Chapter 7

Summary and Discussion

7.1 Summary of results

The objective of this research was to develop new techniques for image segmentation and structure extraction. Several requirements drove this development effort:

- *Demonstrable* reliability in the presence of small-scale noise (small relative to relevant image structure)
- The ability to measure or detect higher-order geometry within a single, general framework
- Geometric invariance under orthogonal group transformations

Much of this research was motivated by the need to segment images and to find the boundaries of discrete patches. A great deal of previous work has focused on the problem of detecting "edges" by isolating places of "high" gradient magnitude. This strategy has two major shortcomings. First, the differential measurements associated with the gradient magnitude are sensitive to noise. Second, such a strategy defines algorithms for segmenting images on the basis of luminance (light versus dark regions), but it does not provide a general framework for the segmentation of images according to higher-order geometric information.

I have proposed a strategy for segmenting images that relies on the local homogeneity of *geometric descriptions*. A geometric description is a set of measurements that characterize the local shape of the intensity surface at every point in the image space—geometric descriptions are dense feature maps. *Regions* are places in an image where variations in the geometric description are small within local neighborhoods, and *boundaries* are the loci of places where variations in the geometric description are large.

I have assumed that scale is an essential aspect of any segmentation algorithm. The construction of a geometric description of an image requires a series of measurements. Measurements of images, as physical observables, must have an associated scale. Furthermore, no single scale can perform properly in all situations, so a segmentation algorithm should include scale as a *free parameter* which controls the size of objects that are of interest. Gaussian kernels are scalable filters that exhibit the appropriate symmetries, and they serve as neighborhood operators for making geometric measurements. Gaussian kernels can also have the desirable impact of reducing uncorrelated noise. However, isotropic Gaussian blurring also has the undesirable effect of undermining interesting image structures and making the boundaries of regions more difficult to locate.

The Gaussian kernel is the fundamental solution of the diffusion equation. Thus, the diffusion equation is an evolution equation that describes the changes in images that result from progressively larger Gaussian kernels. This fact motivates a variable-conductance diffusion equation which accounts for local image structure and preserves interesting structures that might be essential to subsequent segmentation algorithms. This structure-preserving behavior is achieved by selectively lowering the conductance associated with the diffusion process. The primary result of this dissertation is a framework for constructing *variable-conductance* diffusion processes. These processes are "tailored" to capture particular geometric structures that are relevant to the task. This framework, geometry-limited diffusion, derives from a series of generalizations of the variable-conductance diffusion process proposed by Grossberg (1984) as a model of human vision, and later by Perona and Malik (1990) in the form of anisotropic diffusion. I have presented an analysis of the anisotropic diffusion process, as well as the generalizations that I have made which include scale, multi-valued functions, and higher-order geometry.

I have pursued three different *modes* of research in the study of these processes. The first is a mathematical analysis, which shows how the development from Gaussian scale, to anisotropic diffusion, and to multiscale and geometry-limited diffusion results from a number mathematical generalizations. This analysis also gives insight in the general behavior of these process. The second mode of research is the subjective evaluation of specific examples. These examples allow qualitative comparisons of the behavior of various diffusion processes acting on real and artificial images in the presence of noise. The third mode of research in this dissertation is the empirical analysis of Chapt. 5. This is an objective, systematic analysis which relies on a set of metrics to quantify the behavior of ensembles of images with respect to some goal. The combination of the three research strategies; the mathematical analysis, the study of examples, and the empirical analysis; offers a great deal of insight into geometry-limited diffusion and the possible value of this technique in computer vision and medical image processing.

7.1.1 Anisotropic diffusion

Gaussian blurring, while reducing unwanted small-scale structure in images, has the apparent drawback of undermining larger structures. This can make subsequent detection and segmentation tasks difficult. Gaussian blurring is intimately tied to the heat or diffusion equation—the Gaussian kernel is the fundamental solution (also the impulse response) of the heat equation. That is,

$$\frac{\partial f}{\partial t} = \nabla \cdot \nabla f,$$

so that

$$f(\vec{x}, t) = G(\vec{x}, t) \otimes f(\vec{x}, 0) \text{ and } G(\vec{x}, t) = \frac{1}{(4\pi t)^{\frac{n}{2}}} \exp\left(-\frac{\vec{x} \cdot \vec{x}}{4t}\right).$$

Anisotropic diffusion is a variation on the heat equation which includes a space and time varying conductance term that depends on the gradient magnitude:

$$\frac{\partial f}{\partial t} = \nabla \cdot g(|\nabla f|) \nabla f, \tag{7.1}$$

where previous authors and I have used $g(|\nabla f|) = \exp(-\frac{|\nabla f|^2}{2k^2})$. Examples show that anisotropic diffusion can reduce noise and preserve or enhance the boundaries of larger objects in images. The mathematical study of local behavior confirms this edge-enhancing behavior and shows that Eq. (7.1) incorporates a kind of threshold (determined by the conductance parameter k) which draws a division between those edges that are enhanced and those edges that are blurred away.

The empirical analysis of anisotropic diffusion confirms the qualitative observations. Indeed, comparisons of Gaussian blurring and anisotropic diffusion show that anisotropic diffusion offers better rms-error *restorations* of piecewise constant images in the presence of uncorrelated noise. For uncorrelated noise at levels of 25% (of images with 256 × 256 pixels), anisotropic diffusion can offer a 70% advantage in rms error ($\mathcal{M}_{\text{rms}}$) over Gaussian blurring. The relative magnitude of this improvement becomes progressively smaller with increasing noise levels. At the 100% noise level, the improvement associated with anisotropic diffusion drops to less than 50%.

Under the same conditions, anisotropic diffusion also offers a better tradeoff between true and false positive edges. This is indicated by the ROC analysis and the associated metric $\mathcal{M}_{\text{ROC}}$. For Gaussian blurring at small scales, increasing the scale results in a consistent decrease in *both* true and false edges measurements. For very large scales Gaussian blurring distorts the shapes of object boundaries, and it can *increase* the number of false positives. Anisotropic diffusion shows a very different behavior. With increasing time (scale) it offers a consistent decrease in false positives with little or no decrease in true positives. At very large scales true positives diminish while false positives continue to decrease, indicating less distortion of boundaries relative to Gaussian blurring.

An analysis of the conductance parameter shows that poor choices of conductance k result in relatively poor results. However, when k is measured in terms of the rms gradient magnitude of the image, there is a range of k values (0.6–1.2) that offer very good results. Greater levels of noise require greater values of k. This helps to explain the diminishing advantage of anisotropic diffusion with progressively greater levels of noise. For increasing noise, the best performance of anisotropic diffusion is achieved with larger values of k, forcing anisotropic diffusion to behave more like Gaussian blurring.

In cases where the goal is *not* piecewise constant, anisotropic diffusion does not offer measurably better rms restorations. However, an evaluation with respect to subjective, user-defined boundaries shows that anisotropic diffusion offers an advantage over Gaussian blurring in the detection of edges. That is, while anisotropic diffusion does not offer better restorations of the original image (before the introduction of noise), it does provide patches with boundaries that adhere more closely to our subjective impressions of the objects in the image. This analysis supports the strategy of developing "useful" segmentations by first constructing piecewise constant approximations to images.

The effectiveness of the anisotropic diffusion process is limited by the inability to make reliable edge measurements at small scale. Anisotropic diffusion, as presented here, relies on a gradient measurement to control the conductance, $g(|\nabla f|)$, and this gradient measurement has an associated scale. In most discrete implementations (Perona and Malik 1990, Nordstrom 1990, Gerig, Kubler, Kikinis and Jolesz 1991) these measurements are implemented as nearest neighbor differences. Such measurements are essentially "small scale" and are not useful in the presence of certain kinds of noise. In situations with very poor signal to noise, the relative advantage of anisotropic diffusion over Gaussian blurring is diminished. This has been shown qualitatively in Chapts. 2 and 3, and empirically in Chapt. 5. In cases where image structure is discerned only at very large scales, the process is virtually ineffective. The construction of a multiscale diffusion process, which combines information over a range of scales, alleviates this dependency on small-scale structure.

7.1.2 Multiscale diffusion

Scale is an essential aspect of image measurement. Thus, the gradient measurements that control the flow of the anisotropic diffusion process must incorporate some scale. As stated above, anisotropic diffusion implicitly assumes that the relevant scale is the "inner" scale, pixel scale, or otherwise the smallest scale available in the discrete implementation. I have proposed a process, multiscale diffusion, which makes this measurement scale explicit, and I have tied this scale parameter s to the evolution parameter t:

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

This scale $s(t)$ is the size of the Gaussian kernel used in the gradient measurement $|\nabla_{s(t)} f|$, and it determines the size of structures that influence the "flow" of intensity in the diffusion process. The scale $s(t)$ should decrease over time in order to reflect the "increased confidence" that results from the variable-conductance diffusion process.

Multiscale diffusion combines image information over a range of scales by decreasing the scale of the conductance measurements. Examples indicate that this large-to-small-scale strategy allows one to measure small-scale structures that are associated with the boundaries of large-scale objects. Compared to the anisotropic diffusion process, which measures conductance at a single, relatively small scale, the multiscale approach is less sensitive to noise and is particularly effective in the presence of correlated noise. Examples show that the multiscale diffusion process is able to group smaller objects together to create patch boundaries which are consistant with our own subjective impressions of larger objects. This is shown in the grouping of dots in Fig. 3.5 and in the empirical analysis of Chapt. 5.

The empirical analysis showed some advantage in multiscale diffusion over anisotropic diffusion. This advantage was particularly evident in situations with high noise levels or correlated noise. The comparative behavior of theses processes depended on which metrics were used. For the function-based metric, $\mathcal{M}^{\mathcal{O}}_{\mathrm{rms}}$ (the lowest rms intensity error among all possible scales), and low to moderate levels of uncorrelated noise, the multiscale process offers no advantage over anisotropic diffusion. However, the structure-based metric $\mathcal{M}_{\mathrm{ROC}}$, which quantifies the scale-space tradeoff of true and false positive edge measurements, shows that multiscale diffusion does offer a significant advantage over anisotropic diffusion for all levels of uncorrelated noise.

The advantage of multiscale diffusion is particularly evident when evaluating the subjective boundaries that are measured on the *head* image from Fig. 1.1. These results indicated that multiscale diffusion introduces a 50% improvement over anisotropic diffusion. This advantage stems from the fact that the subjective boundaries I have identified in the head image were "medium-scale" objects (the brain stem, cerebellum, scalp, tongue, etc.). Such objects are not easily identified with small-scale measurements, and the anisotropic diffusion process picked up many smaller-scale structures within these objects. The subjective boundaries that I chose for the head image have relegated many smaller anatomical structures to the status of noise, and the anisotropic diffusion equation was confounded by the presence of this correlated noise.

A systematic analysis of correlated noise confirmed these findings. For patchwise-correlated noise, Gaussian blurring was rendered ineffective at improving the rms error of images. The *interference* of correlated noise depends on the size of structures in the original image. Correlated noise is particularly hard to remove when its size approaches the size of interesting structures in the image. In such cases multiscale diffusion did improve rms error and offer a significant advantage in the numbers of true and false boundaries edges that are detected.

7.1.3 Geometry-limited diffusion

Variable-conductance diffusion can also aid in the detection of higher-order geometric structures. I have generalized the anisotropic diffusion process to include *multi-valued images*. Multi-valued images are mappings $\vec{F}$ from the image space to an m-dimensional feature space. The mathematical framework associated with multivalued diffusion is somewhat general. This framework encompasses the application of variable-conductance diffusion to multi-valued images that result from imaging devices capable of producing multiple data values at each point in the image space. Examples are multiecho MRI (Gerig, Kubler, Kikinis and Jolesz 1991, Yoo and Coggins 1993) and multispectral Landsat images. This framework can also apply to multi-valued images that result from applying sets of filters to single-valued images. multi-valued diffusion can apply to *geometric descriptions* of images, which are multi-valued images that describe local image structure. I have called this application *geometry-limited diffusion*.

In much of this dissertation I discuss geometric descriptions that consist of sets of image derivatives. These derivatives represent local image structure through the truncated Taylor series expansion of $f(\vec{x})$, which exists at each point in the image space. In this way, variable-conductance diffusion operates not on intensity but on sets of derivatives of the original (input) image. The conductance $g(\mathcal{D}\vec{F})$ is controlled by the *dissimilarity* operator. The dissimilarity operator $\mathcal{D}$ is a generalization of the gradient magnitude as used in anisotropic diffusion, and it is a measure of the local inhomogeneity of the multi-valued image $\vec{F}$. I propose the Euclidean norm of the Jacobian of $\vec{F}(\vec{x})$ as a definition of the dissimilarity. The dissimilarity also requires a *shape metric*, ϕ, which quantifies differences in geometric descriptions. This metric defines the measure of distance in the feature space and thereby the kinds of structures that are preserved in the diffusion process.

These developments lead to a general framework which has several unspecified items:

1. The particular geometric description or, in the case of image derivatives, the *order* of the derivatives used in the description

2. The shape metric ϕ, which determines the kinds of structures that are preserved under diffusion

3. The scale that is used in the conductance term

4. The value of the conductance parameter k

These undetermined parameters give the geometry-limited diffusion framework its generality. Their specification depends on the application and they are likely to be decided on a case-by-case basis. Chapt. 4 presented several examples of the use of this framework and developed some guidelines that constrain the parameter space.

A diffusion process on first-order derivatives provides a reliable mechanism for detecting *creases* (ridges and valleys) of image intensity. The development of this process follows from the observation that creases are, by at least one definition, places of local inhomogeneity in first-order information. In two dimensions, the pair of first order derivatives, f_x and f_y, form a 2-valued function, $\vec{F}(\vec{x}) = F_1(\vec{x})\ F_2(\vec{x})$, and the dissimilarity measures the derivatives of gradient direction (derivatives of normalized first-order features). This process breaks images into patches based on the local homogeneity of the *gradient direction*. The boundaries of these patches are creases, and several second-order measurements provide a means of characterizing these creases by their mean curvature, deviation from flatness, etc.

Geometry-limited diffusion incorporates a nonlinear blurring of image derivatives, and these derivatives are treated as separate values or *features* (in the statistical-pattern-recognition sense of the word). Despite the separate, nonlinear treatment of the derivatives, the process remains invariant to orthogonal group transformations, as long as the dissimilarity is constructed as a Euclidean norm of a tensor. Orthogonal group invariance means that values of scalar invariants are independent of the initial choice of basis (except that it must be orthogonal) used to construct the derivatives. The invariant behavior of geometry-limited diffusion depends on the invariance of the dissimilarity—this fact provides a further constraint on the choice of dissimilarity measures.

Besides orthogonal group invariance, geometry-limited diffusion can provide dissimilarity measures that are invariant to greylevel (intensity) transformations. The dissimilarity associated with the first-order diffusion process is invariant to monotonic intensity transformations. Measures that are invariant to linear greylevel transformations are achieved by normalizing the geometric features when calculating dissimilarity. Such dissimilarity measures are independent of the units of intensity. This has practical implications in the choice of the conductance parameter, k. Greyscale-invariant dissimilarity measures enable one to choose k without regard for the "brightness" or contrast associated with a particular image. This property makes it easier to choose values of k that are appropriate for a wide range of input images.

These principles apply to the five-dimensional feature space associated with first- and second-order derivatives. While there is a variety of options for choosing a dissimilarity measure, a straightforward multiplication of first- and second-order tensors in conjunction with the appropriate normalization provides interesting results. Such a dissimilarity measure breaks images into patches that have as boundaries both the "edges" and "middles" of objects (assuming objects are light on dark, or vise versa). The combination of first- and second-order information provides a rich selection of geometric invariants for characterizing the boundaries between patches.

7.2 Spectra-limited diffusion

One of the underlying principles of this work is that local inhomogeneity offers a means of constraining the diffusion of local image descriptions. This principle applies to the analysis of textures. I have generalized variable-conductance diffusion to the blurring of *envelopes*. An envelope is a representation of a function in terms of the magnitude and phase of a characteristic function. For this work I used sinusoids as characteristic functions, and the result is a set of reaction-diffusion equations. Each equation has a characteristic frequency (the frequency of the sinusoidal characteristic function), and these equations describe processes that evolve by the "spreading" or "flow" of *wave packets*. These wave packets resemble Gabor blobs, sinusoids modulated by Gaussians. As with geometry-limited diffusion, the process is nonlinear, and the flow of these blobs slows at places of high dissimilarity. A mathematical analysis showed that a group of such processes, each member with its own characteristic frequency, can compute a patchwise Fourier transform of an image.

Further analysis showed that this approach is a special case of the geometry-limited diffusion framework. In this special case, local shape is captured via a nonstationary procedure which is a multiplication of the image with a space-varying sinusoid. The principles of dimensionless metrics suggest a dissimilarity measure that is sensitive to the variations in the *relative* frequency information in the local image structure. The result, spectra-limited diffusion, is a nonlinear, multiscale process for segmenting textured patches in images. Early examples of this approach are promising, but more research is necessary in order to know if this technique is useful for general-purpose texture segmentation.

7.3 Empirical analysis

There is no shortage of computer vision and image processing techniques— new algorithms abound. Which algorithms are better? And how does one decide? I propose a statistical framework for the evaluation of such algorithms. This method quantifies the sensitivity of a process to various kinds of image variations. The strategy is to stochastically generate ensembles of images that contain the same (or similar) *correct answer* for a specific task. Run the algorithm on each image, and quantify the ability of that algorithm to discern that correct answer in a consistent, reliable manner. The success of this evaluation approach depends on the construction of useful metrics. These metrics must be relevant to the task and the noise processes, and they must be sensitive enough to quantify performance differences in various processes. I have asserted that useful computer vision algorithms should be both reliable and accurate, and I have proposed two classes of metrics in order to quantify these criteria.

For the purposes of this dissertation, the empirical analysis of Chapt. 5 provided a means of quantifying the gains in performance associated with the variable-conductance diffusion processes. However, the Monte Carlo analysis that I have presented is interesting in its own right as a general-purpose tool for evaluating image processing algorithms. This Monte Carlo approach is particularly useful in situations where the statistical behavior of an algorithm is difficult to describe analytically. This is often the case with relatively complex nonlinear algorithms, such as geometry-limited diffusion. Because such algorithms proliferate in vision and image processing, techniques for conducting systematic, empirical comparisons will become even more important.

I have presented the Monte Carlo framework in a general form. The definition of noise encompasses a wide range of possible image and scene variations. In a scene, these variations could include such diverse characteristics as viewpoint, lighting, and object location. In an image, interesting characteristics could be rotation, deformation, and translation, as well as the types of additive noise that are studied in Chapt. 5. This Monte Carlo method should be useful for studying any algorithm which is meant to be insensitive to such variations. The challenges are first, to find a means of generating samples of noisy images, and second, to construct insightful metrics which quantify meaningful properties of the algorithms under study.

The metrics are intended to quantify how an ensembles of images S differs from a goal. I have constructed two classes of metrics based on two different image representations. The first class is that of *function-based* metrics which rely on the conventional representation of an image as a function defined over the images space, I. These are metrics which compare images point by point and aggregate pointwise measures over the entire image space. The result is a set of three metrics $\mathcal{M}_{\mathrm{var}}$, $\mathcal{M}_{\mathrm{mean}}$, and $\mathcal{M}_{\mathrm{rms}}$. The variance metric, $\mathcal{M}_{\mathrm{var}}$, measures the average pixelwise variance within the ensemble S, and it quantifies error within the ensemble due to statistical variation, or noise. The metric $\mathcal{M}_{\mathrm{mean}}$ captures the average distance from the goal, and it quantifies systematic biases across the entire ensemble. The rms metric, $\mathcal{M}_{\mathrm{rms}}$, quantifies *average error* and it combines error that results from systematic bias and statistical variation.

The metrics in the second class are *structure-based*; these metrics rely on structure sets as a means of comparing images. A structure set is a subset of the image space which is distinguished by the application of a detector to an image or set of images. Feature-set images are summed over an image ensemble in order to produce an accumulator image, which is a spatial histogram of detected structures. The accumulator is treated as a *fuzzy set* and it can be compared to a fuzzy goal which consists of a structure set and a specified tolerance. I have analyzed the performance of the diffusion processes based on the number of true and false (in the fuzzy sense) structure estimates that result from applying a structure detector to the processed image. The number of true and false measurements over a range of scales generates an *ROC curve*, and the "goodness" of this curve is quantified by evaluating the integral of the area above the curve (denoted $\mathcal{M}_{\mathrm{ROC}}$).

The experience of this research suggests that the construction of metrics is difficult, but several different metrics can give useful insights into situations where a single, all-encompassing metric is not possible. While these metrics lack an explicit characterization of object (segment) shape, they were able to quantify small distortions, and they proved to be adequate for measuring the differences in the various diffusion processes. The Monte Carlo framework promises to have potential applications to the evaluation of other algorithms and visual tasks. However, its usefulness as a general tool can be verified only by applying it to other evaluation problems.

7.4 Relation to previous work

As discussed in previous chapters, this work represents the extension and generalization of previous work (Grossberg 1984, Perona and Malik 1990). These generalizations offer demonstrable advantages over the techniques proposed in this earlier work. These improvements are threefold:

1. The inclusion of image information at multiple scales into the diffusion process, making the results less sensitive to noise.

2. The generalization of the variable-conductance diffusion to multi-valued images and the development of dissimilarity, a generalized boundary measure, as a means of controlling conductance

3. The explicit connection between higher-order image geometry and variable conductance diffusion

Several authors (Nordstrom 1990, Snyder et al. 1992, Geiger and Gerosi 1991) have established the relationship between variable-conductance diffusion and various energy minimization processes. As such, the research in this dissertation has a relationship to a wide range of nonlinear processes that seek to minimize energy functionals. These nonlinear processes include those which attempt to maximize the posteriori probability distributions associated with Bayesian priors. This work includes that of Geman and Geman (1984), Leclerc (1989), Mumford and Shah (1989) and Johnson, Wong, Hu and Chen (1991) among others.

The differences are philosophical as well as practical. These authors typically study the creation of energy functionals and algorithms for minimizing the same. Some of these algorithms resemble anisotropic diffusion, while others do not. Two major differences are worth noting. First, these approaches are usually iterative, but they may include a stochastic term (as in simulated annealing) which improves the chances of finding a global (in the state space) energy minimum. Second, when used for image segmentation, these approaches often apply *explicit region labels* rather than intensities or feature values at each point in the image. Under such circumstances, the fuzzy neighborhood defined by the Gaussian kernel is replaced by the "clique" which is *grid dependent*. That is, it depends on the nature of the discrete sampling. These cliques do not exhibit the scale and space symmetries of the Gaussian neighborhood. This difference reduces the applicability of scale and geometry within the minimization processes. The work in this dissertation focuses on variable-conductance diffusion as a means of achieving an alternate scale space rather than an energy-minimization process. The inclusion of scale space and differential geometry in the diffusion process is an important and novel aspect of this work. The development of a framework for constructing conductance functions is also new.

There is, to date, a body of literature that deals with the notion of scale space. This work addresses a wide range of visual problems and issues, among them the issues of segmentation and structure extraction. Witkin and Tenenbaum (1983) as well as Koenderink and Doorn (1990) have proposed scale space as an important aspect of image analysis. Ter Haar Romeny and Florack (1991a) have discussed the geometric aspects of scale space and have proposed differential invariance as a means of constructing structure detectors. Blom (1992) has examined some statistical aspects of scale space, and he has proposed some specific structure detectors, such as a "corner" detector.

Much of this scale-space research incorporates the assertion that multiscale structure is essential to image analysis, but it begs the central question of how of to combine information at different scales. In this dissertation I have proposed a means of combining information at different scales using variable-conductance diffusion, and then I have addressed the issues of structure detection and geometric structure within images by generalizing the basic diffusion framework. Other efforts to combine information at different scales include Lindeberg (1990), Pizer et al. (1991), Coggins et. al. (1985, 1990), and Haring, Viergever and Kok (1993).

Lindeberg (1990) has proposed *scale-space blobs* for low-level vision, especially object detection. Pizer, Fritsch, Morse and others have constructed the *multiscale medial axis* as a means of object definition (and as a model of human vision). Both of these approaches have aspects that make them different from the multiscale diffusion approach. First, they both consider the full scale space, which is an $(n+1)$-dimensional space that results from an n-dimensional image. In doing so, they both assume some kind of commensurability between space and scale. Multiscale diffusion, on the other hand, treats a range of scales in a sequential fashion, large to small, and achieves the composition of scales via the net effect that the structure at different scales has on the diffused image. The second difference is that both of these other methods assert a basic, well-defined structure for image objects. In the scale-space blob approach, the boundaries of objects or structures are defined as the zero-crossings of the Laplacian (or some other differential invariant operator). In the case of the multiscale medial axis, objects are regions that are "attached" to a medial structure, which is inferred on the basis of multilocal boundary information. Multiscale and geometry-limited diffusion, on the other hand, rely on a simpler, more general definition of a region or object— *an object is a region of homogeneity in local image structure.* The relative merits and shortfalls of these approaches warrant further research.

Another vein of multiscale research is represented by the work of Coggins (1985, 1990), Haring et al. (1993), and Jones and Malik (1992). These authors create multi-valued images via a discrete sampling of scale space. Each point in the image is represented by a finite set of measurements made over different scales. Both authors treat the multi-valued image (each value represents a different scale) as a mapping into a higher-dimensional feature space and attempt to perform classification and segmentation tasks based on the structure of that space. These methods assume very little about the structure of scale space and infer the relationships between structure at different scales based on some supervised clustering or classification scheme. As it stands, the multiscale diffusion approach assumes a continuous, large-to-small progression of scale that results in the "focusing" near object boundaries. These sampled multiscale methods incorporate techniques for analyzing multi-valued images, and they could be integrated into the multi-valued diffusion framework. This a potential topic of future research.

Besides the work that is similar in either philosophical or algorithmic content, there is very large body of literature that deals with the basic visual problems of image segmentation or structure extraction. These various approaches typically exhibit interesting differences in three important aspects: (1) the underlying philosophy or principles from which the method is developed, (2) the particular algorithm or implementation, (3) the behavior and quality of the results. Much of this dissertation has focused on the development of the mathematical formulation, which is based on the underlying philosophy of local homogeneity of multiscale image structure. I do not consider the digital implementation of these formulas (covered briefly in the Appendix) to be an important or novel aspect of this work. Indeed, one of the advantages of the variable-conductance diffusion approach is that it lends it self to straightforward numerical implementations. Perhaps the most distinctive aspect of this work is the quality of the results. This quality is shown in numerous examples, which offer a means of subjective comparison, as well as the quantitative evaluation of Chapt 5, which offers objective, empirical evidence of the usefulness of these methods. I have not made a systematic study of a great number of other methods–I make direct comparisons of Gaussian blurring, anisotropic diffusion, and multiscale diffusion –but I have presented quantitative evidence of the relative effectiveness of these techniques, and I have set down some benchmarks which could aid future development and evaluation work.

7.5 New research directions

The results throughout this dissertation raise many questions. Many of the analyses involve simplifications of complex phenomena. These simplifications leave open the possibility of more in-depth analyses which could offer further insights into the techniques that I have proposed in this dissertation. In many cases I have noted the shortcomings of the research and have explicitly proposed more research on that particular subject. The purpose of this section is *not* to list all of the individual "loose ends" that arise from the limitations of this research. Instead, this section takes a broader view of the problem and examines some of the general issues that could motivate future research. Thus, I paint my descriptions of "new research directions" with an intentionally broad brush.

7.5.1 Energy-minimization algorithms

Anisotropic diffusion has been shown to bear a mathematical resemblance to a family of segmentation and image restoration algorithms that seek to minimize certain kinds of energy functionals (Nordstrom 1990, Snyder et al. 1992, Geiger and Gerosi 1991). Many of these energy functionals include a term which penalizes the variation in luminance, i.e.,

$$E(f) = \int_I |\nabla f|^2 \mathrm{d}\vec{x}.$$

The relationship with diffusion comes from the fact that the *first variation* (which defines the pattern of evolution in gradient-descent algorithms) of this term is the Laplacian:

$$\frac{\mathrm{d}E}{\mathrm{d}f} = \frac{\mathrm{d}\left[\int_I |\nabla f|^2 \mathrm{d}\vec{x}\right]}{\mathrm{d}f} = \nabla \cdot \nabla f.$$

If these energy functionals include a reduction of the penalty term near suspected edges, the resulting first variation takes the form of variable-conductance diffusion process. If $g(\vec{x})$ is an edge indicator, which *decreases* in the presence of an edge, then the first variation becomes

$$\frac{\mathrm{d}E}{\mathrm{d}f} = \frac{\mathrm{d}\left[\int_I |g(\vec{x})\nabla f|^2 \mathrm{d}\vec{x}\right]}{\mathrm{d}f} = \nabla \cdot g(\vec{x})\nabla f,$$

which is the variable conductance diffusion equation. The relationship between energy-minimizing approaches and anisotropic diffusion could offer some interesting insights into the research of this dissertation. Two issues come to mind:

1. How might the energy-minimizing paradigm be used to improve upon the techniques presented in this dissertation?

2. How might the techniques presented in this dissertation be used to improve upon existing energy-minimizing approaches?

With regard to the first item, consider the following. The energy-minimizing paradigm provides a framework for combining *different kinds* of image properties into a single algorithm. In this light, the anisotropic diffusion algorithm seeks to minimize only those energy terms of the form

$$|g(|\nabla f|)\nabla f|^2.$$

The introduction of more energy terms could create more complex evolution processes. Properly constructed energy terms could be used to "fix" some of the shortcomings of the variable-conductance diffusion processes I have proposed. Such shortcomings include the "rounding off" of corners (Sect. 3.3.4), the instability near places of low conductance (Sect. 3.3.3), and the inability to "fill in" very large gaps in image boundaries. Several authors have used the minimization approach to form more complicated evolution processes. Nordstrom, for example, introduces a "bias" energy term that keeps the diffusion process from straying too far from the initial data. Shah (1991) introduces a separate diffusion equation that explicitly models the behavior of boundaries and provides a means of obtaining more complicated behaviors at corners and line ends. There are many possibilities for applying such ideas to geometry-limited diffusion.

The energy-minimizing paradigm offers some other possibilities for improving the techniques of this dissertation. One possibility is the introduction of stochastic processes for energy minimization. How might such stochastic processes improve upon multiscale and geometry-limited diffusion?

Concerning the second item, the techniques of multiscale and geometry-limited diffusion incorporate some principles that could be used in some of the energy-minimizing paradigms. For example, one of the major drawbacks of the energy-minimizing approach is the difficulty in finding global energy minima. Deterministic mechanisms such as graduated nonconvexity generally do not guarantee optimal solutions. Could the multiscale diffusion process be a means of obtaining lower-energy solutions to such functionals? How does it compare with other deterministic algorithms?

The principles of multi-valued and geometry-limited diffusion are also applicable. One of the challenges within the energy-minimizing paradigm is the construction of energy functions. The various guidelines for constructing dissimilarity measures could be used to construct useful energy functions. Energy functions could include higher-order geometric structures, and geometric invariance could be used as a constraining factor in such circumstances. Information across a range of scales could be part of such energy functionals. Many of the techniques I present in this dissertation in the context of geometry-limited diffusion are relevant also to the construction and minimization of energy functionals.

7.5.2 Diffusion of local image descriptions

I have purposely described the techniques of this chapter in somewhat general terms with the hope that these techniques could have a wide range of possible applications. Yet the question remains, "How general is it?". That is, how general is the principle of using the local inhomogeneity of image structure in combination with variable-conductance diffusion. The application of multi-valued diffusion to other problems in computer vision has some promise. The examples of Chapt. 6 show that it has some potential for texture segmentation. A great number of other computer vision problems involve multi-valued data. Segmentation on the basis of color, optical flow, and time-series data are interesting examples.

Generally, such applications lend themselves to multi-valued diffusion if the task indicates a need for a spatially-local averaging (blurring) that preserves selected image structures. A number of deep questions must be addressed. For example, consider the application of these techniques to optic flow fields. Under what conditions does it makes sense to average over local neighborhoods in an optical flow field? Besides the basic issue of applicability, each new application will require a specification of "structure" that can be described by a dissimilarity operator. What structures ought to be preserved when blurring an optic flow field? The general applicability of the techniques in this dissertation depends on the ease with which such questions can be answered in new applications.