There is no conflict between causality and randomness or between determinism and probability if we agree, as we must, that scientific theories are not discoveries of the laws of nature but rather inventions of the human mind. Their consequences are presented in deterministic form if we examine the results of a single trial; they are presented as probabilistic statements if we are interested in averages of many trials. In both cases, all statements are qualified. In the first case, the uncertainties are of the form "with certain errors and in certain ranges of the relevant parameters"; in the second, "with a high degree of certainty if the number of trials is large enough."

- Athanasios Papoulis

Chapter 4

Multiscale Image Statistics

When digital images are considered as arrays of observations made of an underlying scene, the vocabulary and calculus of statistics may be applied to their analysis. If an image is subject to noise in pixel measurement, it should be presented within the context of either known or computed properties of the pixel values. These properties include the sample size or raster resolution and statistics such as the variance of the additive noise.

This is an introduction to the concept of multiscale image statistics. In particular, the next sections describe the generation of central moments of the local probability density of intensity values. A particular model of images as composed of piecewise regions having similar statistical properties (having similar probability distributions of intensity) is assumed for the construction of multiscale statistics. This model of images as samples of piecewise ergodic stochastic processes is presented after a brief introduction to provide a foundation for the rest of the chapter.

Later sections present the construction of multiscale central moments of intensity. An earlier section in Chapter 2 describes the use of central moments to reconstruct the probability density function uniquely. The approach presented here outlines the generation of the central moments of the local intensity histogram of any arbitrary order. Later sections provide examples of these local central moments up to the fourth order. Properties of these moments are explained, and their behavior is compared with other common image processing operators. The multiscale central moments are generalized to images of two dimensions as well as multivalued images containing two values per pixel.

Applications of multiscale central moments are included. In particular, the use of these measurements in the selection of control parameters in nonlinear diffusion systems for image processing are shown.

4.1. Background and Introduction

Statistical pattern recognition is a discipline with a long and well established history. The literature is mature, and several texts have been written describing image analysis through the statistical methods. Filtering methods based on local neighborhood statistics such as *median filtering* can be found throughout the literature. Image contrast enhancement techniques based on histogram equalization have also been explored and are in use in medical as well as other production environments [Pizer 1987]. Numerous

methods for performing segmentation and classification of images based on statistical pattern recognition are well documented in various texts [e.g., Duda 1974]. Statistically based relaxation filters founded on the theory of Markov processes (Markov Random Fields) [Geman 1984, Chellappa 1993, Jain 1985] as well as relaxation strategies based on expectation-maximization methods [Dempster 1980] also have a long history. Geiger and Yuille provide a framework for comparing these and other segmentation strategies, including nonlinear diffusion discussed later in this chapter, in their survey of the common threads shared by different algorithms [Geiger 1991].

Typically, statistical methods in image processing employ the histogram of the image or some other means of representing the probability density function of the intensity values. This representation is most often computed at the maximum outer scale of the image. That is, the histograms, mixture models, or probability distribution approximations are computed across the whole image, including all pixel values equally.

Image-wide probability density functions are commonly approximated as a Gaussian or linear combinations of multiple Gaussians. A maximum likelihood algorithm is usually then applied to classify individual pixel observations. Such methods seldom include local spatial trends or the geometry of the image as part of the statistical classifier. Maximum likelihood classifiers often employ image geometry in a post-process connectivity filter or, in the case of expectation-maximization methods, the classifier iterates between the maximum likelihood calculation and connectivity filtering.

Exceptions to the generalization that statistics are computed at the outer scale of the image include the contrast enhancement method of adaptive histogram equalization. Adaptive histogram equalization (or AHE) and its derivatives (Contrast limited AHE or CLAHE, and Sharpened AHE or SHAHE) construct local histograms of image intensity and compute new image values that generate an equalized local probability distribution. [Cromartie 1995]. Early algorithms for AHE included calculating histograms over non-overlapping rectangular neighborhoods and interpolation between equalized values. [Pizer 1987]. The choice of neighborhood operator was originally made on the basis of computational efficiency.

Other exceptions include Markov random fields and sigma filters. Markov random fields (MRFs) filters apply maximum likelihood estimators over a local neighborhood [Geman 1984]. Techniques using sigma filtering also compute nonlinear smoothing functions based on a local sampling window [Lee 1983]. Local statistics within a well defined neighborhood are computed and the central pixel value is adjusted according to some function of those statistics. Questions often arise over the priors used in sigma filters and smoothing based on Markov random fields. Other questions arise over the selection of the neighborhood function.

This chapter addresses the construction of robust statistics over a principled neighborhood function. The values that are proposed are local means and local central moments of intensity.

4.2. Images and Stochastic Processes

This work assumes a particular model for images. As with most statistical pattern recognition systems, this research is based on the assumption that the input signal follows a Gibbs distribution. Stated loosely, a Gibbs assumption states that the value for the intensity at a particular location has compact local support. This research restates these assumptions using the language of *stochastic processes* (defined below). Restating and further illuminating this common assumption requires the following background material.

4.2.1. Stochastic Processes

Chapter 2 defines an image to be a representation of some scene. The recording of the information within the scene is always subject to error of some kind (e.g., approximation error, measurement error, noise, discretization error, etc.) If the measurement of the scene is repeated, an identical image is not always acquired. However, there is usually a strong likelihood that the corresponding pixel values within two images of the same scene will have similar intensities.

The study of *stochastic processes* enables the quantification and analysis of the predictability of image measurement, the likelihood of obtaining similar images upon repeated acquisition. A more complete version of the following discussion can be found in Papoulis' introduction to random variables and stochastic processes [Papoulis 1991]. What is presented here is his organization modified from a time-based structure to one based on spatial location, transferring it to the framework of image processing.

A stochastic process $\underline{F}$ is a mapping of locations in space to random variables. Representations for $\underline{F}$ include the function notation of $\underline{F}(\mathbf{p},\xi)$ which represents the ξth sample of the random variable located at position $\mathbf{p}$. This notation is often abbreviated as $\underline{F}(\mathbf{p})$ when individual observations are not of interest, but rather the random variable itself. Formally,

> *Definition:* A *stochastic process* $\underline{F}$ (alternatively $\underline{F}(\mathbf{p},\xi)$ or $\underline{F}(\mathbf{p})$) is a continuous mapping $\underline{F} : \mathsf{R}^n \rightarrow \Xi$, where Ξ is a random variable. The domain of $\underline{F}$ is the set of all points $\mathbf{p}$ of an n-dimensional space: $\mathbf{p} \in \mathsf{R}^n$. The range of $\underline{F}$ is a random variable whose probability distribution function is
>
> $$F(x,\mathbf{p}) = P\{\underline{F}(\mathbf{p}) \quad x\}$$
>
> F is a function of the spatial variable $\mathbf{p}$, and it gives the probability of the event $\{\underline{F}(\mathbf{p}) \quad x\}$ consisting of all outcomes ξ such that, at the specific location $\mathbf{p}$, the samples $\underline{F}(\mathbf{p},\xi)$ of $\underline{F}$ do not exceed the value of x. The corresponding probability density function is $f(x,\mathbf{p})$ such that
>
> $$f(x,\mathbf{p}) = \frac{\partial F(x,\mathbf{p})}{\partial x}$$

The definition given above describes $\underline{F}(\mathbf{p},\xi)$ as a *continuous-space* process since the domain of $\underline{F}$ is continuous over R^n. If $\underline{F}$ is a mapping from the space of integers (i.e.,

$\mathbf{p} \in Z^n$), then $\underline{F}$ is a *discrete-space* process. If the values of $\underline{F}(\mathbf{p},\xi)$ are countable, then $\underline{F}$ is a *discrete-state* process; otherwise it is described as a *continuous-state* process.

4.2.2. Images as Samples

Paraphrasing Papoulis, $\underline{F}(\mathbf{p},\xi)$ has four interpretations:

1. $\underline{F}(\mathbf{p},\xi)$ is an ensemble of functions with $\mathbf{p}$ and ξ as variables.
2. It is a single function $\underline{F}(\mathbf{p}_0,\xi)$, where $\mathbf{p}_0$ is a constant, and ξ is allowed to vary. In this case $\underline{F}(\mathbf{p}_0,\xi)$ is called the *state* of the process at $\mathbf{p}_0$.
3. $\underline{F}(\mathbf{p},\xi_0)$ is a single function (or a *sample* of the given process) where ξ_0 is fixed, and $\mathbf{p}$ is allowed to vary.
4. If ξ_0 and $\mathbf{p}_0$ are constant, $\underline{F}(\mathbf{p}_0,\xi_0)$ is a scalar value.

Using the interpretation 3 above, the process of capturing the intensity values of a scene to form a digital image $I(\mathbf{p})$ can be considered to be a sample from a discrete-space discrete-state stochastic process $\underline{F}(\mathbf{p})$. This interpretation assigns $I(\mathbf{p}) = \underline{F}(\mathbf{p},\xi_0)$, for some ξ_0, as a family or ensemble of samples one from each pixel location $\mathbf{p}$.

This view of images is a natural one. Consider the acquisition of still images of a stationary scene using a video camera. If there is noise in the input signal, two images acquired at slightly different times $I_0(\mathbf{p}) = \underline{F}(\mathbf{p}, \xi_0)$, and $I_1(\mathbf{p}) = \underline{F}(\mathbf{p}, \xi_1)$, while not identical, would be subject to the same noise processes. Optical distortions, manifesting themselves as spatial functions, would exhibit themselves identically in each image. Color shifts, variable sensitivity of the detector grid, radio frequency noise and amplifier noise would not generate the same values on repeated sampling, but would follow the same behavior for each location $\mathbf{p}$.

Given an ensemble of images of the same scene $\{I_0(\mathbf{p}), I_1(\mathbf{p}), \ldots I_n(\mathbf{p})\}$, (equivalently a large set of samples $\{\underline{F}(\mathbf{p}, \xi_0), \underline{F}(\mathbf{p},\xi_1), \ldots , \underline{F}(\mathbf{p}, \xi_n)\}$ of process $\underline{F}$) where n is a large number, the expected, average or mean intensity value $M(I(\mathbf{p}))$ of pixel $\mathbf{p}$ can be estimated using the following calculation.

$$\mu_I(\mathbf{p}) = \mu_{\underline{F}}(\mathbf{p}) = \langle \underline{F}(\mathbf{p}) \rangle \approx \frac{1}{n}\sum_{i=1}^{n} I_i(\mathbf{p}) = \frac{1}{n}\sum_{i=1}^{n} \underline{F}(\mathbf{p},\xi_i) \qquad (4\text{-}1)$$

Note the natural association between the expected value or mean of the stochastic process and the mean of the sample set of images. The variance of the sample set of images $V(I(\mathbf{p}))$ can be calculated in a similar fashion, with a corresponding relationship to the variance of $\underline{F}$.

$$\mu_I^{(2)}(\mathbf{p}) = \mu_{\underline{F}}^{(2)}(\mathbf{p}) \approx \frac{1}{n}\sum_{i=1}^{n} \left(I_i(\mathbf{p}) - \mu_I(\mathbf{p})\right)^2 = \frac{1}{n}\sum_{i=1}^{n} \left(\underline{F}(\mathbf{p},\xi_i) - \mu_{\underline{F}}(\mathbf{p})\right)^2 \qquad (4\text{-}2)$$

The notation $\mu_{\underline{F}}^{(2)}(\mathbf{p})$ refers to the second central moment of $\underline{F}$ at location **p**. The order of the moment is indicated by the superscript value. The parenthesized superscript denotes that this is a central moment; that is, that this moment is calculated about the mean of $\underline{F}$ at **p**. A general form for central moments of $\underline{F}$ at **p** given an ensemble of samples is

$$\mu_{\underline{F}}^{(k)}(\mathbf{p}) = \left\langle \left(\underline{F}(\mathbf{p}) - \mu_{\underline{F}}(\mathbf{p})\right)^k \right\rangle = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{F}(\mathbf{p},\xi_i) - \mu_{\underline{F}}(\mathbf{p})\right)^k \tag{4-3}$$

4.2.3. Ergodicity

Stochastic processes are not functions, but mappings to random variables. When using real data it is not always convenient or possible to acquire sufficient samples of a single process $\underline{F}(\mathbf{p})$ to generate accurate information regarding the probability density function of the random variable for each location **p**. In many real examples in image processing, only one image is provided, not several. If $\underline{F}$ may be assumed to be stationary, that is the probability densities of the random elements of $\underline{F}$ are identical independent of **p**, it is possible to use these assumptions or properties to perform spatial averaging in place of averaging across many samples. The concept of *ergodicity* describes these conditions when a practitioner may trade spatial averaging for sample averaging.

Definition: Consider the stochastic process $\underline{F}(x)$ where $x \in \mathsf{R}^1$. A stochastic process $\underline{F}(x)$ is said to be *mean-ergodic* if for some fixed sample ξ_0 as $d \to \infty$ the following condition holds

$$\int_{-\infty}^{\infty} w(\tau;d)\underline{F}(x-\tau,\xi_0)d\tau \xrightarrow[d\to\infty]{} \mu_{\underline{F}}(x) \tag{4-4a}$$

$$w(x;d) = \begin{cases} 0 & \text{if} \quad x \le -\frac{d}{2} \\ \frac{1}{d} & \text{if} \quad -\frac{d}{2} < x \le \frac{d}{2} \\ 0 & \text{if} \quad \frac{d}{2} < x \end{cases} \tag{4-4b}$$

where the definition of the mean value $\mu_{\underline{F}}(x)$ is described in equation (4-1).

Notice that (4-4) is equivalent to a convolution of the function $\underline{F}(x,\xi_0)$ with a zero-centered square pulse function of height $1/d$ and width d centered at x.

The definition of mean-ergodicity as shown above can easily be generalized to processes of higher spatial dimensions. The concept of ergodicity can also be generalized from equation (4-4) to higher order central moments. For example,

Definition: A stochastic process $\underline{F}(x)$ (where $x \in \mathsf{R}^1$), is said to be *variance-ergodic* if for some fixed sample ξ_0 as $a \to \infty$ the following condition holds

$$\int_{-\infty}^{\infty} w(\tau;d)\left(\underline{F}(x-\tau,\xi_0) - \mu_{\underline{F}}(x)\right)^2 d\tau \xrightarrow[d\to\infty]{} \mu_{\underline{F}}^{(2)}(x) \tag{4-5a}$$

$$w(x;d) = \begin{cases} 0 & \text{if} \quad x \le -\frac{d}{2} \\ \frac{1}{d} & \text{if} \quad -\frac{d}{2} < x \le \frac{d}{2} \\ 0 & \text{if} \quad \frac{d}{2} < x \end{cases} \tag{4-5b}$$

where the definition of the second central moment $\mu_{\underline{F}}^{(2)}(x)$ is shown in equation (4-2).

Within the integral, the mean value term is relative to the position x, rather than to the index of integration τ. That is, the right value in the squared term of the integral is $\mu_{\underline{F}}(x)$ and not $\mu_{\underline{F}}(\tau)$.

Given these definitions of mean-ergodicity and variance-ergodicity, it can be shown that if a process $\underline{F}$ is variance-ergodic, it must also be mean-ergodic. The converse of this statement, however, is not true [Papoulis 1991].

Ergodicity may be generalized to even higher moments. If a process is ergodic in the strict sense, increasing the spatial measurement window about a pixel of a single sample of the process uniquely specifies the probability density function for the stochastic process. Further exploration of these ideas is beyond the scope of this dissertation.

If a process has a constant value for some observable moment of its distribution across space, it can be considered to be ergodic in a weak sense. That is, if the mean value of a process varies across space but the variance remains constant about that mean, then the process can be considered to be variance ergodic in the weak sense.

4.2.4. Ergodicity and Images

If an image, which is a representation of a wider scene, is considered to be a sample of a completely ergodic process, where "ergodic" is defined in the strict sense, then the scene itself is of little interest since its expected brightness is constant, essentially a grey field. The image portrays significant information about the noise in the acquisition process, but little other information.

What about images of scenes that have varying brightness and contrast? This section introduces *piecewise* or limited definitions of ergodicity. This distinction and its ramifications make the following definitions more applicable to image processing tasks.

Definition: A process $\underline{F}(x)$ (where $x \in R^1$), is *piecewise mean-ergodic* if it can be partitioned into intervals such that for each interval [a,b] where a x b:

$$\frac{1}{(b-a)} \int_a^b \underline{F}(\tau,\xi) d\tau = \mu_{\underline{F}}(x) + \varepsilon \quad \text{where} \quad \varepsilon \to 0 \text{ as } (b - a) \to \infty \tag{4-6}$$

Definition: A process $\underline{F}(x)$ (where $x \in R^1$), is *piecewise variance-ergodic* if it can be partitioned into intervals such that for each interval [a,b] where a x b:

$$\frac{1}{(b-a)} \int_a^b \left(\underline{F}(\tau,\xi) - \mu_{\underline{F}}(x)\right)^2 d\tau = \mu_{\underline{F}}^{(2)}(x) + \varepsilon \quad \text{where} \quad \varepsilon \to 0 \text{ as } (b - a) \to \infty \tag{4-7}$$

As in the previous definitions, definitions of higher order central moments and for higher dimensions may be inferred from these cases.

Given a single sample $I(x) = \underline{F}(x,\xi_0)$ of a piecewise ergodic process $\underline{F}$, it is not possible to recover either $\mu_{\underline{F}}(x)$ or $\mu^{(2)}_{\underline{F}}(x)$ completely since the partitioned intervals limit the averaging process. However, some reduction of the variance of the estimates of $\mu_{\underline{F}}(x)$ and $\mu^{(2)}_{\underline{F}}(x)$ may still be achieved through spatial averaging. If the boundaries of the partitions are known *a priori*, an optimal estimate of both $\mu_{\underline{F}}(x)$ and $\mu^{(2)}_{\underline{F}}(x)$ can be calculated from a given sample I(x).

If the boundaries of the partitions of $\underline{F}$ are not known, the problem of optimally estimating $\mu_{\underline{F}}(x)$ and $\mu^{(2)}_{\underline{F}}(x)$ from an image sample I(x) is underspecified. Without the size (or scale) of the intervals, $\mu_{\underline{F}}(x)$ may be estimated using equation (4-1) by varying the interval width |b-a| for each location x and selecting an interval size based on some criterion. A regularizing sampling kernel is required to handle these uncertain boundary positions and the randomness of $\underline{F}$. This regularization requirement is the basis for the research presented in this chapter, the development of multiscale techniques for estimating and evaluating the local probability densities in an image.

4.3 Multiscale Statistics

Without *a priori* knowledge of the boundaries and the object widths within an image, locally adaptive multiscale statistical measurements are required to analyze the probability distribution across an arbitrary region of an image. This section presents multiscale image statistics, a technique developed through this research for estimating central moments of the probability distribution of intensities at arbitrary locations within an image across a continuously varying range of scales. The piecewise ergodic nature of the image is an underlying assumption of these developments.

The definitions of mean and variance ergodicity in equations (4-4) and (4-5) imply the measurement of central moments in a local neighborhood of varying size about a point. Consider a set of observed values, $\tilde{I}(x) \subset R^1$, where for purposes of discussion the location $x \in R^1$, but can easily be generalized to R^n. The values of $\tilde{I}(x)$ may be sampled over a local neighborhood about a particular location x using a weighting function, $\omega(x)$, and the convolution operation, $\tilde{I}(x) \otimes \omega(x)$, where

$$\tilde{I}(x) \otimes \omega(x) = \int_{-\infty}^{\infty} \omega(\tau)\tilde{I}(x-\tau)d\tau = \int_{-\infty}^{\infty} \omega(x-\tau)\tilde{I}(\tau)d\tau \qquad (4\text{-}8)$$

A regularizing sampling kernel is desired. To avoid a preference in orientation or location, the sampling function should be invariant with respect to spatial translation and spatial rotation. As with all probability weighting functions it is essential that $\int_{-\infty}^{\infty} \omega(\tau)d\tau = 1$. One function that meets the above criterion is a normalized Gaussian function. Therefore, let

$$\omega(x) = G(\sigma, x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{4-9}$$

where the parameter σ represents the width of the sampling aperture.

4.3.1. Multiscale Mean

Let the scale space measurement comprised of a sum of the original image intensities weighted by a Gaussian sampling kernel be the average or expected value of $\tilde{I}(x)$ over the neighborhood defined by the aperture of size σ. This local mean is

$$\mu_{\tilde{I}}(x \mid \sigma) = \left\langle \tilde{I}(x); \sigma \right\rangle = \sum_{\tau = x}^{\text{neighborhood}(\sigma)} \omega(\tau)\, \tilde{I}(x - \tau) = \int_{-\infty}^{\infty} G(\sigma, x - \tau)\tilde{I}(\tau)\, d\tau \tag{4-10}$$

where $\left\langle \tilde{I}(x); \sigma \right\rangle$ is read as the expected value of $\tilde{I}(x)$ measured with aperture σ. This definition follows from the assumption that the observed values $\tilde{I}(x)$ represent a single sample from a mean-ergodic (or piecewise mean-ergodic) stochastic process.

The effect of a multiscale statistical operator can be viewed through its response to the input of a square pulse function. The resulting *pulse transfer function* is the output of a multiscale statistical operator acting upon a simple piecewise ergodic input signal. A point transfer function, the result of applying the multiscale statistical operator to a Dirac delta function input is not defined; statistics cannot be generalized from a single sample.

For the purposes of this discussion, the assumed input signal is P(d, x), a square pulse function centered at the origin with a spatial width of d and a height of 1/d (See Figure 4.1.). Note that $\lim_{d \to 0} P(d, x) = \delta(x)$.

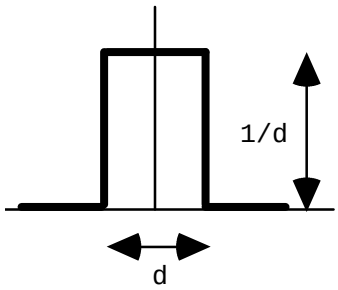


Figure 4.1a. 1D square pulse function P(d, x). Used as the input for generating pulse transfer functions.

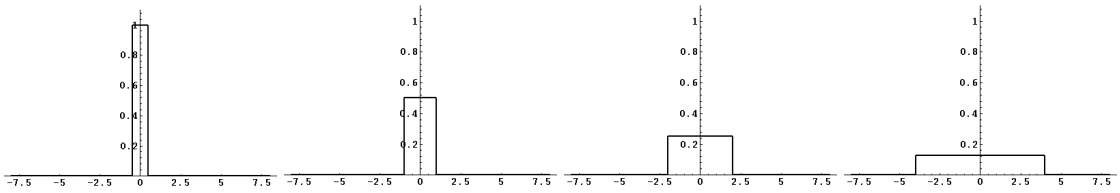


Figure 4.1b. 1D square pulse functions P(1, x), P(2, x), P(4, x), P(8, x). From left to right: d = 1, d = 2, d = 4, d = 8; $\lim_{d\to 0} P(d,x) = \delta(x)$.

The relationship between object width and the aperture of the multiscale statistical operator can be seen by applying the statistical mean operator at a variety of scales. Alternately, a statistical operator may be applied to square pulse inputs of various widths. Throughout this chapter the relationship between object and operator scale of the multiscale mean and higher order multiscale central moment operators will be presented by applying the operator to square pulse inputs of varying widths. An analysis of the relationship between object scale and operator aperture is found in Section 4.5.

The 1D pulse transfer function for the multiscale mean operation is described in the following equation and shown in Figure 4.2 for varying values of d.

$$\mu_{P(d,x)}(x|\sigma) = P(d,x) \otimes G(\sigma,x) = \frac{1}{2d}\,\mathrm{erf}\left(\frac{x+\frac{d}{2}}{\sigma\sqrt{2}}\right) - \frac{1}{2d}\,\mathrm{erf}\left(\frac{x-\frac{d}{2}}{\sigma\sqrt{2}}\right) \qquad (4\text{-}11)$$

erf(x) is the standard error function, $\mathrm{erf}(x) = \int_{-\infty}^{x} G(1,\tau)d\tau$. As the scale of the operator decreases relative to the size of the object or pulse, it provides a better approximation to the original input signal.

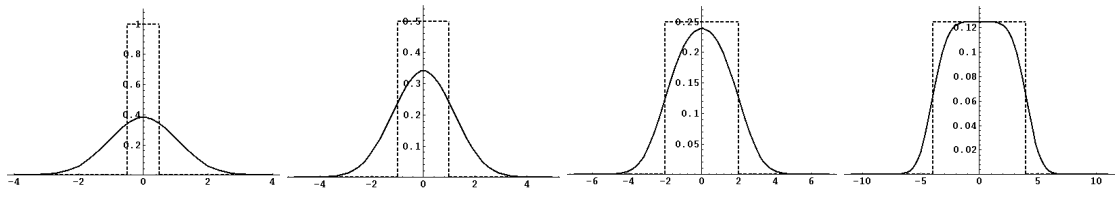


Figure 4.2. 1D Pulse transfer function for the multiscale mean operator $\mu_{P(d,x)}(x|\sigma)$ for σ = 1. From left to right: d = 1, d = 2, d = 4, d = 8. The dashed lines represent the input pulse function P(d,x). Note the difference in spatial and intensity ranges in each plot.

4.3.2. Multiscale Variance

It is straightforward to calculate a value for the local variance over the neighborhood specified by the scale parameter σ. Equation (4-12) describes the local variance of intensity about a point x at scale σ.

$$\begin{aligned}\mu_{\tilde{I}}^{(2)}(x\,|\,\sigma) &= \left\langle \left(\tilde{I}(x) - \mu_{\tilde{I}}(x\,|\,\sigma)\right)^2 ; \sigma \right\rangle \\ &= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau) - \mu_{\tilde{I}}(x\,|\,\sigma)\right)^2 d\tau \\ &= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^2 d\tau - \left(\mu_{\tilde{I}}(x\,|\,\sigma)\right)^2 \\ &= G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^2 - \left(\mu_{\tilde{I}}(x\,|\,\sigma)\right)^2\end{aligned} \tag{4-12}$$

The point transfer function of the local variance operator is not defined for the Dirac delta function $\delta(x)$ (i.e., $\mu_{\delta(x)}^{(2)}(x\,|\,\sigma)$ does not exist). However, the multiscale variance operation can be visually portrayed through its pulse transfer function $\mu_{P(d,x)}^{(2)}(x\,|\,\sigma)$. The multiscale variance of a pulse transfer function is

$$\begin{aligned}\mu_{P(d,x)}^{(2)}(x\,|\,\sigma) &= G(\sigma, x) \otimes (P(d,x))^2 - (G(\sigma, x) \otimes P(d,x))^2 \\ &= \left(\frac{1}{2d^2}\operatorname{erf}\left(\frac{x+\frac{d}{2}}{\sigma\sqrt{2}}\right) - \frac{1}{2d^2}\operatorname{erf}\left(\frac{x-\frac{d}{2}}{\sigma\sqrt{2}}\right)\right) - \left(\frac{1}{2d}\operatorname{erf}\left(\frac{x+\frac{d}{2}}{\sigma\sqrt{2}}\right) - \frac{1}{2d}\operatorname{erf}\left(\frac{x-\frac{d}{2}}{\sigma\sqrt{2}}\right)\right)^2\end{aligned} \tag{4-13}$$

Figure 4.3 shows the multiscale variance operator applied to a square pulse P(d, x) for varying values of d.

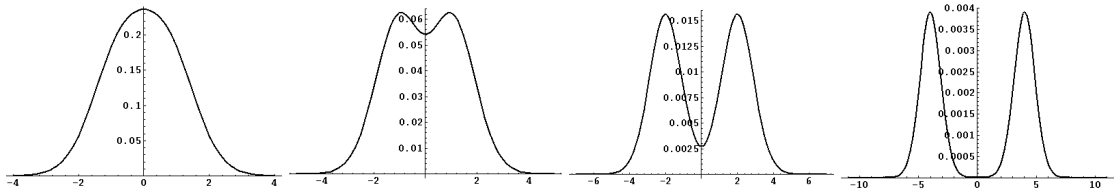


Figure 4.3. 1D Pulse transfer function for the multiscale variance operator $\mu_{P(d,x)}^{(2)}(x|\sigma)$ for σ = 1. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

The function shown in equation (4-13) is interesting in its resemblance to the square of the scale space gradient magnitude function (e.g., in the 1D case, $|\nabla P(d,x\,|\,\sigma)|^2 = (\frac{\partial}{\partial x} P(d,x\,|\,\sigma))^2 = (\frac{\partial}{\partial x} G(\sigma,x) \otimes P(d,x))^2$). Both are invariant with respect to rotation and translation, and both have similar responses to a given input stimulus. For example, Figure 4.4 portrays the variance calculation and the square of the scale-space 1D gradient magnitude of P(d,x).

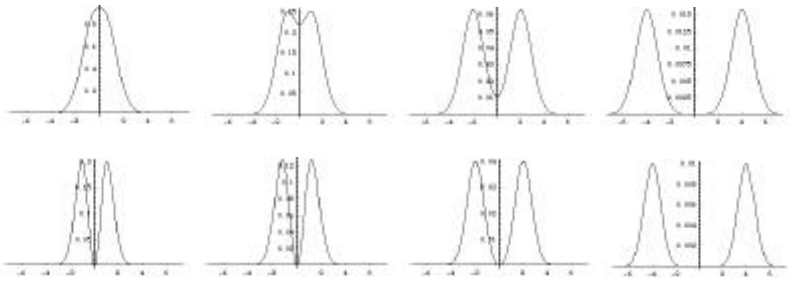


Figure 4.4. Comparison of the 1D Pulse transfer function for the multiscale variance operator $\mu^{(2)}_{P(d,x)}(x \mid \sigma)$ with $\sigma = 1$ to the square multiscale gradient magnitude operator. Top row, $\mu^{(2)}_{P(d,x)}(x \mid \sigma)$. Bottom row: $(\frac{\partial}{\partial x} P(d, x \mid \sigma))^2$. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

4.3.3. Multiscale Skewness and Kurtosis

The third and fourth local central moments are easily calculated in a similar fashion. The multiscale third central moment is

$$\begin{aligned}
\mu^{(3)}_{\tilde{I}}(x \mid \sigma) &= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau) - \mu_{\tilde{I}}(x \mid \sigma)\right)^3 d\tau \\
&= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^3 d\tau - 3\mu_{\tilde{I}}(x \mid \sigma)\int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^2 d\tau \\
&\quad + 3\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^2 \int_{-\infty}^{\infty} G(\sigma, x-\tau) I(\tau) d\tau - \left(\mu_{\tilde{I}}(x \mid \sigma)\right)^3 \\
&= G(\sigma, x) \otimes \left(\tilde{I}(\tau)\right)^3 - 3\mu_{\tilde{I}}(x \mid \sigma)\left(G(\sigma, x) \otimes \left(\tilde{I}(\tau)\right)^2\right) + 2\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^3 \\
&= G(\sigma, x) \otimes \left(\tilde{I}(\tau)\right)^3 - 3\mu_{\tilde{I}}(x \mid \sigma)\mu^{(2)}_{\tilde{I}}(x \mid \sigma) - \left(\mu_{\tilde{I}}(x \mid \sigma)\right)^3
\end{aligned} \tag{4-14}$$

The third central moment is demonstrated visually through its pulse transfer function across a range of pulse widths in Figure 4.5.

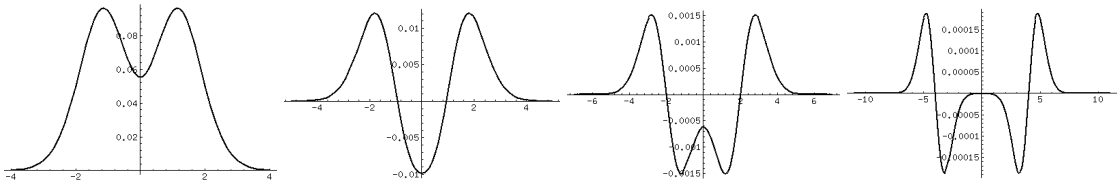


Figure 4.5. 1D Pulse transfer function of $\mu^{(3)}_{P(d,x)}(x \mid \sigma)$ with $\sigma = 1$. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

The response of the multiscale third central moment of a square pulse $\mu^{(3)}_{P(d,x)}(x \mid \sigma)$ is similar to the multiscale first derivative of a pulse stimulus. Although the magnitude of the responses of the two operations are often an order of magnitude apart, the

correspondence between the shapes of the two curves is remarkable. The two functions are compared in Figure 4.6.

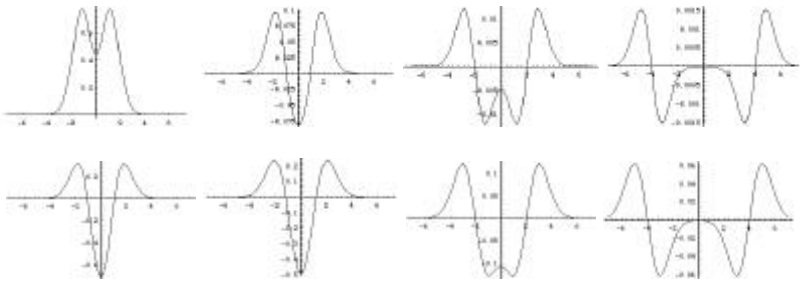


Figure 4.6. Comparison of $\mu^{(3)}_{P(d,x)}(x \mid \sigma)$ with σ = 1 to $\frac{\partial}{\partial x} P(d,x \mid \sigma)$ with σ = 1. Top row, $\mu^{(3)}_{P(d,x)}(x \mid \sigma)$. Bottom row: $\frac{\partial}{\partial x} P(d,x \mid \sigma)$. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

The multiscale fourth central moment is shown and simplified in equation (4-15).

$$
\begin{aligned}
\mu^{(4)}_{\tilde{I}}(x \mid \sigma) &= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau) - \mu_{\tilde{I}}(x \mid \sigma)\right)^4 d\tau \\
&= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^4 d\tau - 4\mu_{I}(x \mid \sigma)\int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^3 d\tau \\
&\quad + 6\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^2 \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau)\right)^2 d\tau - 4\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^3 \int_{-\infty}^{\infty} G(\sigma, x-\tau)\tilde{I}(\tau) d\tau \\
&\quad + \left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4 \int_{-\infty}^{\infty} G(\sigma, x-\tau) d\tau \\
&= G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^4 \\
&\quad - 4\mu_{I}(x \mid \sigma)\left(G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^3\right) + 12\left(\mu_{I}(x \mid \sigma)\right)^2 \left(G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^2\right) \\
&\quad + 16\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4 - 6\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^2 \left(G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^2\right) + 5\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4 \\
&= G(\sigma, x) \otimes \left(\tilde{I}(y)\right)^4 - 4\mu_{\tilde{I}}(x \mid \sigma)\mu^{(3)}_{\tilde{I}}(x \mid \sigma) \\
&\quad - 6\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^2 \left(G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^2\right) + 6\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4 \\
&\quad - \left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4 \\
&= G(\sigma, x) \otimes \left(\tilde{I}(x)\right)^4 - 4\mu_{\tilde{I}}(x \mid \sigma)\mu^{(3)}_{\tilde{I}}(x \mid \sigma) - 6\left(\mu_{\tilde{I}}(x \mid \sigma)\right)^2 \mu^{(2)}_{\tilde{I}}(x \mid \sigma) - \left(\mu_{\tilde{I}}(x \mid \sigma)\right)^4
\end{aligned}
\qquad (4\text{-}15)
$$

The fourth central moment is demonstrated visually through its pulse transfer function across a range of pulse widths in Figure 4.7.

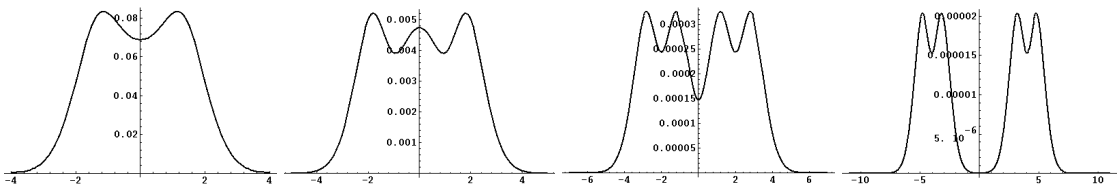


Figure 4.7. 1D Pulse transfer function of $\mu^{(4)}_{P(d,x)}(x \mid \sigma)$ with $\sigma = 1$. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

The function $\mu^{(4)}_{P(d,x)}(x \mid \sigma)$ has a response similar to the square of the scale-space curvature or second derivative measure of a pulse stimulus (e.g., in 1D the square of the multiscale curvature of a pulse P(d,x) is $(\frac{\partial^2}{\partial x^2} P(d, x \mid \sigma))^2 = (\frac{\partial^2}{\partial x^2} G(\sigma, x) \otimes P(d, x))^2$). At relatively large apertures the two curves take on similar properties. The two functions are compared in Figure 4.8.

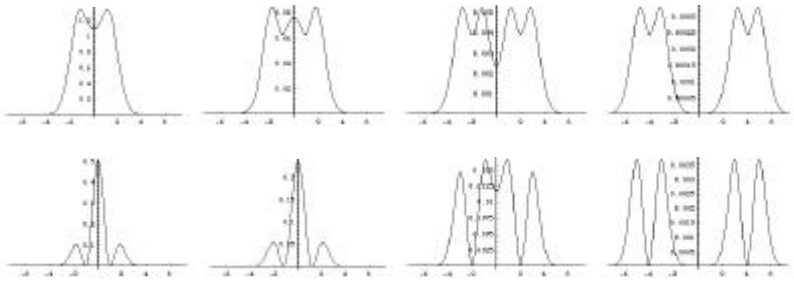


Figure 4.8. Comparison of $\mu^{(4)}_{P(d,x)}(x \mid \sigma)$ with $\sigma = 1$ to $(\frac{\partial^2}{\partial x^2} P(d, x \mid \sigma))^2$ with $\sigma = 1$. Top row, $\mu^{(4)}_{P(d,x)}(x \mid \sigma)$. Bottom row: $(\frac{\partial^2}{\partial x^2} P(d, x \mid \sigma))^2$. From left to right: d = 1, d = 2, d = 4, d = 8. Note the difference in spatial and intensity ranges in each plot.

4.3.4. Invariance with respect to linear functions of intensity

As specified before, the selection of the Gaussian distribution as the sampling kernel was motivated by a desire for the sampling filter to be invariant with respect to particular transformations of x. It may be desirable to analyze the sampled measurements of the array of $\tilde{I}(x)$ values in dimensionless units (i.e., invariant with respect to certain transformations of $\tilde{I}$). The dimensions of thethe third and fourth central moments shown above are subject to exponentiation by the order of the moment calculation. Dimensionless measurements may be obtained by normalizing the central moments with powers of the square root of v_0, the variance of the input noise (if known). The resulting measures are described as *skewness* and *kurtosis*. Their local manifestations, given a sampling aperture σ, are defined as

$$\text{Local Skewness: } g^{(3)}_{\tilde{I}}(x \mid \sigma) = \frac{\mu^{(3)}_{\tilde{I}}(x \mid \sigma)}{\left(\sqrt{v_0}\right)^3} \tag{4-16}$$

$$\text{Local Kurtosis: } g_{\tilde{I}}^{(4)}(x\,|\,\sigma) = \frac{\mu_{\tilde{I}}^{(4)}(x\,|\,\sigma)}{\left(\sqrt{v_0}\right)^4} \tag{4-17}$$

In the normalizations shown above, v_0 is used rather than the calculated second central moment $\mu_{\tilde{I}}^{(2)}(x\,|\,\sigma)$. In the case where the neighborhood about a pixel is contiguous and ergodic, $\mu_{\tilde{I}}^{(2)}(x\,|\,\sigma)$ can be used. However, under the piecewise ergodic assumption, discontinuities introduce bias into the value $\mu_{\tilde{I}}^{(2)}(x\,|\,\sigma)$, making it a poor estimate of v_0 where boundary conditions exist. This suggests a different form of multiscale statistical analysis to overcome this bias. Directional analysis methods that deemphasize the bias in multiscale central moment calculations introduced by local image geometry is the topic of Chapter 5.

4.4. Other Multiscale Central Moments

The general form for the multiscale central moment of order k of $\tilde{I}(x)$ is given by

$$\begin{aligned}\mu_{\tilde{I}}^{(k)}(x\,|\,\sigma) &= \left\langle \left(\tilde{I}(x) - \mu_{\tilde{I}}(x\,|\,\sigma)\right)^k ; \sigma \right\rangle \\ &= G(\sigma, x) \otimes \left(\tilde{I}(x) - \mu_{\tilde{I}}(x\,|\,\sigma)\right)^k \\ &= \int_{-\infty}^{\infty} G(\sigma, x-\tau)\left(\tilde{I}(\tau) - \mu_{\tilde{I}}(x\,|\,\sigma)\right)^k d\tau\end{aligned} \tag{4-18}$$

Although higher moments than the fourth central moment may be also of interest, the remainder of this discussion will address the nature of scale, noise and extensions of this concept of moments to multiple dimensions as well as to images containing multiple values per pixel.

4.5. Characteristics of Multiscale Image Statistics

It is important to recognize multiscale image statistics as central moments of the local *probability distribution* of intensity values taken from the neighborhood about a pixel location. Given the ensemble of all orders of these central moments, it is possible to reproduce the statistical behavior of the input signal and its noise properties at a particular location x in the image $\tilde{I}(x)$. These moments also capture some information of the local image geometry.

Multiscale image statistics may be illuminated by contrasting them with other image processing concepts. Such comparisons can lead to deeper insights into the nature of multiscale central moments of intensity.

4.5.1. Multiscale Statistics vs. Difference of Gaussian Operators

A cursory glance at the mathematical form for the k-th order multiscale central moment of intensity in equation (4-18) might falsely suggest that these moments are simply a form of contrast measurement by a difference of two Gaussian operators (DoG) raised to the k-

th power. There are some crucial differences between an exponentiated difference of Gaussian operation and the multiscale central moments described above.

In difference of Gaussian processing, an image is convolved with two Gaussian operators of differing aperture. The two filtered images are then subtracted to produce a resultant image that emphasizes boundary information within the original image. The process of filtering an image via a difference of Gausian operator and raising the result to the k-th power is formally described as follows:

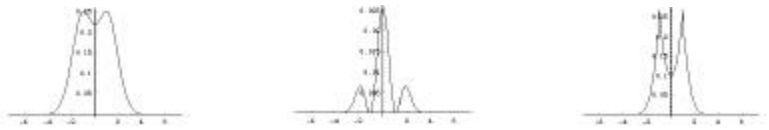
$$\begin{aligned}\left(\mathrm{DoG}\left(\tilde{I}(x);\sigma_a,\sigma_b\right)\right)^k &= \left(G(\sigma_b,x)\otimes\left(\tilde{I}(x)\right)-G(\sigma_a,x)\otimes\left(\tilde{I}(x)\right)\right)^k \\ &= \left(\int_{-\infty}^{\infty}G(\sigma_b,x-\tau)\left(\tilde{I}(\tau)\right)d\tau-\int_{-\infty}^{\infty}G(\sigma_a,x-\nu)\left(\tilde{I}(\nu)\right)d\nu\right)^k\end{aligned} \tag{4-19}$$

To simplify the comparison, equation (4-18) can be further simplified to the following expression.

$$\begin{aligned}\mu_{\tilde{I}}^{(k)}(x\mid\sigma) &= G(\sigma,x)\otimes\left(\tilde{I}(x)-\mu_{\tilde{I}}(x\mid\sigma)\right)^k \\ &= \int_{-\infty}^{\infty}G(\sigma,x-\tau)\left(\tilde{I}(\tau)-\int_{-\infty}^{\infty}G(\sigma,x-\nu)\tilde{I}(\nu)d\nu\right)^k d\tau\end{aligned} \tag{4-20}$$

Contrasting equation (4-19) and equation (4-20), their differences are immediately apparent. The DoG operation has two separate aperture parameters that govern its behavior where multiscale statistics use a single aperture. A more important distinction is the association of the exponential term. In a difference of Gaussian image raised to the k-th power, the difference of two filtered signals is exponentiated. In multiscale statistics the difference between the *original* input image and a filtered image is taken *before* being exponentiated and then filtered. Since convolution is a weighted summation process, exponentiated unsharp masking and multiscale statistics may be distinguished as follows: the exponentiated difference of Gaussian process is a power of the difference of two weighted sums. Multiscale central moments of intensity are weighted sums of an exponentiated difference. More simply, this is another example where the square of the sums does not equal the sum of the squares.

To illuminate the difference between these two forms of image measurement, a comparison between variance and the square of the DoG response to a pulse input is shown in Figuire 4.9. Two different aperture selections are shown for the DoG filter. These results demonstrate that the response of the DoG filter is sensitive to the selection of the aperture size parameters.



a. b. c.

Figure 4.9. Comparisons of $\mu^{(2)}_{P(d,x)}(x \mid \sigma)$ with DoG(P(d,x); σ_a, σ_b). The input function is a pulse P(d,x) with d = 1. a. $\mu^{(2)}_{P(d,x)}(x \mid \sigma)$ with σ = 1, b. DoG(P(d,x); σ_a, σ_b) with $\sigma_a = \sigma_b/\sqrt{2}$, $\sigma_b = 1$, and c. DoG(P(d,x); σ_a, σ_b) with $\sigma_a = 0$, $\sigma_b = 1$

4.5.2. Multiscale Moments of Intensity vs. Moment Invariants of Image Functions

The unmodified term "central moment" is ambiguous when taken in the context of image processing. There is a family of methods for image analysis describing image geometry that includes the concepts of moments and central moments. These measurements are distinct from the concept of statistics of local image intensities.

Hu introduced the family of moment invariants, taking advantage of the moment theorem that provides a bijection from derivatives in image space to moments in frequency [Hu 1962]. In 1D the calculation for computing the regular moment m_k moment of image continuously differentiable function $\tilde{I}(x)$ is shown in equation (4-21).

$$m_k = \int_{-\infty}^{\infty} \tau^k \tilde{I}(\tau) d\tau \tag{4-21}$$

To compute central moments, the spatial index of integration τ is offset to the image centroid calculated in 1D as (m_1/m_0). Central moments $m_{(k)}$ of the input image $\tilde{I}(x)$ are defined as

$$m_{(k)} = \int_{-\infty}^{\infty} (\tau - m_1/m_0)^k \tilde{I}(\tau) d\tau \tag{4-22}$$

It is possible to postulate the existence of $m_{(k)}(x|\sigma)$, a multiscale locally adaptive version of these moment invariants. Using a Gaussian as the neighborhood function and using a normalization consistent with the moment theorem, the formalization of multiscale locally adaptive moment invariants becomes

$$m_{(k)}(x \mid \sigma) = \int_{-\infty}^{\infty} e^{-\tau^2\sigma^2} (\tau - m_1/m_0)^k \tilde{I}(\tau) d\tau \tag{4-23}$$

From these basic equations it is clear that moment invariants and multiscale image statistics are very different. Moment invariants are applied in the spatial domain while image statistics are applied in the intensity domain. Moment invariants capture information about image geometry; the Taylor reconstruction of the infinite set of central moments of the image function yields the original image $\tilde{I}(x)$. Multiscale image statistics capture information about the histogram of pixel values within an image; the Taylor reconstruction of the infinite set of central moments of intensity generates the probability distribution function of $\tilde{I}(x)$.

4.6. Measurement Aperture, Object Scale, and Noise

How does the error associated with the additive noise propagate through multiscale image statistics? In particular, how does noise affect the calculation of the local variance or second central moment? What is the relationship between noise and image geometry?

Assume an image function with additive, zero-mean, Gaussian distributed, spatially uncorrelated, "white" noise $\tilde{I}(x) = (I(x) + \tilde{u}(x))$ where $\tilde{u}(x)$ is a random variable with zero mean, variance of v_0, and no spatial correlation. That is, $\tilde{u}(x) \sim N(0,v_0)$. Also, $\tilde{u}(x_0)$ and $\tilde{u}(x_1)$ are Gaussian distributed, zero-mean, independent, identically distributed random variables for all spatial coordinates $x_0 \neq x_1$. Let the scale-space representation of $\tilde{I}(x)$ where σ is the scale or measurement aperture be $\tilde{I}(x \mid \sigma) = G(\sigma, x) \otimes (I(x) + \tilde{u}(x))$.

Consider $M(\mu_{\tilde{I}}^{(2)}(x \mid \sigma)) = \left\langle \mu_{\tilde{I}}^{(2)}(x \mid \sigma) \right\rangle$, the mean of the local variance $\mu_{\tilde{I}}^{(2)}(x|\sigma)$. Applying the calculus of expected values to $\mu_{\tilde{I}}^{(2)}(x|\sigma)$ generates the following expression.

$$\begin{aligned}
M(\mu_{\tilde{I}}^{(2)}(x \mid \sigma)) &= \left\langle G(\sigma, x) \otimes (\tilde{I}(x))^2 - (G(\sigma, x) \otimes \tilde{I}(x))^2 \right\rangle \\
&= \left\langle G(\sigma, x) \otimes (\tilde{I}(x))^2 \right\rangle - \left\langle (G(\sigma, x) \otimes \tilde{I}(x))^2 \right\rangle \\
&= \left\langle G(\sigma, x) \otimes (I(x))^2 \right\rangle + 2\left\langle G(\sigma, x) \otimes I(x)\tilde{u}(x) \right\rangle \\
&\quad + \left\langle G(\sigma, x) \otimes (\tilde{u}(x))^2 \right\rangle - \left\langle (G(\sigma, x) \otimes I(x))^2 \right\rangle \\
&\quad - 2\left\langle (G(\sigma, x) \otimes I(x))(G(\sigma, x) \otimes \tilde{u}(x)) \right\rangle - \left\langle (G(\sigma, x) \otimes \tilde{u}(x))^2 \right\rangle \\
&= G(\sigma, x) \otimes (I(x))^2 + 0 + \left\langle (\tilde{u}(x))^2 \right\rangle - (G(\sigma, x) \otimes I(x))^2 - 0 - \tfrac{1}{2\sigma\sqrt{\pi}} \left\langle (\tilde{u}(x))^2 \right\rangle
\end{aligned} \tag{4-24}$$

Since the variance of $\tilde{u}(x)$ is defined to be v_0, equation (4-24) simplifies to the following expression.

$$M(\mu_{\tilde{I}}^{(2)}(x \mid \sigma)) = G(\sigma, x) \otimes (I(x))^2 - (G(\sigma, x) \otimes I(x))^2 + \left(1 - \tfrac{1}{2\sigma\sqrt{\pi}}\right) v_0 \tag{4-25}$$

4.6.1. Noise Propagation in Multiscale Statistics of an Ergodic Process

Increasing the aperture of the multiscale statistical measurement operator improves the measurement by decreasing the variance of the reported value through spatial averaging. This trend holds as long as discontinuities in the image are not encountered. In the absence of discontinuities, that is, with an image that is a sample of a complete ergodic process, the relationship between scale and variance can be studied.

Let I(x) be a constant function (i.e., let I(x) = c). Then $\tilde{I}(x)$ is ergodic. With a constant expected value across the image, multiscale statistics reflect the ergodic properties of the image as scale increases. In other words, there is a satisfying correspondence between the scale of the multiscale central moment of intensity operator and the measurement interval described in the definitions of ergodicity described in

equations (4-4) and (4-5). Specifically, it can be shown that multiscale statistics can be used to demonstrate mean-ergodicity using the following two relations.

$$\mu_{\tilde{I}}(x \mid \sigma) = G(\sigma, x) \otimes \tilde{I}(x) \xrightarrow[\sigma \to \infty]{} \mu_I(x) = c \tag{4-26}$$

Consider $V\left(\mu_{\tilde{I}}(x \mid \sigma)\right) = \left\langle \left(\mu_{\tilde{I}}(x \mid \sigma) - (G(\sigma, x) \otimes I(x))\right)^2 \right\rangle$, the variance of $\mu_{\tilde{I}}(x \mid \sigma)$.

$$V\left(\mu_{\tilde{I}}(x \mid \sigma)\right) = V\left(G(\sigma, x) \otimes \tilde{I}(x)\right) = \frac{1}{2\sigma\sqrt{\pi}} v_0 \xrightarrow[\sigma \to \infty]{} 0 \tag{4-27}$$

Equation (4-27) shows how $\mu_{\tilde{I}}(x \mid \sigma)$ converges to I(x) as a function of the initial variance v_0 and scale σ. The relationship in equation (4-27) is derived in Chapter 3.

Using multiscale statistics, it is also possible to show that $\tilde{I}(x)$ is variance-ergodic. Moreover, a closed form for the convergence of the local variance measure $\mu_{\tilde{I}}^{(2)}(x \mid \sigma)$ can be derived. Consider the expected value of the local variance given a constant function I(x).

$$\begin{aligned}
M\left(\mu_{\tilde{I}}^{(2)}(x \mid \sigma)\right) &= G(\sigma, x) \otimes (I(x))^2 - (G(\sigma, x) \otimes I(x))^2 + \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \\
&= G(\sigma, x) \otimes c^2 - (G(\sigma, x) \otimes c)^2 + \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \\
&= \int_{-\infty}^{\infty} G(\sigma, \tau) c^2 d\tau - \left(\int_{-\infty}^{\infty} G(\sigma, \tau) c \, d\tau\right)^2 + \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \\
&= c^2 \int_{-\infty}^{\infty} G(\sigma, \tau) d\tau - \left(c \int_{-\infty}^{\infty} G(\sigma, \tau) \, d\tau\right)^2 + \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \\
&= c^2 - c^2 + \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \\
&= \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0
\end{aligned} \tag{4-28}$$

Equation (4-28) implies that as scale increases the expected value of the multiscale variance approaches a constant value.

$$M\left(\mu_{\tilde{I}}^{(2)}(x \mid \sigma)\right) = \left(1 - \frac{1}{2\sigma\sqrt{\pi}}\right) v_0 \xrightarrow[\sigma \to \infty]{} v_0 \tag{4-29}$$

As scale decreases, $\mu_{\tilde{I}}^{(2)}(x \mid \sigma)$ becomes unstable. If $\sigma < \frac{1}{2\pi}$, $\mu_{\tilde{I}}^{(2)}(x \mid \sigma)$ is negative, an undesirable attribute for a measure of the variance of a random variable. However, in the context of discrete statistics, it is consistent with an attempt to compute central moments from small numbers of discrete samples. It is impossible to generalize statistics of a population from a single sample. Estimating statistics from a fraction of a sample can yield nonsensical negative values for variance and for all even order moments.

4.6.2. Noise Propagation in Multiscale Statistics of a Piecewise Ergodic Process

Most images are not ergodic in the strict sense; they contain discontinuities or boundaries denoting separate regions and objects within the image. If an image is a sample of a piecewise ergodic process, it is not possible to increase the aperture of a measurement operator to infinity without introducing bias from object boundaries. This section extends the previous discussion on the interaction between scale and noise to include boundary information.

Consider the simplest piecewise ergodic 1D image, a step function. Unlike the earlier pulse transfer function examples which were chosen to reflect the symmetry of the multiscale statistical operators, this example uses a single step. The mathematics are more easily presented and the effects of the discontinuity remain clear with this type of input function.

Let I(x) be a step function T(h, x), such that $\tilde{I}(x) = (T(h,x) + \tilde{u}(x))$ where

$$T(h,x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ h & \text{if} \quad x \geq 0 \end{cases} \tag{4-30}$$

The shape of this threshold step function is portrayed in Figure 4.10.

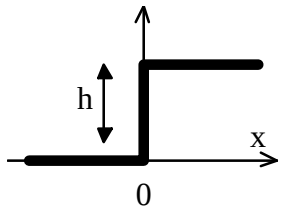


Figure 4.10. Test function T(h,x)

The following two relations follow.

$$\mu_{T(h,x)}(x \mid \sigma) = G(\sigma, x) \otimes T(h, x) = (h)\ \mathrm{erf}(\tfrac{x}{\sigma\sqrt{2}})$$
$$\text{and} \tag{4-31}$$
$$G(\sigma, x) \otimes (T(h, x))^2 = (h^2)\ \mathrm{erf}(\tfrac{x}{\sigma\sqrt{2}})$$

Inserting (4-31) into equation (4-25) yields a closed form for $M(\mu_I^{(2)}(x \mid \sigma))$ that is dependent on scale σ, height of the step function h, the initial variance v_0 of u(x), and proximity to the discontinuity expressed by $\mathrm{erf}(\tfrac{x}{\sigma\sqrt{2}})$.

$$M(\mu_I^{(2)}(x \mid \sigma)) = \underbrace{(h^2)\ \mathrm{erf}(\tfrac{x}{\sigma\sqrt{2}}) - \left((h)\ \mathrm{erf}(\tfrac{x}{\sigma\sqrt{2}})\right)^2}_{\text{relative proximity}} + \underbrace{\left(1 - \tfrac{1}{2\sigma\sqrt{\pi}}\right)v_0}_{\text{noise}} \tag{4-32}$$

There are essentially two components to the formal expression for $M(\mu_I^{(2)}(x \mid \sigma))$: (1) a term that measures the proximity of the boundary relative to the measurement aperture and (2) a term that reflects the estimate of the variance of the intensity distribution. These

two terms must be balanced. As scale σ increases, the relative distance to the boundary decreases, so to reduce the influence of the bias introduced by the nearby step edge, a small measurement aperture is desired. However, the noise term is not defined for very small scale σ. To achieve the best estimate of the variance of $\tilde{I}(x)$, a large aperture is desired. This tradeoff can be resolved if the position of the boundary is known *a priori*; a scale can be selected to deemphasize the bias from the step edge while estimating the variance of the local image intensities.

These results suggest a scale-space approach to measuring the variance within an image in the absence of *a priori* boundary information. For an arbitrary image $\tilde{I}(x)$, a continuously varying scale-space representation of $\mu_{\tilde{I}}^{(2)}(x \mid \sigma)$ is easily generated. In large contiguous regions, larger and larger measurement apertures may be used to measure the variance of the noise within the region. Near object boundaries, smaller apertures are required. When boundaries begin to affect the response of $\mu_{\tilde{I}}^{(2)}(x \mid \sigma)$ as scale increases, we can be sure that the value is being dominated by the relative proximity term. Thus by analyzing the variance of the multiscale mean operator $\frac{\partial}{\partial\sigma}(\mu_{\tilde{I}}^{(2)}(x \mid \sigma))$. through scale, minima in the multiscale variance response can be isolated.

4.7. Multiscale Statistics of 2D Images

Extending the construction of multiscale statistics to images of two dimensions is straightforward. For this work the central moments are constrained to be invariant with respect to rotation as well as translation. These constraints specify an isotropic Gaussian as the sampling kernel given by

$$G(\sigma, \mathbf{p}) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\mathbf{p}|^2}{2\sigma^2}} \tag{4-33}$$

where $\mathbf{p} = [p_x, p_y]$.

4.7.1. Multiscale 2D Image Mean

The multiscale 2D image mean is a local weighted average of image values. This measurement can be expressed as a convolution of the image with a 2D Gaussian kernel.

$$\begin{aligned} \mu_{\tilde{I}}(\mathbf{p}|\sigma) &= \left\langle \tilde{I}(\mathbf{p}); \sigma \right\rangle = G(\sigma, \mathbf{p}) \otimes \tilde{I}(\mathbf{p}) \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{((p_x-\tau)^2+(p_y-\nu)^2)}{2\sigma^2}} \tilde{I}([\tau, \nu])d\tau d\nu \end{aligned} \tag{4-34}$$

4.7.2. Multiscale 2D Image Variance

The multiscale 2D image variance measurement, like the multiscale 2D image mean, generalizes easily from the 1D case.

$$\begin{aligned}\mu_{\tilde{I}}^{(2)}(\mathbf{p}\mid\sigma) &= \left\langle \left(\tilde{I}(\mathbf{p}) - \mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^2;\sigma \right\rangle \\ &= G(\sigma,\mathbf{p}) \otimes \left(\tilde{I}(\mathbf{p}) - \mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^2 \\ &= G(\sigma,\mathbf{p}) \otimes \left(\tilde{I}(\mathbf{p})\right)^2 - \left(\mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^2\end{aligned} \tag{4-35}$$

4.7.3. Other Multiscale 2D Image Statistics

The general form for the k-th multiscale central moment for 2D images is

$$\begin{aligned}\mu_{\tilde{I}}^{(k)}(\mathbf{p}\mid\sigma) &= \left\langle \left(\tilde{I}(\mathbf{p}) - \mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^k;\sigma \right\rangle \\ &= G(\sigma,\mathbf{p}) \otimes \left(\tilde{I}(\mathbf{p}) - \mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^k \\ &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{((p_X-\tau)^2+(p_Y-\nu)^2)}{2\sigma^2}} \left(\tilde{I}([\tau,\nu]) - \mu_{\tilde{I}}(\mathbf{p}\mid\sigma)\right)^k d\tau d\nu\end{aligned} \tag{4-36}$$

4.7.4. Some 2D Examples of Multiscale Image Statistics

The equations of section 4.6 suggest that the height of the step edge found at a boundary as compared with the variance of the initial input signal is a critical element in image analysis. In the examples presented in this chapter, the term signal to noise ratio (SNR) will refer to the difference of the foreground intensity and the background intensity divided by the standard deviation of the additive spatially uncorrelated noise. The pixel is the atomic image element, so the additive noise and consequently the relative measurement of noise to signal is expressed as the SNR per pixel. Figure 4.11 shows a noisy computer generated image of a teardrop shape. The measured SNR per pixel within that image has been set to 4:1 on a raster resolution of 128 x 128 pixels.

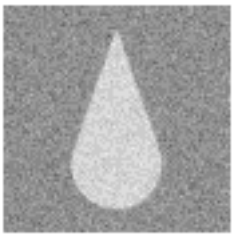
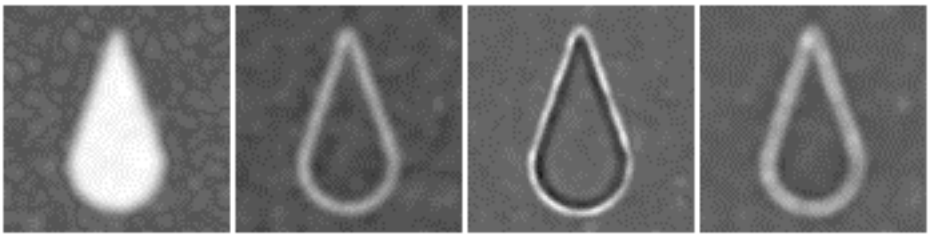


Figure 4.11. A 128×128 pixel Teardrop with Signal to Noise of 4:1

The images of Figure 4.12 are four local statistical measurements made of the teardrop using an aperture whose spatial aperture is 3 pixels wide. Figure 4.12a shows the local mean values. Figure 4.12b shows the measured local variances. Figure 4.12c shows local skewness. Figure 4.12d shows local kurtosis.



a. b. c. d.

Figure 4.12. Local statistical measures of the teardrop from Fig. 4.11.
(a: local means, b: local variance, c: local skewness, d: local kurtosis)

The important aspects of Figure 4.12 are that the multiscale statistical representations are significant with respect to image processing tasks. The local mean image is simply a multiscale measurement of the intensity values within the image. The variance image reflects edge strength and is analogous to the squared multiscale gradient magnitude of intensity. The local skewness measurement has a locus of zero crossings along the boundary of the teardrop shape. This behavior is similar to the response from applying the Laplacian of the Gaussian as a filter kernel to the same image.

4.8. An Application: Statistical Nonlinear Diffusion

A primary early focus of this research was automatic parameter selection for nonlinear diffusion functions. An extension beyond isotropic diffusion as an analytical tool is the notion of multiscale nonlinear or Variable Conductance Diffusion (VCD)[Whitaker 1993ab]. The work of Whitaker is preceded by the concepts of "anisotropic diffusion," an analytical process where images are often treated at pixel resolution [Perona 1990]. Other early work in segmentation by fixed-scale nonlinear diffusion has been pursued by Mumford and Shah [Shah 1991]. The fixed inherent scale of these early processes generate instabilities arising from noise with high spatial frequency. The multiscale properties of VCD allow it to perform smoothing operations in the presence of noise while preserving boundary information inherent in the image.

The control of VCD systems has often been difficult; the nature of the diffusion parameters are only partly understood. Seemingly insignificant changes in control constants can drastically modify the behavior of the image as the nonlinear diffusion process progresses. The VCD process as well as the properties of the noise contained within the image can be studied through statistical pattern recognition. Traditional statistical pattern recognition is performed at the maximum outer scale of the image, where histograms and clustering are analyzed across the entire range of observed pixels [Duda 1973]. As suggested in section 4.6, multiscale statistical analysis illuminates elements of the relationship between scale and object shape. By using critical values revealed in the analysis, images may be reconstructed at "natural" boundaries, accounting for the size and shapes of the objects within the image.

To achieve boundary preserving smoothing and nonlinear multiscale image analysis, Whitaker used the following diffusion equation with a boundary measuring conductance function.

$$\frac{\partial I(\mathbf{p},t)}{\partial t} = \nabla \bullet \left(g\left(\left| \nabla I(\mathbf{p},t);\sigma(t) \right| \right) \left(\nabla I(\mathbf{p},t);\sigma(t) \right) \right)$$

and (4-37)

$$g\left(\left| \nabla I(\mathbf{p},t);\sigma(t) \right| \right) = e^{-\frac{\left| \nabla I(\mathbf{p},t);\sigma(t) \right|^2}{k^2}}$$

The control parameter k is the conductance parameter and it remains constant. It controls the rate of the variable conductance given a perceived intensity gradient at aperture $\sigma(t)$. The aperture $G(\sigma(t),\mathbf{p})$ is a Gaussian sampling kernel whose scale or aperture, $\sigma(t)$ is monotonically decreasing with time t.

Conductance should be relative to the probability that the perceived intensity gradient is either noise or an object boundary [Perona 1990]. Boundariness is thus measured using the gradient magnitude normalized by the standard deviation of the noise process. In particular, Gerig suggested that the standard deviation be estimated from the whole image and used as the control parameter k used in the conductance equations shown in equation (4-37). In his application of VCD, Gerig uses a grid to sample patches, typically 8×8 pixels square, of an image, seeking the mean and variance of the sample population of 64 pixel intensities. The intensity histogram of the whole image is divided into subranges, and the variances of the grid squares whose means fall within the same subrange are compared. The minimum variance of comparable grid squares are then evaluated, and two variances, corresponding to noise in background and noise within objects, are extracted to be used as normalizing parameters in intensity gradient measurement [Gerig 1992].

While these methods generate useful measures of the noise process within an image, they are based on non-overlapping square partitions of the image. The scale of the sample grid is also fixed and not relative to the scale of the underlying objects. Through a multiscale approach, the scale of the variance and mean measurements can be made invariant with respect to the Cartesian coordinates of the image and invariant with respect to changes in object scale. Moreover, the largest scale reflecting a piecewise ergodic region may be used to provide the best measure of the probability distribution of noise within the region.

Work performed early in this research has yielded encouraging results [Yoo 1993, 1994]. Using the local statistical analysis presented in this chapter, a modification of the Whitaker VCD method has been developed. By analyzing the images with the local statistical operators of variance and means, a local measure of the SNR may be made in a spatially adaptive fashion. The sample aperture may be controlled by seeking scale values where changes in local variance are relatively insensitive to changes in aperture size. This process selects the largest scale where noise may be accurately estimated without the interference of intensity gradients introduced by boundaries. These arguments suggest replacement of $\sigma(t)$ with $\sigma(I(\mathbf{p},t))$ where

$$\sigma\left(I(\mathbf{p},t)\right) \Rightarrow \frac{\partial \mu_I^{(2)}(\mathbf{p},t \mid \sigma)}{\partial \sigma} = 0 \qquad (4\text{-}38)$$

This in effect allows the best estimate of both the mean intensity as well as the variance of the noise. In practice, a zero value is never achieved, so a small threshold is used to select values of σ that represent local minima of noise.

Given $\sigma(I(\mathbf{p},t))$, we can compute $\mu_I^{(2)}(\mathbf{p},t|\sigma(I(\mathbf{p},t)))$, the local variance at the optimal measurement scale. Gradient measurements are normalized with the expected noise distribution. That is, the conductance parameter k is replaced with the measured variance, making the conductance function

$$g(|\nabla I(\mathbf{p},t);\sigma(I(\mathbf{p},t))|) = e^{-\frac{|\nabla I(\mathbf{p},t);\sigma(I(\mathbf{p},t))|^2}{\mu_I^{(2)}(\mathbf{p},t|\sigma(I(\mathbf{p},t)))}} \quad (4\text{-}39)$$

Figure 4.13 shows a test object used to evaluate VCD systems. The object reflects features of differing scale, with the connecting bar between the two circles often dropping out during isotropic analysis.

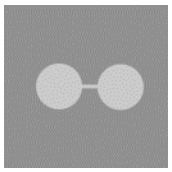


Figure 4.13. A test object. The figure contains structures at different scale. The raster resolution of the object is 128 x 128 pixels.

Figure 4.14 shows the results of the modified VCD system on the object shown in figure 4.13 with SNR of 1:1 on a 128 × 128 raster grid. Through modified VCD, no individual parameters were set for the image. Instead, measurement aperture was selected at each iteration for optimal estimation of mean and variance values.

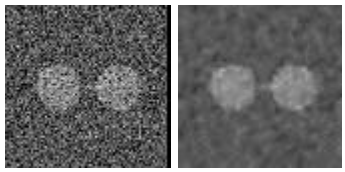


Figure 4.14. Results from the modified multiscale statistical approach to VCD (left: initial image, right: after 75 iterations of VCD)

A weakness of the current implementation is that it can get trapped in local minima as the multiscale variance measurement varies across scale. The patchwork appearance of the output image in Figure 4.14 is due in part to the algorithm getting stuck in a local minimum in scale rather than selecting a larger scale in the large contiguous regions of the background. However, the results are still encouraging since they represent automatic parameter selection for VCD.

Another weaknesses in this implementation foreshadows the developments of the next chapter. It is impossible to obtain a good estimate of the signal to noise ratio near object boundaries since the bias introduced by the geometry of the image dominates the multiscale statistical measurement. Chapter 5 introduces the acquisition of directional second moment values. Singular value decomposition generates the major and minor

axes, the directions in which the greatest and the least variance are measured given a scale σ. The resulting variance measured in the minor axis direction is likely to be a less biased measure of the local noise than the isotropic $\mu_I^{(2)}(\mathbf{p} \mid \sigma)$ statistic. Isophote curvature still introduces some bias, but an investigation of these properties is the topic of Chapter 5.

4.9. Multiscale Statistics of Multivalued Images

There are many sources of real data containing multiple values per pixel. The easiest means of manipulating these data is to invoke the language and tools associated with vector calculus, casting each pixel containing n elements as an n-element vector.

In many cases, the data values at each pixel are *incommensurable*. That is, they lack common basis for comparison. For instance, registered multimodal medical images containing both CT and MRI information, while extremely valuable, have little correspondence between the brightness and contrast at individual pixel locations. There is no basis for casting pixels as vectors since there is no useful distance metric to measure the norm of each vector. Even strongly related data are often not commensurable. Multiple echoes of MR images and even the separate color channels of color images are not commensurable.

However, statistical measurements of multiple incommensurable values per pixel can be made, and meaningful results can be drawn. Statistical methods enable the tracking of correlations among the multiple values, as well as detecting shifts or derivatives in these correlations. Moments and joint moments of image intensity distributions provide invariants that illuminate and normalize incommensurable values and make possible linear image processing methods with multivalued data.

4.9.1. The Multiscale Multivalued Mean

To extend the view of multiscale statistics presented in this chapter to images that have more than one value per pixel, first consider a two-valued image as an ordered pair of values at each location $\tilde{\mathbf{I}}(x) = \left(\tilde{I}_1(x), \tilde{I}_2(x)\right)$ where $x \in \mathsf{R}^1$. The following discussion may be generalized both to higher dimensions and to images of more than two values per pixel. For clarity and brevity, this presentation is limited to images of one dimension and two values per pixel.

The multiscale multivalued mean of $\tilde{\mathbf{I}}(x)$ is an ordered pair the multiscale means of the individual values.

$$\mu_{\tilde{\mathbf{I}}}(x|\sigma) = \left(\mu_{\tilde{I}_1}(x|\sigma), \mu_{\tilde{I}_2}(x|\sigma)\right) \tag{4-40}$$

4.9.2. Multiscale Multivalued Joint Moments

In the case of data containing multiple values the joint distribution must be considered. Consequently the distribution is uniquely defined by the set of its joint moments (see

Chapter 2). The general form for the measurement of multiscale joint moments is given by

$$\begin{aligned}\mu_{\tilde{\mathbf{I}}}^{(j,k)}(x\,|\,\sigma) &= \left\langle \left(\tilde{I}_1(x) - \mu_{\tilde{I}_1}(x\,|\,\sigma)\right)^j \left(\tilde{I}_2(x) - \mu_{\tilde{I}_2}(x\,|\,\sigma)\right)^k ; \sigma \right\rangle \\ &= G(\sigma, x) \otimes \left(\left(\tilde{I}_1(x) - \mu_{\tilde{I}_1}(x\,|\,\sigma)\right)^j \left(\tilde{I}_2(x) - \mu_{\tilde{I}_2}(x\,|\,\sigma)\right)^k \right)\end{aligned} \tag{4-41}$$

4.9.3. Multiscale Multivalued Variance

The second moment of a multivalued function deserves further illumination since it is so prominent in the literature and in image processing algorithms. Collectively, there are three joint moments that comprise the second moment or variance of a two valued image. They are $\mu_{\tilde{\mathbf{I}}}^{(2,0)}(x\,|\,\sigma) = \mu_{\tilde{I}_1}^{(2)}(x\,|\,\sigma)$, $\mu_{\tilde{\mathbf{I}}}^{(0,2)}(x\,|\,\sigma) = \mu_{\tilde{I}_2}^{(2)}(x\,|\,\sigma)$, and $\mu_{\tilde{\mathbf{I}}}^{(1,1)}(x\,|\,\sigma)$.

An alternate representation for the second moment of a multivalued function is to arrange these three individual moments in a symmetric matrix, called the *covariance matrix.*

$$\begin{bmatrix} \mu_{\tilde{\mathbf{I}}}^{(2,0)}(x\,|\,\sigma) & \mu_{\tilde{\mathbf{I}}}^{(1,1)}(x\,|\,\sigma) \\ \mu_{\tilde{\mathbf{I}}}^{(1,1)}(x\,|\,\sigma) & \mu_{\tilde{\mathbf{I}}}^{(0,2)}(x\,|\,\sigma) \end{bmatrix} \tag{4-42}$$

This matrix is invariant with respect to spatial rotation and translation of the image. This matrix is also invariant with respect to the addition of a constant intensity to either of the image values. It is not invariant with respect to multiplication of the intensities with a constant coefficient.

Another important invariant is the *correlation coefficient.* This value is calculated by

$$r_{\tilde{\mathbf{I}}}(x\,|\,\sigma) = \frac{\left(\mu_{\tilde{\mathbf{I}}}^{(1,1)}(x\,|\,\sigma)\right)}{\sqrt{\mu_{\tilde{\mathbf{I}}}^{(2,0)}(x\,|\,\sigma)\mu_{\tilde{\mathbf{I}}}^{(0,2)}(x\,|\,\sigma)}} \tag{4-43}$$

The correlation coefficient $r_{\tilde{\mathbf{I}}}(x\,|\,\sigma)$ is invariant with respect to spatial rotation and translation of the image, and it is also invariant with respect to linear functions of intensity applied to either of the image values. Note that $-1 \le r_{\tilde{\mathbf{I}}}(x\,|\,\sigma) \le 1$.

Finally, the covariance matrix may be normalized in a fashion similar to $r_{\tilde{\mathbf{I}}}(x\,|\,\sigma)$. If the entire matrix is divided by a scalar, the root of the product of the diagonal elements the matrix simplifies to

$$\begin{bmatrix} \sqrt{\dfrac{\mu_{\tilde{I}}^{(2,0)}(x\,|\,\sigma)}{\mu_{\tilde{I}}^{(0,2)}(x\,|\,\sigma)}} & r_{\tilde{I}}(x\,|\,\sigma) \\ r_{\tilde{I}}(x\,|\,\sigma) & \sqrt{\dfrac{\mu_{\tilde{I}}^{(0,2)}(x\,|\,\sigma)}{\mu_{\tilde{I}}^{(2,0)}(x\,|\,\sigma)}} \end{bmatrix} \tag{4-44}$$

This *normalized covariance matrix* exhibits the same invariances as the correlation coefficient. Its determinant is guaranteed to be between 0 and 1, inclusive.

4.9.4. Multiparameter VCD, a foreshadow of future work

Section 4.8 describes statistically controlled nonlinear diffusion for scalar images. The parameters controlling the nonlinear process are automatically selected using a scale-space analysis. The generalization of nonlinear diffusion to multivalued images is an important field of research today and is a likely application area for multiscale statistics.

Gerig and Whitaker both have generalized some forms of VCD to higher dimensions. Gerig has demonstrated vector-valued nonlinear diffusion on medical images with some success [Gerig 1992]. Whitaker has shown that invariants other than zeroth order intensities can be diffused; his resulting geometry limited diffusion has been able to generate ridge structures that describe the general form of objects within images [Whitaker 1993ab].

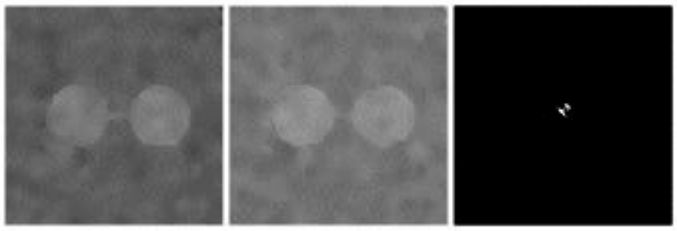
There are sources of data where the concept of vector valued diffusion does not apply because the different within-pixel values are incommensurable. An example of such data include environmental data such as levels of pollutants (measured in parts per million) and average wind velocity (measured in miles per hour). Demographic data also comprise a significant body of multivalued data that are not representable as vectors (e.g., census information reflecting local per capita income (in dollars per annum) and education level (in years of secondary and post secondary schooling). Registered multimodal medical datasets also represent an important type of data where the multiple values within a pixel do not have common units, origins, or noise properties. These types of data exhibit spatially dependent variables, but the values measured in any location are only loosely correlated. A pixel is therefore not a vector. However, the correlations among the values within a pixel may be of greater interest than the individual values themselves.

Early research related to this dissertation suggested how statistics can be applied to VCD in multivalued images, in particular those produced by medical magnetic resonance scanners. The technique where the idea was first developed had as its objective the incorporation of correlations of image intensities into VCD and was based on measurements of the dissimilarity of multivalued intensities through normalized distance measures in the multivalued histogram. Covariances among the multiple values within a pixel were calculated from user-supplied canonical pixels. The covariances were employed in the control of a nonlinear diffusion process. Specifically, in this process the user was asked to supply locations denoting sets pixel values characteristic of one or more desired image segments. Each individual set of pixels was used as a sample of the

population of image intensities within that particular segment. Central moments of each sample population were computed, and a multivalued probability distribution was modeled for each pixel set. Mahalanobis distances from individual pixel values to the surrounding probability distributions were then used as metrics of pixel similarity. The resulting method was applied at varying scale and performed reasonably well at reducing noise while preserving boundary information in multivalued magnetic resonance images [Yoo 1993].

Figure 4.15 shows the results of a user supervised VCD method to filter, segment and classify multivalued data. The individual data values exhibit significant noise, whose standard deviation is significantly greater than the difference between the mean background and mean foreground intensities of the objects. However, the values within each pixel exhibit a strong negative correlation. As such, that correlation can be employed in a similarity measure and used to diffuse the intensity values. As a result, the outcome is a clearly identifiable object of constant intensity on a constant intensity background. Figure 4.15a shows from left to right, both input image values and their corresponding scatterplot histogram. Figure 4.15b shows the image after processing with VCD, both values from left to right and the subsequent scatterplot histogram. Note the differences in the scatterplots. In the input histogram, there is little discernible separation between the two clusters, but the correlation is clear. In the histogram of Figure 4.15b, the clusters are easily seen, and their separation is possible with a linear discriminant with greatly reduced error.

a.



b.

Figure 4.15. Early work in statistically driven multivalued VCD. A synthetic multivalued image where the values are subject to significant Gaussian white noise and with a strong negative correlation between intensity values. (a) - original two valued input image and its scatterplot histogram. (b) - image after processing with VCD and resulting histogram.

The example in Figure 4.15 demonstrates that the information contained in the value of a pixel can be supplemented with information about the correlations among the values within an individual pixel. In the example above these correlations were obtained by asking a user to identify representative pixels, one set for the foreground objects and one

set for the background. Statistics were calculated of the pixel sets and generalized to the larger populations of background and foreground object pixels. It is this idea that I propose can be generalized to multiparameter images.

The next step is to use a scale-space analysis to adaptively select a measurement aperture for each pixel location. With an optimum aperture, correlations among the multiple values can be estimated and automatic nonlinear diffusion of multivalued images attempted.

4.10. Summary and Conclusions

This chapter has explored the derivation of isotropic multiscale image statistics. The stochastic structure of images has been discussed, and the model of an image as a 2D piecewise ergodic stochastic process has been adopted, upon which the mathematics and analysis of these multiscale image statistics has been founded.

A complete family of multiscale central moments has been introduced here with the properties that they (or their normalized relatives) are invariant under

(1) Rotation

(2) Translation

(3) Scale

(4) Linear Functions of Intensity

These moments were generalized to 2D images. Multiscale variance has been successfully applied in nonlinear smoothing algorithms using a scale-space analysis. Finally, the family of multiscale central moments was generalized to multivalued images.