STATISTICAL IMAGE PROCESSING

by
Paul Switzer
Stanford University

TECHNICAL REPORT NO. 15
JANUARY 1987

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
STATISTICAL IMAGE PROCESSING

by

Paul Switzer
Stanford University

TECHNICAL REPORT NO. 15
JANUARY 1987

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
STATISTICAL IMAGE PROCESSING

Paul Switzer
Stanford University

This is a personal overview of a subject whose literature is largely outside the domain of widely read statistical publications. For this statistician much of the literature is unsatisfying because of weak articulation of image processing problems and fragmentary discussions of the properties of image processing procedures. While in no sense a survey, it is hoped that this brief article, by means of example, will draw attention to strengths and weaknesses of some common approaches. It also suggests alternatives and makes proposals.

The article is divided into six sections which deal respectively with definitions of univariate and multivariate images, geometric and other properties of images, pixellation and observational data grids, probabilistic descriptions of images, local and global criteria of image estimation quality, and a classification of estimation procedures.
1. IMAGES

A sufficiently broad definition of an image would be a function $X(t)$ defined for a geographically meaningful index $t$ taking values in a bounded region $R$. We distinguish between an image, regarded as truth, and the data which are available to represent or estimate the image.

1.1. Geographic Support

Usually we think of an image as two-dimensional as in the case of photographic style imagery. In this case the location $t$ will be a vector with a horizontal and vertical coordinate. In some applications the image could be three-dimensional such as an ore body, or the evolution of a two-dimensional image over some time interval. It is also useful to consider single transects, paths, bore holes, etc., to be one-dimensional images, as well as the evolution in time at a single geographic point.

In this presentation the geographic support of an image will always be a compact region in a Euclidean space of one, two, or three dimensions, although further generalizations are possible. Even if the available data are on a discrete lattice, as is typical, it is important not to lose sight of the fact that a true continuous image exists apart from the observation lattice.

1.2. Types of Images

The type of image is characterized by the space of possible values of the image function $X(t)$. Some examples are:

(i) Dichotomous images: \{present, absent\}, \{0, 1\}, \{black, white\}

(ii) Non-negative images: \{intensity\}, \{concentration\}, \{grey level\}

(iii) Polychotomous images: \{colors\}, \{rock types\}, \{species\}

(iv) Multivariate combinations of the above: \{multi-band intensities\}, \{multi-element assays\}, \{hue & intensity\}

We commonly will want photographic-like representations of the images which have been estimated from available data. For two-dimensional grey-level or dichotomous images this representation is fairly obvious. For three-dimensional or multivariate images, representations
will commonly take the form of multiple two-dimensional grey-level images. Three-variate images could be represented by intensity-weighted mixtures of the three primary colors to produce false-color photographic imagery.

1.3. Image Size

Image size is an elusive concept. It could refer to the extent of the geographic domain which one wishes to consider as a whole for a particular application. Or it may be determined by physical limitations on display capabilities or by mapping conventions. Even when there are conventions it would seem reasonable to choose image sizes related to the spatial scales appropriate to applications, even when the full resolution of the physical display are not exploited. The effect of image size on estimation procedures is discussed briefly later.

2. IMAGE PROPERTIES

The image properties here discussed could be regarded as summary statistics reflecting global textural characteristics. A fairly general notion of a global property is that of a spatial frequency distribution function, \( F_Q \), of a locally defined pointwise property, \( Q(t) \). Specifically, for a real-valued property defined on images of unit area,

\[
F_Q(q) = \text{image area where } Q(t) < q,
\]

\[
= \text{Prob}[Q(t') < q] \text{ where } t' \text{ is a random location in the image domain } R.
\]

From the global distribution \( F_Q \) one can calculate average values, extrema, etc., of the local property \( Q(t) \). Generalizations are possible to joint global frequency distributions of two or more locally defined image properties. Global distributions of image properties clearly depend on the choice of image domain and subimages of \( R \) will differ from one another with regard to specific global properties. Different image estimating procedures will characteristically differ with respect to global properties of the image estimate.

2.1. Elementary Properties

The elementary global properties are those for which the local defining function \( Q(t) \) depends on the underlying image only through \( X(t) \) at the single point \( t \). Thus any reshuffling of the image will not change the values of elementary global properties. Some examples are:
(i) For univariate intensity images take \( Q(t) = X(t) \); then the global property \( FQ \) is just the frequency distribution of intensity values over the image. A variant would be a recentered and/or rescaled distribution, where the transformation may depend on other elementary global statistics. For bivariate (or multivariate) intensity images we could have bivariate frequency distributions of intensity values.

(ii) For bivariate (or multivariate) intensity images take \( Q(t) = X_1(t) \times X_2(t) \) where \( X_1(t) \) and \( X_2(t) \) are two standardized (zero mean, unit variance) components of the image vector at \( t \). The frequency distributions, \( FQ \) of such products of image components can then be averaged to get pairwise correlation coefficients between variables.

(iii) For dichotomous images \( Q_h(t) = 0 \) if \( X(t) = X(t + h) \), and \( Q_h(t) = 1 \) otherwise, where \( h \) is a lag vector. The global \( FQ_h \) is then essentially the relative frequency of same-color point pairs separated by lag \( h \). If \( h \) is varied then the plot of \( FQ_h \) against \( h \) is comparable to the variogram. For polychotomous images one could define a variogram for each color pair.

(iv) For differentiable intensity images let \( Qm(t) \) be the magnitude of the image gradient at location \( t \). The global distribution function, \( FQm \), contains rich textural information. For example, the variance of this global distribution, normalized by the image intensity variance, may be interpreted as a measure of image patchiness; it is related to the derivative at the origin of the isotropic variogram described above.

(v) For line-segment images \( Q_h(t) = 0 \) if no segment in the image crosses the segment \( (t, t+h) \), and \( Q_h(t) = 1 \) otherwise. The resulting global frequencies may once again be plotted as functions of \( h \); the local derivatives may be used to find segment-length distributions within directional class intervals.

3. DATA

We assume that the data are not definitive of the image \( X(t) \). Either the variable \( X \) is not directly observed, or it is not observed at all locations \( t \), or both. The set of data locations is denoted \( R' \) and is a subset of the image domain \( R \).
3.1. Pixels

Image data, typically obtained from scanning devices, are most often measurements recorded on a geographic rectangular grid. Each small rectangle in the grid is called a pixel (picture element). By convention, the pixel located at \( t \) means the pixel whose central point is \( t \). Pixels are a creation of the data reporting procedure rather than an intrinsic property of the underlying image.

Usually we think of the data grid as two-dimensional as in the case of photographic style imagery such as satellite imagery. In some applications the grid would be three-dimensional such as data from parallel sections of a rock or satellite imagery repeated at equal time intervals. It may also be useful to think of time series data or data recorded along geographic transects (lines, curves) as pixellations of a one-dimensional image.

Image data not obtained from scanning devices may not be geographically gridded, as is often the case with mineral assay data and other geologic field data. In such cases the data will not be associated with a regular lattice but they will still be spatially and/or temporally indexed by \( t \).

3.2. Measurements

The measurement recorded for the pixel at \( t \) will be denoted by \( Y(t) \). The types of possible measurements are very diverse. Furthermore, the space of possible values of \( Y(t) \) may or may not coincide with the space of possible values of \( X(t) \) for the underlying image. Some examples are given below:

(i) \( Y(t) \) is a multivariate intensity measurement, as in multi-band satellite data, while the underlying image \( X(t) \) is a dichotomous or polychotomous field corresponding to rock types, vegetation categories, land-use classes.

(ii) \( Y(t) \) and \( X(t) \) are both multivariate intensities, as in multi-band satellite data, where \( X(t) \) is interpreted as the true reflectance at \( t \) and \( Y(t) \) is the degraded and integrated reflectance as actually observed.

(iv) \( Y(t) \) is dichotomous, as in very low quality measurements of an intensity field \( X(t) \).

The measurement \( Y(t) \) for the pixel at location \( t \) will commonly be the result of some
integration over the $t$ domain, particularly for intensity data. The region of integration is not necessarily coincident with the pixel. There are several possibilities:

(i) The region of integration is small relative to the grid spacing. This is the common situation for field data such as monitoring data, mineral assay data, soil sampling, etc., using small sample volume.

(ii) The region of integration is comparable to the grid spacing. This is a common situation for satellite data where scanners are designed to provide this level of spatial integration. Sometimes, but not as often as supposed, the region of integration exactly corresponds to the rectangular pixel boundary. More commonly there is overlap across pixel boundaries.

(iii) The region of integration is large relative to the grid spacing. This is a situation typical of reflection seismology data.

3.3. Image Size and Data Sparseness

Image size is an elusive concept. For example, there are certain conventions in satellite data analysis which seem to prescribe square images which are $512 \times 512$ pixels. Such conventions, which may be related to scanner specifications, data storage and retrieval modes, or to capabilities of CRT display devices, are useful for image identification purposes. However, conventional image sizes may not be related to the sizes of image features which may be of interest.

The related question, whether the image data are sparse or dense, should not be gauged by image size conventions. Usually it is possible to make more or less ambitious demands on the data so that the same data may be viewed as either sparse or dense. Statistical processing of image data may use parameters reflecting both a local spatial scaling and a global spatial scaling related respectively to feature size and image size. It may be said that the most interesting problems for statistical approaches to image processing are those where the data are somewhat sparse rather than very dense.

3.4 Data Models

In the derivation of a formal Bayes image estimation procedure corresponding to a specified estimation error criterion, one combines a probability model for the class of possible images
(the ‘prior distribution’) with a model for the observable data \( Y(t) \) conditional on the actual image \( X(t) \).

For typical multi-band gridded satellite data, for example, a fairly general representation of such a conditional distribution is

\[
Y(t) = \int y[X(t + u)]dK(u) + N(t) \quad \text{for} \quad t \in R',
\]

where \( N(t) \) is a spatial lattice of unwanted noise vector random variables which model atmospheric effects, measurement error, illumination effects, surface texture variation, scan line effects, quantization, etc. The integrating kernel \( K(u) \) represents the spatial resolving capability of the measuring device and the effects of pixellation.

The transformation \( y(x) \) represents the conversion of the underlying object image property to a vector of light intensities corresponding to the frequency bands of the measuring device. For example, the object image may represent a patchwork of rock or vegetation species, i.e. a polychotomous image. Sometimes the object image is itself a multi-band intensity image so that the transformation \( y(x) \) is trivial. The degree to which the transformation \( y(x) \) is specified as well as the integrating kernel \( K(u) \) and the distribution of the noise field \( N(t) \) will depend on the physical situation as well as the formal requirements of the estimation procedure. For sparse data lattices the \( N(t) \) may sometimes be treated as mutually independent random variables. However, typically the \( N(t) \) would have some spatial continuity due to the persistence of atmospheric effects, for example. But the success of many estimation procedures depends on having an image, \( X(t) \), which changes more slowly spatially than the unwanted noise, \( N(t) \).

4. IMAGE MODELS

4.1. Image Models and Bayes Estimation

An image model is used to express global prior information about the true but unknown image. Models may be used to constrain the class of possible estimated images derived from available data. Or, via a Bayesian analysis, they may be used to assign posterior probabilities to image estimates, to find ‘most probable’ estimates which ‘minimize’ expected values of estimation error criteria, or more informally to guide estimation procedures.
For a proper Bayes estimation procedure, a unique probability measure needs to be associated with the set of all possible images on the domain of interest $R$. Since these prior measures describe global properties, it is reasonable to suppose that they could and should depend on $R$. In particular, if we have information specific to subregions of $R$, then we might use different models for each subregion, all of which are different from the model we would use for $R$ taken as a whole.

Image models, in the form of complete prior probability measures, are most usefully described constructively in ways which facilitate Monte Carlo image generation. Sample images, i.e., simulated images, are sometimes evaluated visually to assess the appropriateness of a selected image model. Of course, many very different image models may all provide visually satisfying simulations but have different implications for image estimation.

A parametric class of image models is often specified as a means of introducing flexibility into the choice of a prior measure; examples are described a little later. To select a unique measure for a Bayesian calculation, values of the parameters must be assigned. It might seem reasonable to infer parameter values from the available data so that the prior measure might somehow be adapted to the specific image, even though one no longer has a genuine Bayesian analysis.

It is desirable if parameters of an image model can be interpreted as tangible global statistics of the image. Examples are transition frequencies for specified spatial lag, class boundary perimeter, mean size of ‘objects’, class frequencies, mean gradient length and direction, etc. If such textural properties of the image are allowed to vary slowly over the image then additional parameters could describe this slow variation.

While the model-dependent optimality of a Bayes estimation procedure should not be taken too seriously, such procedures nevertheless do incorporate information which could establish, for example, the degree of localization of the data used for estimation. Further uses of image models such as calculations of average estimation errors can be quite speculative. However, it will always be interesting to use image models for evaluation of the kind of image properties described in Section 2.

We distinguish between models for the true image and models for the observable data,
given the true image. Both are needed for a Bayes estimation procedure. The latter have been mentioned briefly in Section 3; examples of the former are described below.

4.2. An Image Model Example for Univariate Intensity Images

The following example of an image model allows for slowly changing average intensities and intensity variability. The model is usefully written as the sum of two components:

\[ X(t) = m(t) + r(t) \quad \text{for} \quad t \in R, \quad \text{where} \]

- \( m(t) \) is the slowly changing smooth part of the intensity image,
- \( r(t) \) is a zero-mean non-stationary gaussian random function

\[ \text{Cov}[r(t), r(t')] = \sigma(t)\sigma(t')\rho(t - t') \quad \text{where} \]

- \( \sigma(t) \) is a slowly varying intensity standard deviation and
- \( \rho(h) \) is a stationary autocorrelation function.

The specified functions \( m, \sigma, \rho \) may each have adjustable parameters and these become the global parameters of the image model. The richness of the intensity model depends then on the richness of the \( m, \sigma, \rho \) specification. Of course, even for very parsimonious specifications, the class of model realizations is large because \( r(t) \) is a random process.

The gaussian property allows one to calculate some global textural measures. For example the marginal frequency distribution of intensity values will be a mixture of gaussian distributions. The mean of this mixture distribution is the spatial average value of \( m(t) \) and its variance is the spatial average of \( \sigma^2(t) \) plus the spatial variance of \( m(t) \). The average magnitude of intensity changes for lag \( h \) could also be calculated in a straightforward way as a function of the global parameters of the model.

4.3. Models for Dichotomous Images

A probability model for a dichotomous image is the same as a model of a random partition of the region \( R \) into two subregions, say black and white. Such partitions may be generated, for example, in one of the following ways:
(i) Poisson centers. Points are located at random within $R$; these serve as the centers of regions of influence; these regions of influence may have random or fixed sizes and shapes; the union of the regions of influence is the black subset of the partition and its complement is the white subset (1).

(ii) Cell structure. The region $R$ is first partitioned into cells, say by means of random lines or the Voronoi polygons of a point process; the cells are then independently colored black or white with either fixed probabilities or spatially varying probabilities. Cell structure models generalize in a straightforward manner to polychotomous images (2).

(iii) Dichotomization of continuous intensity images. The black subregion corresponds to the subregion where intensity exceeds a threshold.

Properties of dichotomous images may be related to parameters of the generating model. For example, the mean length of the boundary of the black subregion is proportional to

$$\frac{\partial p(h)}{\partial h} |_{h=0}$$

where $p(t, t') = \text{probability that the image is black at locations } t \text{ and } t'$

$$p(h) = \text{average } \{p(t, t')\} \text{ with } |t - t'| = h.$$  

Sometimes models are used which generate image values only on a discrete lattice. Often this lattice is made to correspond to the data observation lattice, but this can be both confusing and unnecessary. The underlying true image lattice should be a refinement of the data lattice, say with a spacing of one-half of the data lattice. Properties of lattice models for a given spacing do not typically carry over to other spacings so lattice models must usually be regarded only as models of convenience.

For dichotomous images on a finite region $R$, a lattice model corresponds to a discrete probability assignment to each of the finitely many possible images. Lattice models may be obtained by starting with a continuous space model and restricting it to the chosen lattice. Alternatively, one may have a direct specification such as the one described below:

Let $bb = \text{the proportion of all pairs of adjacent lattice points which are both black}$

$ww = \text{the proportion of all pairs of adjacent lattice points which are both white}$
Assign probability $G(bb, ww)$ to the image, where $G$ is a specified non-negative function on the unit square.

If $G$ is an increasing function for both arguments, then roughly images with less black-white boundary will be more probable than images with more black-white boundary. If we specify a model of this type for a given lattice then generally no such model would be implied for other lattice spacing. The above model is an example of a first-order spatial Markovian lattice process; see (3) and references therein.

5. ESTIMATION CRITERIA

Available data are used to obtain an estimate $\hat{X}(t)$ of the underlying image field $X(t)$ for all points $t$ in the spatial domain of interest. Goals of estimation are

(i) undo the effects of spatial integration of the measurement process

(ii) interpolate between measurement locations

(iii) filter noise components and other irrelevancies in the data.

Assuming the data are not definitive, we want estimating procedures which try to minimize meaningful error criteria. Here we consider mainly criteria which are local in that we wish to minimize a measure of error at each geographic point $t$ of the image domain. Of course, an estimation procedure which is ‘optimum’ for local estimation criteria may produce estimated images whose global properties do not match well the global properties of the true image; more on this point later.

5.1. Examples of Pointwise Error Criteria

Let $e(t)$ denote the measure of error at $t$; a more explicit notation is

$$e(t) = e(\hat{X}(t), X(t)) \quad \text{for} \quad t \in R.$$

It is important to note that an error can be defined at all points of the region $R$ and not just at the set of data points $R'$. A standard example for univariate intensity images is the squared error criterion $[X(t) - \hat{X}(t)]^2$. However, it is often the case that the absolute intensity at each location is less critical for interpretation than the intensity relative to some standard
scaling for the image as a whole. For example, we may agree that the average brightness and contrast of any estimated image will always be adjusted to standard levels. If errors are calculated from globally standardized true images and standardized estimates, then this could substantially affect the choice of estimation procedures.

The squared error criterion, even in its globally standardized form, is an example of a non-contextual criterion in that it pays no attention to neighboring points. By contrast, a contextual criterion is one which might do local standardization of the image and its estimate before calculation of the error $e(t)$. Hence, an intensity error of a given magnitude would be penalized more heavily in a locality of low variability and less heavily in a locality of high variability.

The pointwise errors, $e(t)$, will generate a frequency distribution as the location index $t$ ranges over the region of the image $R$. The mean of this error frequency distribution (or median, maximum, etc.) might serve as a measure of the overall quality of the restored image. This spatially averaged error we denote $e(R)$. It is clear that the spatially averaged error depends on the choice of averaging region $R$.

5.2. Comments on Optimality

Explicit image restoration quality criteria serve several useful ends such as

(i) judging whether the available data are adequate or bounteous

(ii) assisting the design of further data collection or data compression

(iii) permitting comparisons between competing estimating procedures.

Of course the direct calculation of errors is not possible except for experimental situations or simulations where the true image is known. Therefore, errors must themselves be estimated. Such estimated errors will necessarily involve some form of data modelling, for example a probabilistic description of how the observable data depend on the underlying image.

Probabilistic models which relate data to the underlying image might also be used to obtain estimation procedures wich ‘optimize’ restored image quality according to selected error criteria. However, the optimality property is not only tied to the selected error criteria and the adopted data model, but is also tied to a particular image domain: restoration algorithms
which try to minimize the spatially averaged error \( e(R) \) may not be the same as those which try to minimize \( e(R^\sim) \) where \( R^\sim \) is some subregion of \( R \), say. Basically, the statistical properties descriptive of the region \( R \) may be very different from those which are descriptive of subregions.

Estimation procedures which optimize a specified error criterion may necessarily produce estimated images which have properties quite different from the underlying true image or image probability model. As an extreme example consider the case of a polychotomous image with uninformative data; then the most probable estimate is the image with the highest prior probability which might well be a trivial single-color image. A more typical phenomenon is the reduction in local contrast when trying to optimize estimates of pointwise intensity, especially seen as a blurring of edges. One may wish to introduce side conditions to the optimization of the image estimate which try to preserve important local and global properties of the type described in Section 2, for example.

6. IMAGE ESTIMATION PROCEDURES

Image estimation procedures may be roughly classified into three types which are described more fully later:

(i) Bayes estimates;

(ii) estimates which minimize an empirical function of the data;

(iii) ad hoc estimates which have certain desirable properties.

In ‘easy’ problems many competing estimation procedures will perform similarly so comparisons should be made on challenging problems. This does not only mean low nominal signal-to-noise ratios since the effects of high noise levels are quite easily mitigated if the image signal is changing slowly relative to the noise. A challenging problem is one, for example, in which the noise and signal power spectra are not grossly different or in which the noise is correlated with the signal.

6.1. Bayes Estimates

Bayes procedures make use of a prior probability measure on the class of all possible images, a conditional probability specification of the observable data for each possible under-
lying image, and an explicit estimation loss criterion. The estimated image is the one which minimizes the expected value of the loss criterion; expectation is taken with respect to the calculated posterior distribution on the class of all possible images, given the data. In principle, a complete probability specification is needed for true Bayesian inference. This principle is often compromised by allowing for a parametric specification with parameter values somehow extracted from the data themselves.

The estimated image with maximum posterior probability or probability density corresponds to the use of an all-or-nothing error criterion. Another criterion is the sum of pointwise errors which requires the calculation of the posterior marginal probability distribution at each point of the image domain. Since a Bayes procedure depends so heavily on a complete probability specification it becomes important to examine the robustness of Bayes estimates to alternative specifications. There is no particular reason why a Bayes estimated image should have the same statistical properties as the true underlying image, so a Bayes estimate might be visually misleading even if the probability specification is 'correct'. Examples of prior probability specifications are described briefly in Section 4 for a dichotomous image and for an intensity image.

6.2. Image Estimates Which Minimize Empirical Functions of the Data

The notion that the underlying image to be estimated, $X(t)$, is spatially less changeable than the observable data $Y(t)$ leads to various estimation criteria based on regularization ideas; examples are smoothing splines (4), maximum entropy estimates (5), and principal component projections (6). The first two have been primarily developed for univariate images while the last of these is for multivariate images. As in the case of Bayes procedures, interesting properties of the underlying true image, such as those described in Section 2 may be systematically distorted by the estimate.

In the case of splines the idea is to minimize average curvature of the intensity field for reasons good or bad. If the data are error-free observations of the field on a discrete lattice then the spline estimate can be used to interpolate the field. When the observations are subject to error then allowing for the data points themselves to be reestimated can further reduce the curvature of the estimated intensity image. Smoothing splines are a compromise between
image curvature and fidelity to the observation points.

In the case of maximum entropy estimates of an image, it is only the data lattice itself which is reestimated and the procedure is not well adapted for interpolation. A compromise is made between fidelity to the observation points and minimization of the entropy of the marginal intensity frequency distribution. Unlike the spine curvature criterion, there is nothing spatial about the entropy criterion so it will not perform efficiently where the spacing of the observation grid is moderately close relative to changes in the intensity of the image. Indeed, if the data were permuted then the resulting estimation algorithm would be unaffected.

For multivariate gridded data such as multi-band satellite observations, a common estimation procedure to recover a noise-contaminated underlying multivariate image is to use principal components in the following manner. The modulation or overall spatial variation of the underlying true image is assumed to be much greater than the added noise so that the directions in space of variables corresponding to minimum variation are identified as noise. Hence, the projection of the multivariate data on the subspace orthogonal to the low-variance subspace is thought to suppress the noise while sacrificing some dimensionality.

The problem with the principal components procedure described above is the assumption that small global variability corresponds to noise. This criterion is non-spatial as it does not pay attention to spatial integrity and is invariant to spatial permutations of the data. Instead one may use a criterion explicitly related to spatial structure such as unit lag autocorrelation. Then one looks for directions in the space of variables which have minimum autocorrelation, rather than minimum variance, and identifies those as noise (7).

6.3. Ad Hoc Image Estimation Procedures

These procedures do not necessarily differ from those described above nor are they necessarily less effective in certain classes of problems. Rather ad hoc image estimates are distinguished in that they are not derived explicitly as solutions to optimality problems. Instead they are perceived to have desirable properties when applied to selected examples.

Certain ad hoc procedures are motivate by apparent inadequacies of 'optimal' procedures. For example, pointwise optimal procedures will not accurately reproduce derivative behavior
of an intensity image. In the case of images with sharp gradients, optimal linear filters will induce blurring. To remedy such effects one might consider median-type filters (8) or similarity type filters (9).

A general description of a median type filter begins with the assigning of weights to observations at locations \( t' \) in relation to estimation at location \( t \). Denote these weights by \( w(t', t) \). A cumulative frequency distribution of the observations is formed with jumps proportional to the assigned weights. The usual linear filter would take the mean of this distribution as the estimated intensity at fixed location \( t \); instead we take the median of this distribution to be the estimated intensity. The assignment of weights will reflect proximity to \( t \) so for each estimating location a new set of weights is needed, for example a moving window with fixed weights inside the window and zero weights outside the window.

A similarity type filter allows the weights \( w(t'; t) \) to depend on the similarity of the observations at \( t' \) and \( t \) assuming the image is being estimated only at data grid points. If \( Y(t') \) and \( Y(t) \) are the respective observed values, then \( w(t'; t) \) would be taken to be a nonincreasing function of the difference \( |Y(t) - Y(t')| \). In this way, if there is a steep intensity gradient such as a fault line in the vicinity of the estimation point \( t \), then little weight will be attached to data values on the far side of the fault.

7. REFERENCES


