BAYESIAN STATISTICS 5

Proceedings of the Fifth Valencia International Meeting

June 5–9, 1994

Edited by

J. M. Bernardo
J. O. Berger
A. P. Dawid
and
A. F. M. Smith

CLARENDON PRESS · OXFORD
1996

BAYESIAN STATISTICS 5, pp. 801–808
J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.)

Models for Shape Deformation

A. G. WILSON and V. E. JOHNSON
Duke University, USA

SUMMARY

Much work in Bayesian image analysis has focused on incorporating vague prior knowledge about an image into its analysis and on the calculation of appropriate estimates of the resulting posterior distribution. However, in the field of medical imaging, there is a need to incorporate specific prior information into many image analysis tasks. This paper discusses various models for shape deformation and proposes a model that accounts for features at multiple spatial resolutions.

Keywords: SCALE-SPACE; PRIOR DISTRIBUTION; HIERARCHICAL MODEL; DEFORMABLE TEMPLATE.

1. INTRODUCTION

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

In the field of medical imaging, specific prior information is often available. When a diagnostic image is created, there is knowledge about what part of the body has been imaged and about why the image has been made. Additionally, a template is often available in the form of a medical atlas, or as a cross-correlated high-resolution magnetic resonance (MR) or computed tomography (CT) image of the same patient. The question that we address here is how to incorporate this prior information into image analysis.

Space limitations preclude a thorough discussion of common Bayesian image analysis models, but the interested reader might refer to Geman and Geman (1984) and Besag (1974, 1986) for a discussion of Markov random fields, Bookstein (1986) and Goodall (1991) for landmark methods, Grenander and Miller (1994) for deformable templates, and Staib and Duncan (1992) for Fourier methods.

2. SCALE-SPACE METHODS

It has been recognized for about 25 years that multi-scale approaches can improve image analysis. Various multi-scale representations have been proposed, including pyramids (Burt, 1981; Crowley, 1981), wavelets (Mallat, 1989), and multi-grid methods (Hackbusch, 1985). One of the most promising representations, *scale space*, has been developed recently in the computer vision literature (Lindeberg, 1994).

Scale space was developed to model the processing at the front end of the visual system. Its basic premise is that the visual system must initially be able to handle image structure at all scales and resolutions (Lindeberg, 1994). This is accomplished by embedding an image in a one-parameter family of derived images, with resolution or "scale" as the parameter. This representation is not "data-reducing" like a wavelet decomposition, but is a highly redundant representation of the image that allows efficient calculations using features at multiple scales.

At first, the task of creating a "multi-scale" representation of image data seems somewhat arbitrary. Considered carefully, however, it is clear that when constructing a scale space representation it is critically important that the transformation from fine scale to coarse scale actually represent a simplification of the data, so that fine-scale features vanish monotonically with increasing scale. If new artificial structures could appear at coarse scales, it would be impossible to determine whether these structures arose from finer scale features or by accident—for example, by the amplification of noise.

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

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

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

$$L(x, \sigma) = f(x) * G(x; 0, \sigma).$$

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

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

If causality is combined with a prohibition of space-variant blurring, it can be shown that the derived family of images must satisfy the diffusion equation, with the initial condition that the derived image at scale zero is the initial image. This happens when the initial image is convolved with the Gaussian density function or one of its derivatives. Scale is taken to be the standard deviation of the Gaussian. If the image is thought of as an initial heat distribution, the scale space shows the heat distribution over time as diffusion occurs in a homogeneous medium.

Figure 1 shows a one-dimensional signal and slices from its scale space at increasing scales. Notice that small-scale features are suppressed as σ increases.

Koenderink (1984) notes that the prohibition of space-variant blurring is made primarily for computational convenience. If space-variant blurring is allowed, scale space methods known as *anisotropic diffusion* result (Perona and Malik, 1990). It should also be noted that *morphological* scale spaces (Jackway, 1992, 1993; van den Boomgaard, 1992) share many of the desirable properties of Gaussian scale spaces and additionally have the property that the number of local extrema is monotonically decreasing in scale, even in n dimensions.

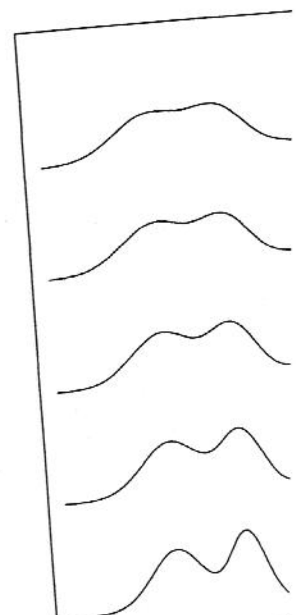


Figure 1. *Sample*

Intuitively, in Gaussian scale spaces, the sca
spatial dimensions. Eberly (1994) develops a me
He assumes that measurements in scale space sh
invariance, and zoom invariance, where *zoom* i
equivalently, a uniform magnification in the scale
the metric for scale space is

$$ds^2 = \frac{dx \cdot d}{\sigma^2}$$

where s denotes arc length, x denotes spatial c
With this metric, scale space has hyperbolic ge
Shape is often defined as those characteri
figure is translated, rotated, or uniformly sca
zoom invariance is particularly useful, becaus
when the figure changes scale.

3. MULTI-SCALE

This section develops a class of prior distribu
from a template into image analysis. Many of
discussed in the previous section. The overal
the space of images that assigns high proba
probability to images that are unlike a templ
prior probability distribution over the space
Notice that this is the form of a Gibbs dist
formulated in terms of cliques and potential

wavelet decomposition, but is a highly redundant
ent calculations using features at multiple scales.
ıle" representation of image data seems somewhat
s clear that when constructing a scale space repre-
ısformation from fine scale to coarse scale actually
t fine-scale features vanish monotonically with in-
ld appear at coarse scales, it would be impossible to
n finer scale features or by accident—for example,

ale space for continuous one-dimensional signals.
le space image $L : R \times R^{+} \rightarrow R$ is defined so that
inal signal,

$$0) = f(x)$$

:n by the convolution of the signal with the Gaussian
) and standard deviation σ:

$$f(x) * G(x; 0, \sigma).$$

ı-crossings of the second derivative of the Gaussian
eases monotonically with scale. Yuille and Poggio
ıperator that commutes with the diffusion equation;
' holds for derivatives of arbitrary order. Since the
zero-crossings of its first derivative, the number of
scale.
:here is no non-trivial linear shift-invariant (convo-
ıl extrema (Lifshitz and Pizer, 1990). To generalize
ension, a different "monotonically decreasing fea-
derives a multi-dimensional Gaussian scale space
:ant to capture the idea that "any feature at a coarse
(not necessarily unique) 'cause' at a fine level of
: true." As developed by Koenderink, the causality
not created as the scale parameter is increased, i.e.,
not "pop up out of nowhere" (Lindeberg, 1994).
tion of space-variant blurring, it can be shown that
ıe diffusion equation, with the initial condition that
al image. This happens when the initial image is
on or one of its derivatives. Scale is taken to be the
nage is thought of as an initial heat distribution, the
time as diffusion occurs in a homogeneous medium.
ıl and slices from its scale space at increasing scales.
sed as σ increases.
ıition of space-variant blurring is made primarily for
t blurring is allowed, scale space methods known as
ik, 1990). It should also be noted that *morphological*
:n Boomgaard, 1992) share many of the desirable
litionally have the property that the number of local
le, even in n dimensions.

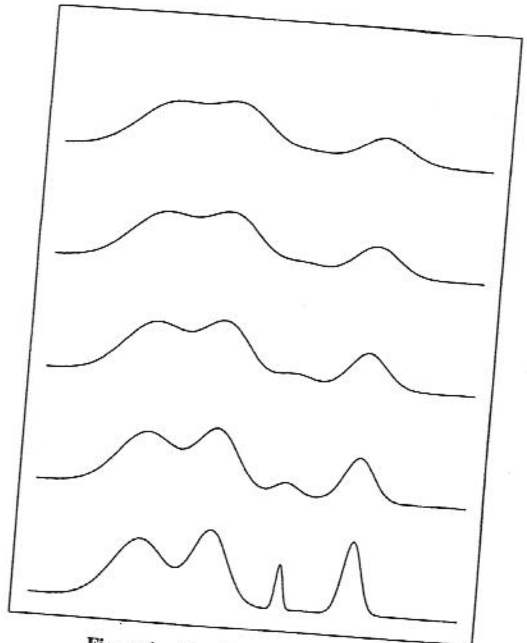


Figure 1. *Sample scale space.*

Intuitively, in Gaussian scale spaces, the scale dimension is not commensurate with the spatial dimensions. Eberly (1994) develops a metric for scale space that formalizes this idea. He assumes that measurements in scale space should have rotational invariance, translational invariance, and zoom invariance, where *zoom* is a change in the units of measurement, or equivalently, a uniform magnification in the scale and space variables. Under these conditions, the metric for scale space is

$$ds^2 = \frac{dx \cdot dx}{\sigma^2} + \frac{d\sigma^2}{\sigma^2},$$

where s denotes arc length, x denotes spatial coordinates, and σ denotes the scale coordinate. With this metric, scale space has hyperbolic geometry.

Shape is often defined as those characteristics of a figure that are unchanged when the figure is translated, rotated, or uniformly scaled (Goodall, 1991). When describing shape, zoom invariance is particularly useful, because distances within the figure remain unchanged when the figure changes scale.

3. MULTI-SCALE TEMPLATE PRIORS

This section develops a class of prior distributions that can be used to incorporate information from a template into image analysis. Many of the ideas are drawn from the scale space methods discussed in the previous section. The overall goal is to develop a probability distribution over the space of images that assigns high probability to images that are like a template and low probability to images that are unlike a template. Let $y = f(x)$ denote a candidate image. The prior probability distribution over the space of images has the form $P(y) = \frac{1}{Z}\exp(U(y))$. Notice that this is the form of a Gibbs distribution, although the probability is not explicitly formulated in terms of cliques and potential functions.

The class of priors is specified hierarchically via a branching structure that starts at large-scale features and terminates at small-scale detail. More specifically, the first node in the branching structure corresponds to the large-scale center of the template. The feature that identifies the center is usually some measure of medialness. Next, n locations $(\boldsymbol{x}_i, \sigma_i)$ in scale space are chosen where there are features, $\phi_i()$, of interest. Locations of interest are typically at medial, corner, boundary, or junction points in the image and its background. Finally, a hierarchical branching structure is established for the locations. The hierarchical structure groups the locations of interest into "branches" that contain ordered sets of related features; often, moving from the top of the branch to the bottom corresponds to moving from large-scale features to small-scale features. For example, one branch for a hand template might contain the large-scale center of the hand followed by the center of the index finger followed by the center of the index fingernail.

Three components are important for specifying $U(\boldsymbol{y})$: feature similarity functions, V_i; location similarity functions, K_i; and feature weights, w_i. The V_i penalize features in a candidate image where they differ from features in the template. Let $L^{(\boldsymbol{y})}(\boldsymbol{x}, \sigma) = f(\boldsymbol{x}) * G(\boldsymbol{x}; 0, \sigma^2 I)$ denote the Gaussian scale space of the candidate image, $L^{(t)}$ the Gaussian scale space of the template, $(\boldsymbol{x}_i, \sigma_i)$ the ith point of interest in the template, and $\phi_i()$ the features of interest at $(\boldsymbol{x}_i, \sigma_i)$. Define

$$V_i(\boldsymbol{y}) = V_i(\boldsymbol{x}, \sigma) = \exp((\phi_i(L^{(\boldsymbol{y})}(\boldsymbol{x}, \sigma)) - \phi_i(L^{(t)}(\boldsymbol{x}_i, \sigma_i)))' \Sigma_i^{-1} (\phi_i(L^{(\boldsymbol{y})}(\boldsymbol{x}, \sigma)) - \phi_i(L^{(t)}(\boldsymbol{x}_i, \sigma_i)))).$$

Evaluating $V_i(\boldsymbol{x}, \sigma)$ at each point in the scale space provides a measure of how closely the features at each point correspond to the features expected at $(\boldsymbol{x}_i, \sigma_i)$. This formulation of V_i requires that features of interest be scale invariant, since V_i is evaluated in scale space. Σ_i can be approximated by considering affine transformations of the template (Wilson, 1995).

The primary purpose of the distance similarity function $K_i(\boldsymbol{x}, \sigma)$ is to capture how far it is reasonable for features in the template to move and still have an image that is "like" the template. If scale space had a Euclidean metric, a natural choice for $K_i()$ would be a multivariate normal distribution. However, since scale space has hyperbolic geometry, it makes more sense to base K_i on the scale space metric. Define

$$K_i(\boldsymbol{x}, \sigma) = \exp(-\frac{1}{2\gamma_i^2} d_{ss}^2((\boldsymbol{x}, \sigma), (\boldsymbol{x}_i, \sigma_i))),$$

where $d_{ss}((\boldsymbol{x}, \sigma_i), (\boldsymbol{x}_i, \sigma))$ denotes the scale space distance between $(\boldsymbol{x}, \sigma)$ and $(\boldsymbol{x}_i, \sigma_i)$. More details on measuring scale space distances can be found in Eberly (1994).

The feature weights, w_i, for each location of interest allow the assignment of more or less important nodes along the branches of the hierarchical structure. It may be the case that certain features are more or less important in the description of an image. For example, for something to be a smiling face, it should have two eyes and a smiling mouth—these features should have large weights. It is less important that the smiling face have a nose, although to exactly match a template a nose may have to be present. The nose would receive correspondingly less weight. The weights can also be adjusted to preserve the template's probability if a location is added or removed.

Let the subscripts i and j denote the ith node along the jth branch, (x^t, σ^t) the location of a feature of interest in the template, (x^m, σ^m) the location of a feature of interest in the candidate image, and (x^e, σ^e) the expected location of a feature of interest in the candidate image. Then

ally via a branching structure that starts at large-
detail. More specifically, the first node in the
-scale center of the template. The feature that
of medialness. Next, n locations $(\boldsymbol{x}_i, \sigma_i)$ in scale
(), of interest. Locations of interest are typically
nts in the image and its background. Finally, a
ed for the locations. The hierarchical structure
es" that contain ordered sets of related features;
e bottom corresponds to moving from large-scale
one branch for a hand template might contain the
center of the index finger followed by the center

ecifying $U(\boldsymbol{y})$: feature similarity functions, V_i;
eights, w_i. The V_i penalize features in a candidate
emplate. Let $L^{(\boldsymbol{y})}(\boldsymbol{x}, \sigma) = f(\boldsymbol{x}) * G(\boldsymbol{x}; 0, \sigma^2 I)$
date image, $L^{(t)}$ the Gaussian scale space of the
the template, and $\phi_i()$ the features of interest at

$$)) - \phi_i(L^{(t)}(\boldsymbol{x}_i, \sigma_i)))'\Sigma_i^{-1}(\phi_i(L^{(\boldsymbol{y})}(\boldsymbol{x}, \sigma)) - \phi_i(L^{(t)}(\boldsymbol{x}_i, \sigma_i)))).$$

le space provides a measure of how closely the
ures expected at $(\boldsymbol{x}_i, \sigma_i)$. This formulation of V_i
riant, since V_i is evaluated in scale space. Σ_i can
ormations of the template (Wilson, 1995).
larity function $K_i(\boldsymbol{x}, \sigma)$ is to capture how far it is
and still have an image that is "like" the template.
al choice for $K_i()$ would be a multivariate normal
hyperbolic geometry, it makes more sense to base

$$\frac{1}{\gamma_i^2} d_{ss}^2((\boldsymbol{x}, \sigma), (\boldsymbol{x}_i, \sigma_i))),$$

space distance between $(\boldsymbol{x}, \sigma)$ and $(\boldsymbol{x}_i, \sigma_i)$. More
n be found in Eberly (1994).
n of interest allow the assignment of more or less
rarchical structure. It may be the case that certain
cription of an image. For example, for something
and a smiling mouth—these features should have
ling face have a nose, although to exactly match a
nose would receive correspondingly less weight.
the template's probability if a location is added or

ode along the jth branch, (x^t, σ^t) the location of a
he location of a feature of interest in the candidate
a feature of interest in the candidate image. Then

$U(\boldsymbol{y})$ can be defined as

$$\begin{aligned}
U(\boldsymbol{y}) = \sum_{j \in \textit{branches}} & w_{1j} \int V_{1j}(x_{1j}, \sigma_{1j}) K_{1j}(x_{1j} - x_{1j}^t, \sigma_{1j} - \sigma_{1j}^t)\, dx_{1j}\, d\sigma_{1j} + \\
& w_{1j} w_{2j} \int\!\!\int V_{1j}(x_{1j}, \sigma_{1j}) K_{1j}(x_{1j} - x_{1j}^t, \sigma_{1j} - \sigma_{1j}^t) V_{2j}(x_{2j}, \sigma_{2j}) \\
& K_{2j}(x_{2j} - (x_{1j} + x_{2j}^e - x_{1j}^m), \sigma_{2j} - (\sigma_{1j} + \sigma_{2j}^e - \sigma_{1j}^m))\, dx_{1j}\, d\sigma_{1j}\, dx_{2j}\, d\sigma_{2j} + \ldots + \\
& w_{1j} \ldots w_{n_j j} \int \ldots \int V_{1j}(x_{1j}, \sigma_{1j}) K_{1j}(x_{1j} - x_{1j}^t, \sigma_{1j} - \sigma_{1j}^t) \ldots V_{n_j j}(x_{n_j j}, \sigma_{n_j j}) \\
& K_{n_j j}(x_{n_j j} - (x_{(n_j-1)j} + x_{n_j j}^e - x_{(n_j-1)j}^m), \sigma_{n_j j} - (\sigma_{(n_j-1)j} \\
& \quad + \sigma_{n_j j}^e - \sigma_{(n_j-1)j}^m))\, dx_{1j} d\sigma_{1j} \ldots dx_{n_j j} d\sigma_{n_j j}.
\end{aligned}$$

According to this definition, the probability of a candidate image can be evaluated through the following steps:

(1) Use the feature similarity function, V_{1j}, to find places in the image with features like the template at the first node on the branch.
(2) Multiply the location similarity function, K_{1j}, by the similarity function to identify places similar in location and features.
(3) Search for the next set of features at the appropriate distance from the location of the maximum, $(x_{1j}^m, \sigma_{1j}^m)$, of $V_{1j}K_{1j}$. $(x_{1j}^m, \sigma_{1j}^m)$ is taken to be the location of the feature of interest. The expected location of the next point along the branch, $(x_{2j}^e, \sigma_{2j}^e)$ is calculated relative to $(x_{1j}^m, \sigma_{1j}^m)$. The location function for the second set of features is $V_{1j}K_{1j}$, translated so that its maximum is at $(x_{2j}^e, \sigma_{2j}^e)$, and then convolved with the *a priori* location function, K_{2j}.
(4) Repeat (2) and (3) along each branch in the tree structure and sum over the branches.

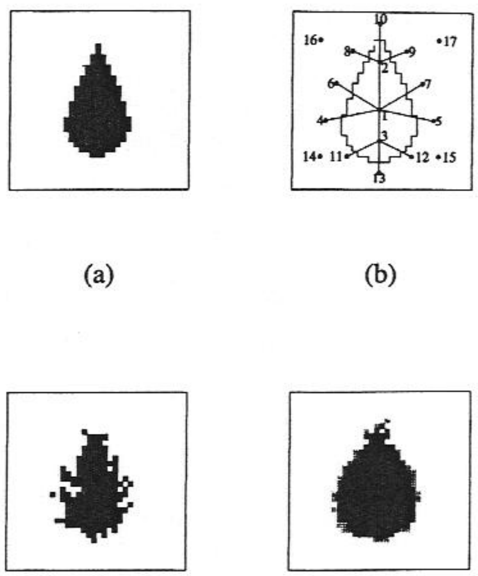
Several properties of $U(y)$ should be noted. Features of interest do not have to be found precisely at the n locations of the template. Since the V_i are multiplied by a location function, images that have the appropriate features near where the template has them are downweighted slightly, as is appropriate for a slightly deformed image. Local scale variability is accounted for since the location function allows movement in both scale and space.

Aside from the first node in a branch, location variability depends both on the prior specification of K_i and on what was seen at the previous node along the branch. If a large area matched well in location and features at the previous node, the expected location of the features at the next scale is less certain and the location variability increases, allowing the uncertainty about the location of features to increase with decreasing scale.

This prior assigns positive probability along a branch until matching nodes are no longer found. Since features are incorporated multiplicatively along a branch, once a feature at a location is deemed "not like the template," the calculation of $U(y)$ is unaffected by features seen further down the branch.

The branching description of the template allows the prior to handle obstructed views. If one branch of the template is blocked, the probability can still be relatively high due to contributions from other branches.

To demonstrate the utility of this model, consider the following simulation results. The first experiment used the template in Figure 2(a). The hierarchical description in Figure 2(b) was chosen. Notice that the points of interest are both medial (nodes 1, 2, 3, 14, 15, 16, 17) and boundary points. The description in Figure 2(b) captures only the spatial locations of the features of interest: the scales of interest range from 1 to 7 pixels. The features at each node were taken to be the scaled gradient magnitude and the scaled Laplacian. Samples from the prior were obtained using the Metropolis-Hastings algorithm. A sample image is given in Figure 2(c), and an estimate of the prior mean is given in Figure 2(d).



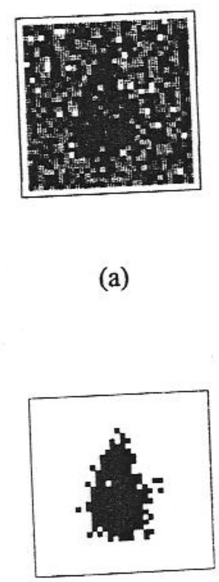
(a) (b) (c) (d)

Figure 2. *Prior simulation: (a) template; (b) template description; (c) sampled image; (d) prior mean.*

The second experiment demonstrates the utility of the model for segmentation. *Segmentation* identifies and labels structures within images. Suppose that the image in Figure 3(a) needs to be segmented or divided into figure and background. Suppose also that the likelihood function for the observed image is known: conditional on the true image, any pixel inside the region of interest is Poisson distributed with mean five and any background pixel is Poisson distributed with mean one. To perform a Bayesian analysis, a prior distribution must be specified on the space of true images. Take the prior to be the one simulated in the first example. Samples from the posterior distribution (Figure 3(c)) were obtained using the Metropolis-Hastings algorithm. Figure 3(d) is an estimate of the posterior mean. Areas of Figure 3(d) that are dark have high probability of being in the region of interest. This can be compared to Figure 3(b), which shows the "true" image that was used to generate Figure 3(a).

4. DISCUSSION

There are a variety of issues that can be explored in conjunction with this class of priors. A central concern is the development of effective template descriptions. There are a variety of differential geometry shape descriptors that can be applied in scale space, but heuristics for choosing features, feature locations, and scales need to be explored and formalized. It will



(a)

(c)

Figure 3. *Segmentation: (a) noisy image; (b) t*

also be important to explore the sensitivity of
preliminary work in this direction has been p
describe the selection of features for use in mu

Another issue under investigation is hyper-
tation, the hyper-parameters are fixed within
of models would permit these parameters to
images to define reasonable values.

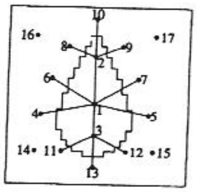
ACKNOWI

This material is based upon work supported
Fellowship, PHS NCI grant CA 56671, and N

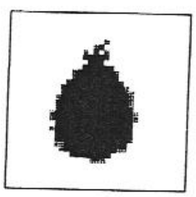
REFE

Besag, J. (1974). Spatial interaction and the statistical a
Besag, J. (1986). On the statistical analysis of dirty pic
Bookstein, F. (1986). Morphometrics: size and shape
Burt, P. (1981). Fast filter transforms for image proces
Coggins, J. and Huang, C. (1993). Defining optimal fe *Mathematical Methods in Medical Imaging II, Pro Vol. 2035 held in San Diego, CA, 15-16 July 199* 80–88.
Crowley, J. (1981). *A Representation for Visual Infor* Institute, Pittsburgh, PA.

e from 1 to 7 pixels. The features at each node
le and the scaled Laplacian. Samples from the
ıgs algorithm. A sample image is given in Figure
ı in Figure 2(d).



(b)



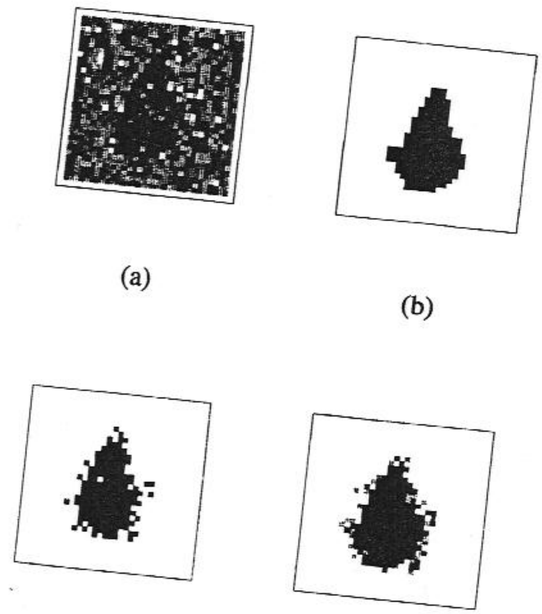
(d)

ıte description; (c) sampled image; (d) prior mean.

ity of the model for segmentation. *Segmenta-*
:. Suppose that the image in Figure 3(a) needs
ıund. Suppose also that the likelihood function
the true image, any pixel inside the region of
l any background pixel is Poisson distributed
a prior distribution must be specified on the
simulated in the first example. Samples from
ıed using the Metropolis-Hastings algorithm.
Areas of Figure 3(d) that are dark have high
can be compared to Figure 3(b), which shows
3(a).

SION

l in conjunction with this class of priors. A
ımplate descriptions. There are a variety of
be applied in scale space, but heuristics for
ıeed to be explored and formalized. It will



(a) (b) (c) (d)

Figure 3. *Segmentation: (a) noisy image; (b) template; (c) sampled image; (d) posterior mean.*

also be important to explore the sensitivity of the results to the choice of descriptors. Some preliminary work in this direction has been performed by Coggins and Huang (1993), who describe the selection of features for use in multi-scale statistical pattern recognition.

Another issue under investigation is hyper-parameter estimation. In the current implementation, the hyper-parameters are fixed within the template description. A more flexible class of models would permit these parameters to vary and would use information from observed images to define reasonable values.

ACKNOWLEDGMENTS

This material is based upon work supported under a National Science Foundation Graduate Fellowship, PHS NCI grant CA 56671, and NSF SCREMS DMS 9305699.

REFERENCES

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. B* **36**, 192–236.
Besag, J. (1986). On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. B* **48**, 259–302.
Bookstein, F. (1986). Morphometrics: size and shape spaces. *Statist. Sci.* **1**, 181–242.
Burt, P. (1981). Fast filter transforms for image processing. *CVGIP* **16**, 20–51.
Coggins, J. and Huang, C. (1993). Defining optimal features sets for segmentation by statistical pattern recognition. *Mathematical Methods in Medical Imaging II, Proceedings of the International Society for Optical Engineering Vol. 2035 held in San Diego, CA, 15-16 July 1993* (D. Wilson and J. Wilson, eds.), Bellingham, WA: SPIE, 80–88.
Crowley, J. (1981). *A Representation for Visual Information*. Ph.D. Thesis, Carnegie Mellon University, Robotics Institute, Pittsburgh, PA.

Eberly, D. (1994). A differential geometric approach to anisotropic diffusion. *Geometry-Driven Diffusion in Computer Vision* (B. ter Haar Romeny, ed.), Boston: Kluwer Academic.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intelligence* **6**, 721–741.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. B* **53**, 285–339.

Grenander, U. and Miller, M. (1994). Representations of knowledge in complex systems. *J. Roy. Statist. Soc. B* **56**, 549–603.

Hackbusch, W. (1985). *Multi-Grid Methods and Applications*. New York: Springer-Verlag.

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

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

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

Lifshitz, L. and Pizer, S. (1990). A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. Patt. Anal. Mach. Intelligence* **12**, 529–541.

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

Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intelligence* **11**, 674–693.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intelligence* **12**, 629–639.

Staib, L. and Duncan, J. (1992). Boundary finding with parametrically deformable models. *IEEE Trans. Patt. Anal. Mach. Intelligence* **14**, 1061–1075.

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

Wilson, A. (1995). *Statistical Models for Shapes and Deformations*. Ph.D. Thesis, Duke University, Institute of Statistics and Decision Sciences, Durham, NC.

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

Yuille, A. and Poggio, T. (1986). Scaling theorems for zero crossings. *IEEE Trans. Patt. Anal. Mach. Intelligence* **8**, 15–25.

BAYESIAN STATISTICS 5, pp. 809–813
J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith
© Oxford University Press, 1996

On Improving a Moc
Experts' F

M. P. WIPER and
Goldsmiths C

SUMM

In Wiper and French (1994), a model was introduced
light of hearing an expert's quantiles for the foreca
assumption that the Decision Maker believed it possi
Here, we use the Gibbs sampler to remove this flaw i

Keywords: NORMAL-GAMMA MODEL; GIBBS SAM

1. INTROD

A fully Bayesian Decision Maker (DM) wishes
uous variables $X_1, X_2, \ldots$. We assume that the
variable in the form of quantiles $(q_1, q_2, \ldots, q$
The problem facing the DM is to update her ini
information.

This problem was called "The Expert Prob
models have been proposed, e.g. Morris (19
none has been entirely successful. One model
introduced in Wiper (1990) and Wiper and Fre
indicate a theoretical flaw.

Following Wiper and French (1995), we
mation as follows:

$$q_{ij} \rightarrow z_{ij} = $$

where q_{ij} is the expert's p_j quantile for X_i an
function for X_i. Transforming the data in this
variables being measured on different scales a
forecasts of the DM and those of the expert: s
for details.

After transforming the variables

$$X_i \rightarrow Y_i = $$

Wiper (1990) and Wiper and French (1995)

$$(Z_i|M = m,$$
$$(M|y_i, \mathcal{R} = R) = (M|$$
$$(\mathcal{R}|$$