image. If this is true, the analysis would consistently overestimate the number of false positives by an amount equal to the number of true positives. Throughout the analysis of this section, I subtract these "missing" boundaries from the number of false positives in order to obtain results of the metric $\mathcal{M}_{\text{ROC}}$ that are more consistent with the examples from earlier sections. This 50% "fudge factor" does not change the relative behavior of the different processes, only the overall magnitude of $\mathcal{M}_{\text{ROC}}$.

The *shape* image

The structure-based analysis, conducted with the subjective edges in Fig. 5.26, confirms the earlier findings on the *shape* image. Figure 5.27 shows the ROC curves associated with the Gaussian blurring and the variable-conductance diffusion processes over a range of scales or times. The greater

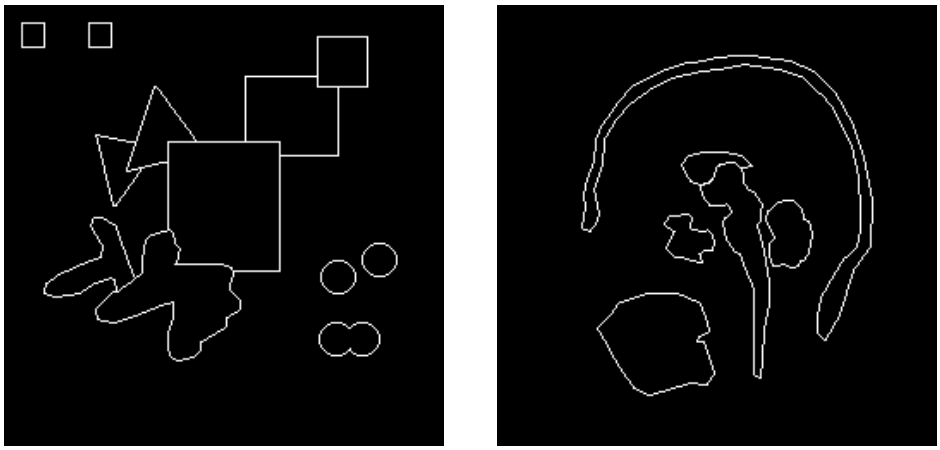


Figure 5.26: Subjective goals are defined for the *shape* and *head* images by hand drawing boundaries of selected objects.

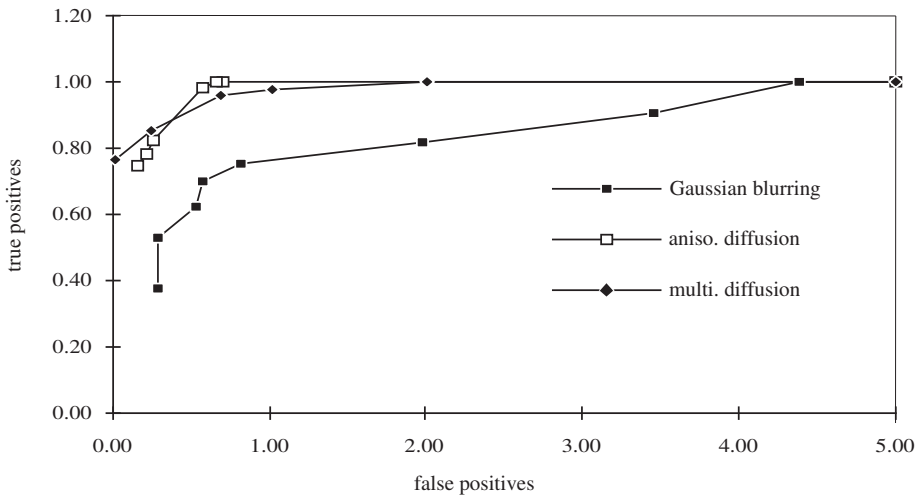


Figure 5.27: The ROC curves for the *shape* image, with 100% GWN and subjectively-define boundaries indicate that $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ and $\mathcal{G}^{\text{aniso}}_{t,k}$ outperform $\mathcal{G}^{\text{iso}}_{s}$.

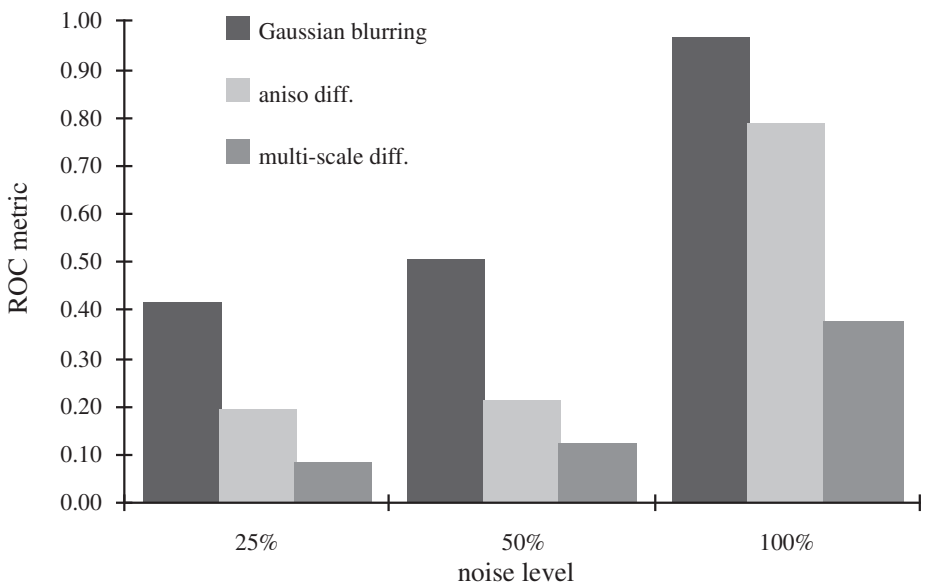


Figure 5.28: The ROC metric $\mathcal{M}_{\text{ROC}}$ shows $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ outperforms both $\mathcal{G}^{\text{aniso}}_{t,k}$ $\mathcal{G}^{\text{iso}}_{s}$, especially at high levels of noise.

numbers of true positives associated with both multiscale and anisotropic diffusion indicate the that these processes (in conjunction with the Canny edge detector) can find more of the correct edges for a given number of spurious edges. The ROC curves for $\mathcal{G}^{\text{multi}}_{k,s_0,\alpha}$ and $\mathcal{G}^{\text{aniso}}_{k,t}$ are comparable, but they both show a consistent improvement over $\mathcal{G}^{\text{iso}}_{s}$. Figure 5.28 shows the values of $\mathcal{M}_{\text{ROC}}$ for several levels of uncorrelated noise. These results show the relative advantage of anisotropic and multiscale diffusion over Gaussian blurring. The subjective edges do not include all of the boundaries that are indicated by the small-scale image structure, yet the ROC analysis shows that this simplification does not change the qualitative relationships among the various processes. Multiscale diffusion is still "better" than anisotropic diffusion, which is "better" than Gaussian blurring.

The *head* image

The use of subjective boundaries provides a means of analyzing the performance of various processes on the *head* image. Figure 5.29 shows the ROC curves associated with $\mathcal{G}^{\text{iso}}_{s}$, $\mathcal{G}^{\text{aniso}}_{k,t}$, and $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$. This analysis confirms the relative performance of these process on the *head* image. Figure 5.30 shows a bar graph of $\mathcal{M}_{\text{ROC}}$ for each process for different levels of noise. These results indicate that for subjectively-defined edges both diffusion processes offer significant improvements over Gaussian blurring *even when one is considering less-than-flat test images.* The multiscale diffusion provides a relatively greater advantage at higher levels of noise. These results provide evidence to support the strategy of finding piecewise approximations to images as preprocessing step in edge detection. The *head* image is not piecewise constant, yet the edges that result from a piecewise constant approximation are good relative to those that result from either the identity process or Gaussian blurring.

5.11 Correlated noise

The larger structures in the *head* image (such as the cerebellum or cortex) are obscured by the presence of smaller structures (the individual sulci, for instance) which represent a kind of correlated

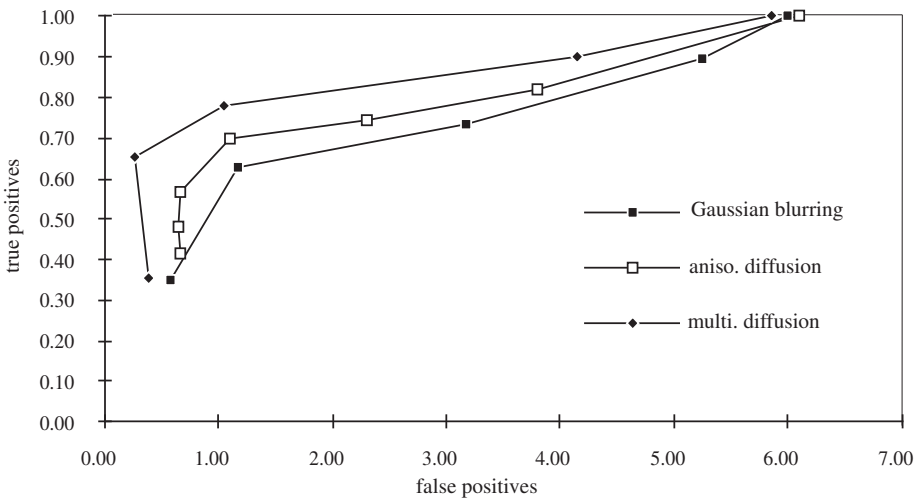


Figure 5.29: The ROC curves for the *head* image, with 100% GWN, indicate $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ outperforms both $\mathcal{G}^{\text{aniso}}_{t,k}$ $\mathcal{G}^{\text{iso}}_{s}$.

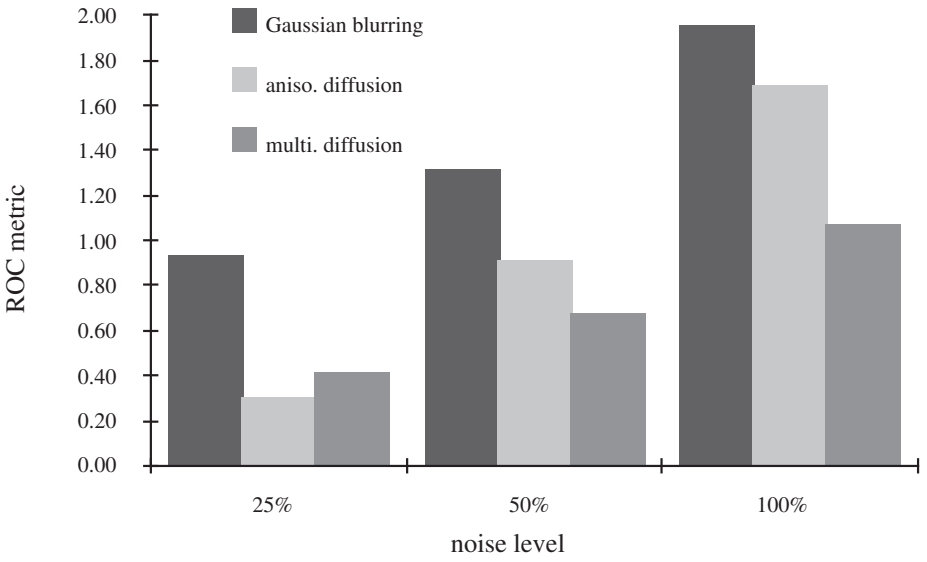


Figure 5.30: The ROC metric $\mathcal{M}_{\text{ROC}}$ shows $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ outperforms both $\mathcal{G}^{\text{aniso}}_{t,k}$ $\mathcal{G}^{\text{iso}}_{s}$, at higher levels of noise.

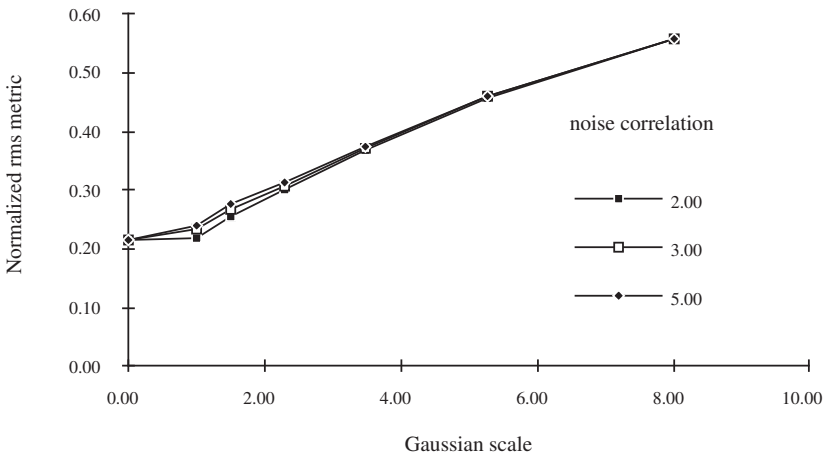


Figure 5.31: The metric $\mathcal{M}_{\text{rms}}$ as a function of isotropic scale $\mathcal{G}_s^{\text{iso}}$ for Gaussian blurring of the *shape* image and three different degrees of correlated noise.

noise. By *correlated noise*, I mean noise for which the pixel intensities share a statistical dependence within local neighborhoods in the image space. In this section, I model correlated noise explicitly and study the statistical behavior of the diffusion processes. Correlated noise is important in its own right, because the scale-space behavior of correlated noise is different from that of Gaussian white noise. Specifically, the pixelwise variance of correlated noise does not decrease as rapidly with increasing scale; correlated noise does not "blur away" as easily.

There is a range of possibilities for producing correlated noise in a stochastic manner. One option is to generate a field of GWN and blur this field with a low-pass filter (such as a Gaussian). For the experiments of this section I use a kind of "patch" noise that is produced by an anisotropic diffusion of blurred GWN (called CGN) as described in Sect. 5.5. This noise is relevant because it includes relatively sharp boundaries between patches unlike the CGN which produces only "smooth" structures. These sharp boundaries in the patchwise noise produce high gradient gradient values (when measured at small scales), making the edge detection in the presence of such noise difficult.

For this analysis I use patch noise with three different *degrees of correlation* but with a single pixelwise intensity variance of 25% (which is measured as a percentage of the average contrast). The degree of correlation is expressed as the average patch size (p from Sect. 5.5), and I use patch sizes 2.0, 3.0, and 5.0 pixel units.

Figure 5.31 shows the behavior of the $\mathcal{M}_{\text{rms}}$ metric for the *shape* image over a range of scales for these three degrees of correlation. These results indicated that *Gaussian blurring does not improve the rms error associated with patchwise correlated noise*. The smallest scale shown on this graph (1.0) has a normalized rms error of 0.228. The error associated with no processing is 0.223 (error introduced by pixelwise variance alone). This indicates that one is better off with the identity process under these circumstances. The reason is apparent in Fig. 5.32, which shows the *unnormalized* metrics $\mathcal{M}_{\text{mean}}$ and $\mathcal{M}_{\text{var}}$ as a function of scale. I present these metrics without normalization so that one can observe the magnitude of the error introduced by variance and bias. The mean difference (bias) increases more quickly than the variance decreases. Small-scale structures in the *shape* image blur away more quickly than noise.

The situation is different for a test image with less small-scale structure. The *block* image (Fig. 5.7b) contains a single "large-scale" object and does not contain the numerous small-scale structures (dots and squares) of the *shape* image (Fig. 5.7a). Because the structure of the *block* image is more robust with respect to blurring, it offers the opportunity for improved rms error with use of Gaussian blurring in the presence of correlated noise. Figure 5.33 shows the rms error

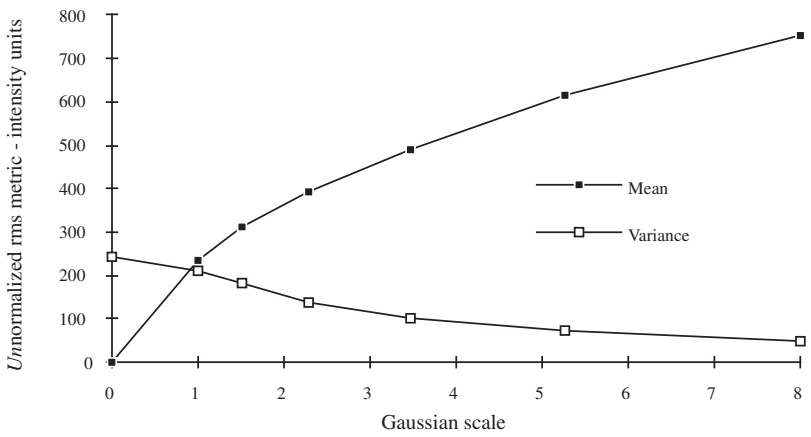


Figure 5.32: $\mathcal{M}_{\text{mean}}$ and $\mathcal{M}_{\text{var}}$ as a function of isotropic scale for Gaussian blurring of the *shape* image and correlated noise of $p = 3.0$.

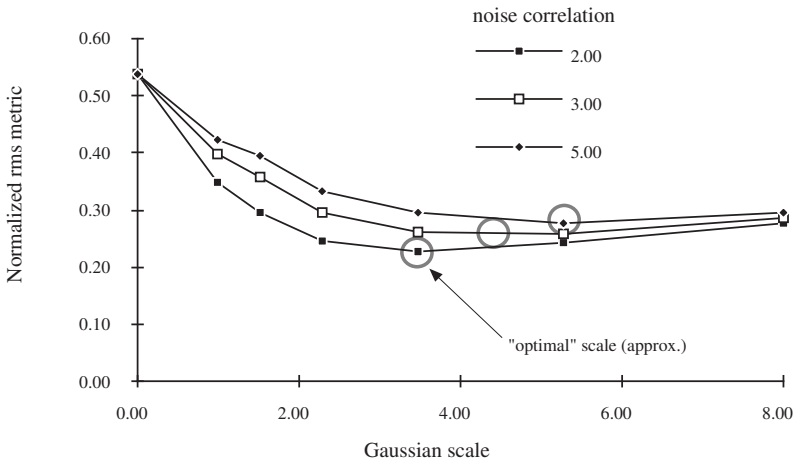


Figure 5.33: $\mathcal{M}_{\text{rms}}$ as a function of isotropic scale for the *blocks* image in the presence of correlated noise.

associated with Gaussian blurring the *block* image in the presence of uncorrelated and correlated additive noise. The results show that the optimal scale increases with the degree of correlation (patch size). While uncorrelated noise requires only small amounts of Gaussian blurring (as in Fig. 5.10), correlated noise requires large amounts of blurring in order to optimize the error metric $\mathcal{M}_{\text{rms}}$.

Multiscale diffusion exhibits a behavior that is significantly different from Gaussian blurring. Experiments indicate (Fig. 5.34) that $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ provides a means of obtaining improved rms error in the presence of both small-scale structure (as in the *shape* image) and correlated noise. Figure 5.34 shows a graph of the rms error versus scale (fixed conductance $k = 0.8$) for three different degrees of correlated noise. Unlike Gaussian blurring, multiscale diffusion preserves enough of the large-scale structure in the *shape* image to provide a net gain in rms error.

Figure 5.35 shows the *smallest* rms error ($\mathcal{M}^{O}_{\text{rms}}$) results for both Gaussian blurring and multiscale diffusion for various degrees of correlated noise in both the *blocks* and *shape* images. These results indicate that multiscale diffusion is better than Gaussian blurring in the presence of correlated noise. The ROC analysis confirms these observations. Figure 5.36 shows $\mathcal{M}_{\text{ROC}}$ for both Gaussian blurring and multiscale diffusion. The analysis of anisotropic diffusion in the presence of correlated noise is consistent with earlier results on uncorrelated noise; $\mathcal{G}^{\text{aniso}}_{k,t}$ is better than $\mathcal{G}^{\text{iso}}_{s}$ but not as effective as

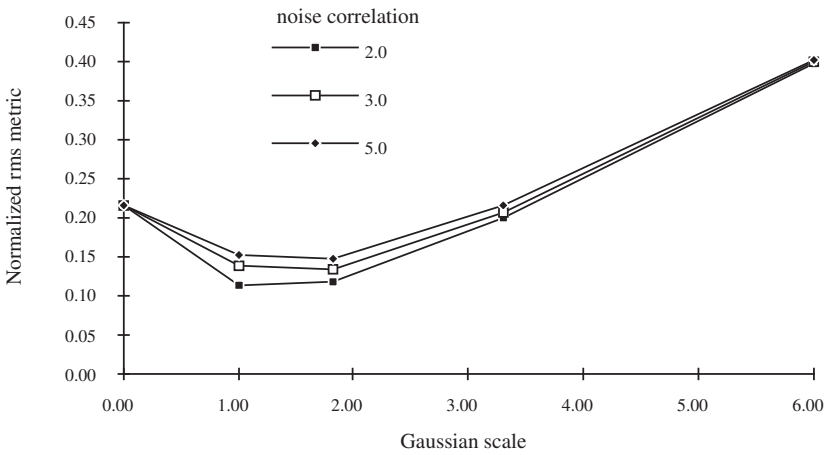


Figure 5.34: For multiscale diffusion applied to the *shape* image, the metric $\mathcal{M}_{\text{rms}}$ as a function of starting scale s_0 in the presence of correlated noise.

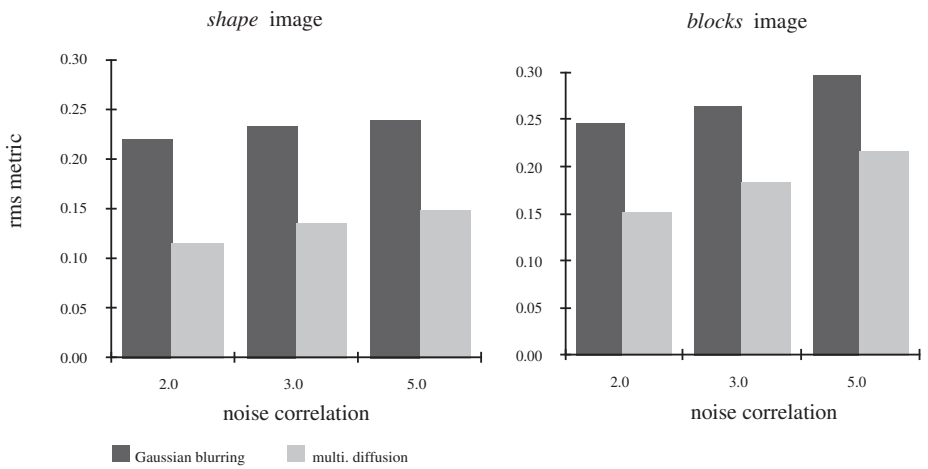


Figure 5.35: Comparisons of $\mathcal{M}_{\text{rms}}$ from Gaussian blurring and multiscale diffusion show the advantage of multiscale diffusion in the presence of correlated noise.

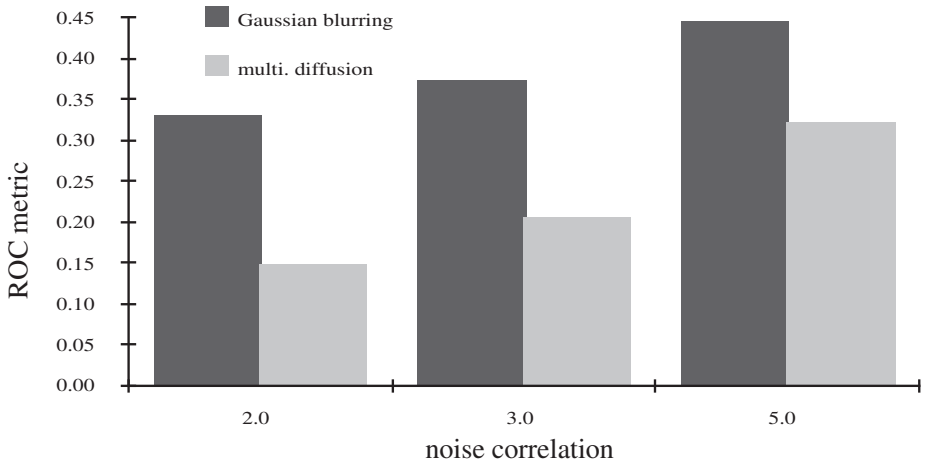


Figure 5.36: For the *blocks* image, comparisons of $\mathcal{M}_{\text{ROC}}$ for Gaussian blurring and multiscale diffusion show the advantage of multiscale diffusion in the presence of correlated (patchwise) noise.

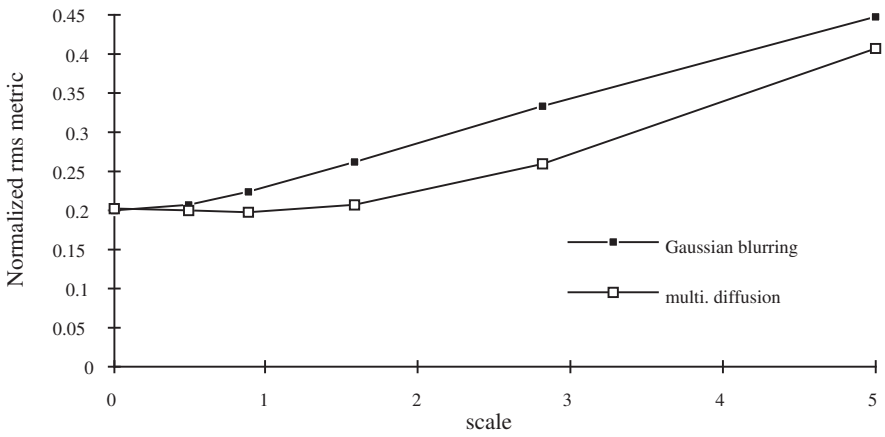


Figure 5.37: Comparisons of $\mathcal{M}_{\text{rms}}$ from Gaussian blurring and multiscale diffusion show that neither process is "effective" in the presence of boundary noise.

$\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$.

5.12 Boundary noise

The noise processes that are examined in previous sections are linear, that is a noisy image f' is constructed from a true image f by

$$f'(\vec{x}) = f(\vec{x}) + N(\vec{x}),$$

where N is the noise, both correlated and uncorrelated. Such processes act homogeneously over the entire image space, and they model only a small subset of the noise processes that are encompassed by Definition 2 from Sect. 5.2. This section examines a kind of noise, called *boundary noise*, which acts on the boundaries of grey-scale objects. The boundary noise described in Sect. 5.5 is used in order to study the effects of *variations in object shape* on Gaussian blurring and multiscale diffusion.

Experiments with boundary noise show that neither Gaussian blurring nor multiscale diffusion are capable of reducing the rms error associated with changes in object shape. Boundary noise can vary in both the scale (frequency of boundary perturbations) and degree (amplitude of those perturbations). Figure 5.37 shows the rms error associated with $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ and $\mathcal{G}^{\text{iso}}_{s}$ for one type of boundary noise. These results are indicative of wide range of amplitudes and frequencies of boundary noise. *For a range of boundary noise amplitudes and scales both Gaussian blurring and multiscale diffusion systematically* increased *rms error*. The multiscale diffusion process offers lower rms error than does Gaussian scale (especially at larger scales), but rms error systematically increases with the starting scale, so $\mathcal{G}^{\text{multi}}_{s_0,k,\alpha}$ does not offer any improvement over the identity process. The ROC analysis (Fig. 5.38) shows multiscale diffusion and Gaussian blurring to have comparable behaviors.

These results are somewhat unexpected. Because Gaussian blurring is a smoothing process that operates in both the gradient and level-curve directions ($\nabla \cdot \nabla f = f_{ww} + f_{vv}$), one expects such a process to "straighten out" the level curves of the image f and reduce the effects of boundary noise. One explanation for this poor performance is that the information lost in the gradient direction smoothing (f_{ww}) counteracts the improved reliability that should result from the level-curve smoothing (f_{vv}). Alternatively, one would expect the multiscale diffusion to perform significantly *worse* for this kind of noise, because the variable-conductance diffusion is constructed to preserve small-scale boundary information (which in this case is dominated by noise). Yet both $\mathcal{M}_{\text{rms}}$ and

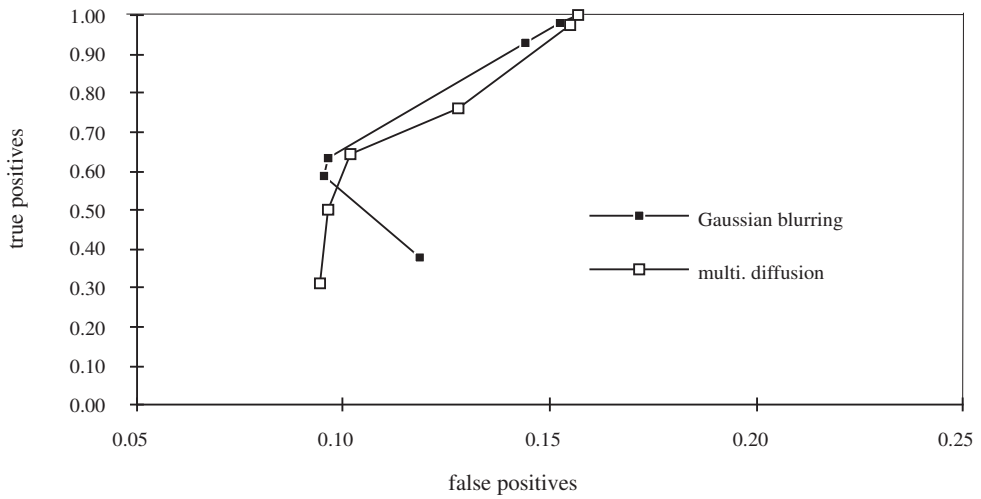


Figure 5.38: The ROC curves associated with Gaussian blurring and multiscale diffusion show the processes are comparable in the presence of boundary noise.

$\mathcal{M}_{\text{ROC}}$ indicate that variable-conductance diffusion does no worse than Gaussian blurring.

These results raise some questions about the effectiveness of the Monte Carlo analysis for this type of noise. It is possible that the *nonlinear* nature of this kind of noise renders either the particular metrics I use or the Monte Carlo simulation itself inappropriate. These issues warrant further research.

5.13 Discussion

This chapter begins by asking the question: How good is the geometry-limited diffusion process? In particular, how good is it with respect to noise? The chapter then proceeds to answer the question for zeroth-order geometry by giving a general definition of noise and then a framework for analyzing noise which relies on a Monte Carlo simulation. The Monte Carlo analysis is a strategy for studying the noise sensitivity of image processing algorithms. Results from the experimental analysis hinge on three important factors, the types of noise, the metrics that are used to quantify the behavior, and the definition of "truth".

I proposed several types of noise, uncorrelated, correlated (patch), and boundary noise. I defined metrics which fall into two classes: function-based metrics and structure (edge) based metrics. The first class has been reduced to a single metric $\mathcal{M}_{\text{rms}}$, which is the root-mean-squared error of intensity, measured pixelwise and then summed over the image. The second class of metrics involved the detection of structures (in this case edges) and then classified structures as true positives and false positives using fuzzy sets for graded notions of truth. True and false positive responses were summarized into a single metric, $\mathcal{M}_{\text{ROC}}$, using an ROC analysis.

A study of these measures under various noise conditions and test images leads to the following conclusions:

- For piecewise-constant test images in the presence of uncorrelated noise, the optimal rms error associated with Gaussian blurring occurs for relatively small kernels (≈ 1.0 pixel units). The appropriate scale increases with increasing noise levels.
- For images that are piecewise constant, anisotropic diffusion offers a better restoration (rms error) of an image than does Gaussian blurring. The relative magnitude of this effect, however, diminishes with increasing levels of noise.

- For images that are piecewise constant anisotropic diffusion offers a better tradeoff (compared to Gaussian blurring) between true and false positive edge estimates ($\mathcal{M}_{\mathrm{ROC}}$).

- For piecewise-constant test images and uncorrelated noise, multiscale diffusion appears to be equivalent to anisotropic diffusion in rms error, but offers a better tradeoff between true and false positive edge estimates—especially at higher levels of noise (100%).

- For images that are not piecewise constant (such as the *head* image) the comparisons between processes depend on the definition of "truth" (the goal). When the goal is taken to be the small-scale structure of the image itself, neither anisotropic diffusion nor multiscale diffusion offer any advantage over Gaussian blurring. When the goal is taken to be a set of subjectively-defined object boundaries, the ROC analysis indicates that multiscale diffusion is better than anisotropic diffusion, which is better than Gaussian blurring. Multiscale diffusion offers a relatively greater advantage with increasing levels of noise. *These results help to validate the strategy of breaking an image into piecewise-constant patches before segmentation.*

- In the presence of correlated noise, the effects of Gaussian blurring depend on the size of structures in the image. For images with small scale objects (the *shape* image), the rms error associated with Gaussian blurring increases monotonically—one is better off to not blur. For images without small-scale objects, Gaussian blurring achieves some optimal level of rms error at scales that are related to the size (degree of correlation) of the noise and the signal.

- In the presence of correlated noise, multiscale diffusion can improve the rms error of images, *even* those images which contain small-scale objects (*shape* image). *These data support the proposition that the multiscale diffusion process can preserve small-scale structure.* ROC analysis also shows that multiscale diffusion performs better than Gaussian blurring in the presence of correlated noise.

- Neither Gaussian blurring nor multiscale diffusion were able to improve on the rms error associated with boundary noise. The ROC analysis shows that Gaussian blurring and multiscale diffusion offered similar trade-offs between true and false positive edges. These results do not adhere to expectations, and they suggest that the issue of boundary noise (and more generally nonlinear noise processes) needs further study.

Besides an analysis of various segmentation strategies, this chapter provides some insight into the new approach for analyzing a range of low-level computer vision (particularly segmentation) algorithms. Generally, there are two main criteria for deciding if an evaluation strategy is worthwhile. The first is the principles and assumptions on which the strategy is founded, and the second is the usefulness (specificity, consistency, etc.) of the results that it produces. The method that I have proposed here makes a number of tenuous and, in some cases, arguable assumptions about low-level computer vision algorithms. This is especially true in the construction of the metrics, which isolate specific aspects of the segmentation problem without providing a comprehensive description of shape. Despite this, the approach appears to have merit. It has offered some very useful insights into the algorithms that are proposed in this dissertation, and many of the results are consistent with our subjective observations and experiences. The Monte Carlo analysis appears to be capable of providing objective, quantitative insights into complex processes and offers the possibility of a *scientific* approach to computer vision. Using this scientific approach, a researcher conducts a systematic empirical evaluation of an complex algorithm in order gain insights into its general behavior. Complex algorithms are understood by designing experiments which isolate particular aspects of their behavior.

This Monte Carlo analysis strategy requires carefully (and creatively) constructed noise processes and metrics in order to provide useful results. The metrics used in this chapter are useful, but

decidedly incomplete. While the fuzzy-set ROC approach offers a means of evaluating structure detection, it is lacking in its ability to quantify the global shape properties of image patches. Future research in this area should include the evaluation of more metrics that capture more complex properties of image segmentations. Also promising is the possibility of using *subjective* goals for evaluating the performance of a process. While I have used a single hand-drawn set of contours, this approach could incorporate accumulated sets of contours (which could also prescribe a tolerance). These contours could be specified by groups of subjects in order to free the analysis from the personal bias in the hand drawings of a single researcher.

One of the greatest limitations of the results of this chapter is their dependence on the particular test image. How general are these results, and how much would vary from one test image to another? One solution to this problem is to calculate the metrics by averaging over a "bread basket" of relevant test images. These groups of test images could be drawn from an area of application ("sagittal-MRI-head images", for instance), or they could be generated stochastically (according to some statistical model of expected image properties (Johnson 1993)) so that they have properties that are relevant to the area of application.

Despite these challenges, the strategies proposed in this chapter provide a relatively rare opportunity to study complex, nonlinear algorithms in a systematic manner. These techniques promise to be an area of future research.