Published in *Proc. SPIE* **2701**, pp. 145-155 (1996)
Medical Imaging: Image Processing, K. M. Hanson, ed.

Improved convergence of gradient-based reconstructions using multiscale models

Gregory S. Cunningham[a], Igor Koyfman[b], and Kenneth M. Hanson[a]

[a]Los Alamos National Laboratory, MS P940, Los Alamos, NM 87545
[b]Department of Applied Mathematics, UCLA

ABSTRACT

Geometric models have received increasing attention in medical imaging for tasks such as segmentation, reconstruction, restoration, and registration. In order to determine the best configuration of the geometric model in the context of any of these tasks, one needs to perform a difficult global optimization of an "energy" function that may have many local minima. Explicit models of geometry, also called deformable models, snakes, or active contours, have been used extensively to solve image segmentation problems in a non-Bayesian framework. Researchers have seen empirically that multi-scale analysis is useful for convergence to a configuration that is near the global minimum. In this type of analysis, the image data are convolved with blur functions of increasing resolution, and an "optimal" configuration of the snake is found for each blurred image. The configuration obtained using the highest resolution blur is used as the solution to the global optimization problem. In this article, we use explicit models of geometry for a variety of Bayesian estimation problems, including image segmentation, reconstruction and restoration. We introduce a multi-scale approach that blurs the geometric model, rather than the image data, and show that this approach turns a global, highly nonquadratic optimization into a sequence of local, approximately quadratic problems that converge to the global minimum. The result is a deterministic, robust, and efficient optimization strategy applicable to a wide variety of Bayesian estimation problems in which geometric models of images are an important component.

Keywords: snakes, active contours, deformable models, Bayesian analysis, global optimization

1. INTRODUCTION

Geometric models have received increasing attention in medical imaging for tasks such as segmentation, reconstruction, restoration, and registration. Explicit models of geometry, e.g. snakes[1] or deformable models,[2,3] are desirable when some user interaction can be tolerated. The parameterization of an explicit model can easily be changed to control its "stiffness", which is useful, e.g. to allow a segmentation algorithm to segment regions of the image where the edge disappears. Controlled continuity splines[4,5] can be used to match edges with sharp corners. Simple calculations can be performed on the optimal configuraton of an explicit model, e.g., one can easily compute the area contained by the model.[6] Finally, if the geometric model is part of a Bayesian solution to the analysis problem, the confidence one should have in the final configuraton can be assessed using directed[7] or stochastic[8] searches of the probability of nearby configurations. One could compute the posterior probability that the area contained by the geomtric model is the calculated value plus or minus 10 percent, for example.

On the other hand, when a completely automatic optimization is needed, implicit models, i.e. Markov random field (MRF) prior probability distributions,[9] may be preferrable because they require almost no user involvement and can handle very complex geometrical configurations. MRF models, since they are part of a Bayesian solution to the problem, can be queried to obtain credible intervals as mentioned above, or can be used to generate more robust estimators than the MAP solution, e.g. the mean posterior or median posterior solution. However, calculations tied to the optimal configuration may be more difficult than for an explicit model, since the geometry is not a set of explicit, closed paths. Furthermore, control of the "stiffness" does not appear to be meaningful for these types of models.

Further author information –
G.S.C.(correspondence): Email: cunning@lanl.gov; Telephone: 505-667-2562; Fax: 505-665-3359
K.M.H.: Email: kmh@lanl.gov; WWW: http://www.lanl.gov/home/kmh/; Telephone: 505-667-1402
I.K.: Email: ikoyfman@math.ucla.edu

In this article, we use explicit models of geometry in the context of a Bayesian approach to many 2D image analysis problems, including reconstruction, restoration and segmentation. A crucial component in the way that we accomodate geometric models in the fitting of image data is the transformation of a geometric model into image space. We show that this transformation is highly nonlinear over the scale of the image pixel spacing, and that it makes global optimization using deterministic strategies essentially impossible without some modification. We show that blurring the result of this nonlinear transformation increases the scale over which the blurry transformation is linear. Since linear models of images produce quadratic log likelihood functions (in the case of Gaussian noise), we show that, if a blur is used that is much larger than the largest perturbation to the geometric model which could result in a more probable configuration, then the optimization problem at that scale is quadratic, and can be solved efficiently and automatically. We propose a multi-scale approach that starts at coarse resolution and proceeds carefully to fine resolution, which may sound very similar to multi-scale approaches applied to image segmentation using snakes. It should be stressed, though, that we are applying the blur to the geometric model, rather than the image data, and that what we are proposing is a robust, deterministic, global optimization strategy that is useful for a whole variety of Bayesian estimation problems which rely on explicit models of geometry.

2. THE BAYESIAN APPROACH

The Bayesian approach provides a complete methodology for model building from measurements. Given a set of measurements, it can be used to answer the fundamental questions: What are the "best" values of the parameters used in a particular model? What is the uncertainty in the parameter values implied by uncertainty in the measurements. Which of two models is more appropriate to use? If using prior information, how strongly should one rely on it? The reader is encouraged to learn more about the fundamentals of Bayesian analysis by studying some of the references available.[10–13]

In Bayesian analysis, the posterior probability summarizes the full state of knowledge concerning a given situation. Given the data $\mathbf{d}$, the posterior probability of parameters $\mathbf{a}$ for a particular model M is given by Bayes' law

$$p(\mathbf{a}|\mathbf{d}, M) \;=\; \frac{p(\mathbf{d}|\mathbf{a}, M)\, p(\mathbf{a}|M)}{p(\mathbf{d}|M)} \;, \tag{1}$$

where $p(\mathbf{d}|\mathbf{a}, M)$, the probability of the observed data given $\mathbf{a}$, is called the likelihood, $p(\mathbf{a}|M)$ is the prior probability of $\mathbf{a}$, and $p(\mathbf{d}|M)$ is called the evidence. The likelihood is the probability distribution of the fluctuations in the measurements about their values predicted by the parameters. The prior probability $p(\mathbf{a})$ encompasses the full prior information about the parameter set $\mathbf{a}$. The prior information might be experimentally based or subjective. The evidence provides the normalization for the posterior, and is not important in most analyses. It is often convenient to deal with the minus log posterior:

$$-\log[p(\mathbf{a}|\mathbf{d})] = \Psi(\mathbf{a}) = \Phi(\mathbf{a}) + \Pi(\mathbf{a}) - E \;\;, \tag{2}$$

where Φ, Π, and E are the negative logarithms of the likelihood, the prior, and the evidence, respectively. Φ and Π depend on the parameters, while E does not. We seek to find the parameter values that maximize the posterior, called the maximum *a posteriori* (MAP) solution. Of course, the MAP estimate is found by minimizing Ψ with respect to the parameters, yielding the estimated parameter values $\hat{\mathbf{a}}$. The condition for the MAP solution is $\frac{\partial \Psi}{\partial a_i} = 0$ for all parameters a_i, providing there are no constraints on the parameters themselves.

Under the assumption that the measurements are subject to additive, Gaussian noise, the minus log-likelihood is half of the familiar chi-squared, $\Phi(\mathbf{a}) = \frac{1}{2}\chi^2 = \frac{1}{2}\sum_i \sigma_i^{-2}(d_i - \hat{d}_i)^2$, the sum of the squared residuals (the difference between observed measurements and their values predicted by the the estimated parameters $\hat{\mathbf{a}}$) divided by the variance of the noise, σ^2.

The tool we have developed for implementing the Bayesian approach for analysis of 2D image data using geometric models is called the Bayes Inference Engine (BIE). This versatile application allows one to develop complex geometric models for the objects under study as well as complex models of the radiographic measurement process (see Fig. 1). The BIE can be used to draw a data-flow diagram that produces a predicted image given a configuration of the potentially complex object model, which could include geometric model(s) coupled to density variation models. The predicted data is compared with the raw data to produce a minus log likelihood (the box with the Φ in it). Minus

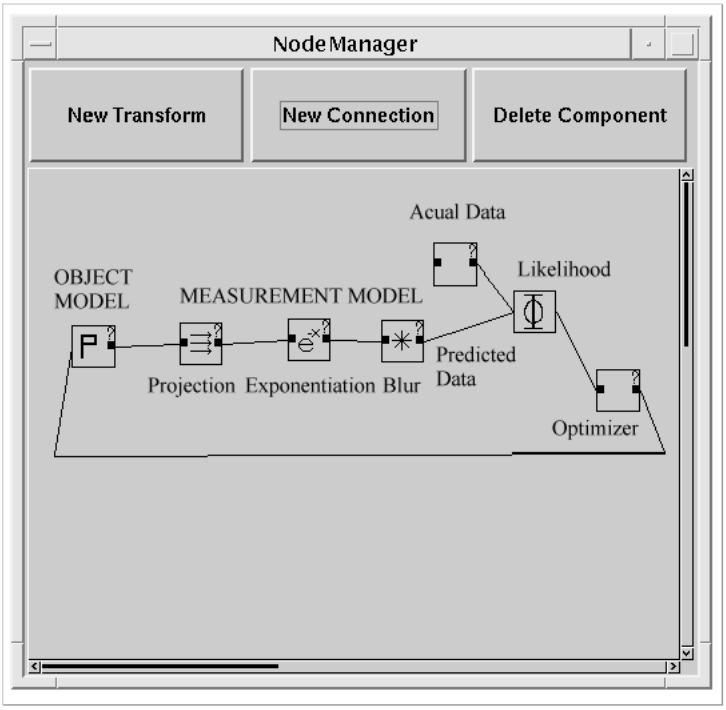

Figure 1. The BIE allows a user to model objects and measurements on objects by constructing a visual data-flow diagram. The box with a **P** in it specifies the model of the object, while the first 3 transforms to the right simulate a simple radiographic measurement system to produce a prediction of the 2D image data. The prediction is to be matched with a real 2D image by minimizing the minus log likelihood, **Φ**, w.r.t. anything connected to the right side of the Optimizer.

log priors can be hooked up directly to components of the object model. The minus log likelihood and minus log prior can be summed to produce a minus log posterior. Any of the three minus log probabilities can be minimized if connected to the left side of an Optimizer. Models that are connected to the right side of the Optimizer are manipulated in order to achieve the minimization. Various aspects of the BIE are described elsewhere.[7,14–17]

Since the predicted data comprise an image, at some point the geometric models have to be transformed into images. This transformation is highly nonlinear over the scale of the image pixel size, and is the major source of difficulty in the global optimization problem - finding the MAP solution (see Fig. 2).

3. EXPLICIT MODELS OF GEOMETRY

Explicit models of geometry have been used in Bayesian and non-Bayesian problem formulations. Snakes were originally introduced to solve a non-Bayesian formulation of the 2D image segmentation problem.[1] The optimization algorithm therein[1] relies on an interpretation of the derivative of the energy function (the function to be minimized) as exerting a force on the parameters of the snake. Dynamic evolution equations are constructed using this interpretation and then time-advanced until the snake reaches equilibrium, when the net force on the snake is zero everywhere and the model has achieved a local minimum in energy.

A multi-scale approach is useful for finding the configuration of the snake that yields a global minimum in energy. In the multi-scale approach suggested by Kass et. al.,[1] the energy function is Gaussian-blurred, and the snake that comes to equilibrium for that blurred energy function is obtained. Then, the snake that results from this coarse-scale energy distribution is "re-equilibrated" using an energy distribution that is blurred slightly less than in the previous iteration. This process can be continued until the blur is as small as the desired accuracy on the snake. Empirically, it has been found that this type of multi-scale approach is useful in avoiding local minima in the global optimization problem, and the explanation given for this desirable behavior is called scale-space continuation.[18,19] In more recent

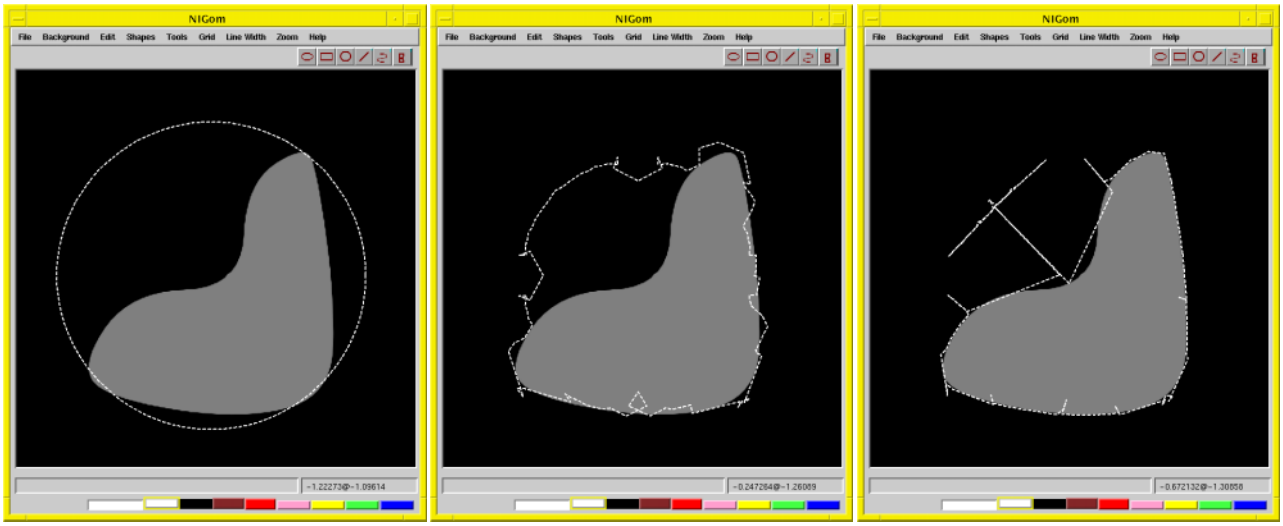

Figure 2. Naive optimization of the polygonal vertices for a noiseless image segmentation problem results in very poor behavior. The initial configuration on the left is a 100-sided polygon. Two global steps using gradient descent results in the configuration in the middle, and the configuration on the right is the local minimum after a dozen global steps. Resampling the polygon after every global step so that it is uniform in arclength reduces the number of self-intersections, but results in excruciatingly slow convergence.

work,[20,21] scale-space continuation is used to argue that the sequence of configurations share desirable relationships that are useful in defining a hierarchical representation of the geometric model.

The use of snakes as originally proposed[1] is not easily formulated as a Bayesian estimation problem, since the force that derives from the image data, which one would like to interpret as the gradient of a minus log likelihood w.r.t. the snake model, is difficult to interpret as coming from a "goodness-of-fit" of the image predicted by the snake model. The snake alone can not easily be used to predict a complete image in which the snake is just the boundary that separates "inside" from "outside", because there is no explicit model for the values of the inside and outside in the snake approach. Thus, snakes seem to be more of a feature-detecting model than an image-fitting model. On the other hand, the force that derives from the snake model itself is easily interpreted as a regularizing prior. In fact, when snakes are extended to surfaces for the purpose of fitting 3D scattered surface data,[4] the energy function that depends on the data is chi-squared, and the optimization problem is identical to a Bayesian MAP estimation problem.

The snake approach appears to be easy-to-use and flexible, since no model of image variations inside or outside the snake is required of the data analyst, and many energy functions have already been proposed to solve a variety of problems in image segmentation. However, snakes are inappropriate if one has a good deal of information about what the image variations should be inside and outside the snake. For example, the knowledge that the image should be near value a inside the snake and near value b outside the snake is powerful information, and can be incorporated easily in a Bayesian approach. Even knowledge that the image must be smooth inside the snake and smooth outside, but can have a discontinuity along the snake boundary can easily be incorporated into a Bayesian approach with simple quadratic log priors on the image variation inside and outside. Finally, if one wants to ask probabilistic questions about the solution, e.g. what is the probability that the area contained by the geometry is between A_1 and A_2, a Bayesian approach is necessary.

Explicit models of geometry have been used in a Bayesian formulation of the limited-view tomography problem,[5,22,23] where a 2D slice of a 3D object is reconstructed using two views, as well as in image segmentation and recognition using deformable models.[2]

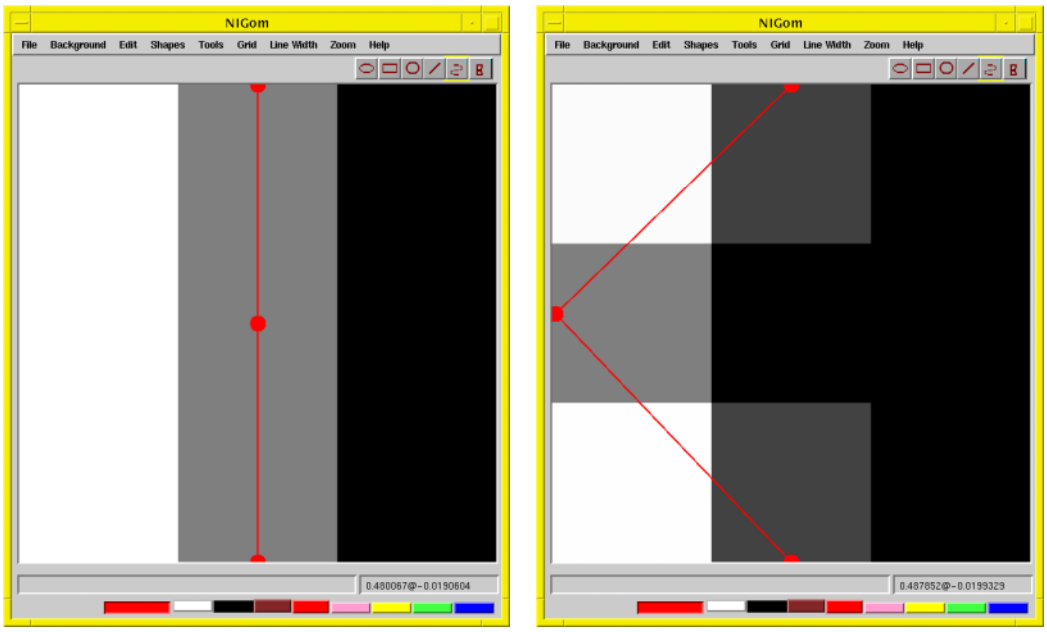

Figure 3. The transformation of a polygon into an image is nonlinear over the scale of the image pixel width. The polygon is filled to the left of the two line segments seen in the above 3x3 subsamples of the output image. The center pixel (in the left image) has a value of $\frac{1}{2}$ times the value of the polygon interior, but as the vertex in the center is pushed 1.5 pixels to the left, the center pixel value (in the right image) goes to zero and stays zero for further movement of the vertex to the left.

4. MULTI-SCALE AS A GENERAL-PURPOSE GLOBAL OPTIMIZATION STRATEGY

As mentioned in Section 2, in a Bayesian formulation of an image analysis problem that utilizes a geometric model, the geometric model has to be transformed into an image at some point. For the following discussion, assume that the geometry is defined by a set of vertices $\mathbf{P} = \{P_k = (x_k, y_k)\}_{k=1}^{N}$ that are connected by straight-line segments, and that the image is defined by a set of pixel values, $\mathbf{I} = \{I_{ij}\}_{i,j=1}^{M}$. More complicated geometric models are discussed in the next section. The discrete image $\mathbf{I}$ is shorthand for a representation of the continuous image $I(x, y)$ that has value I_{ij} in the region $u_{i-1} \leq x \leq u_i, v_{j-1} \leq y \leq v_j$, where $\{u_i\}_0^M$ and $\{v_j\}_0^M$ are uniformly distributed over intervals in $\Re$. The transformation $\mathbf{P} \rightarrow \mathbf{I}(\mathbf{P})$ is accomplished by computing, for each pixel (i, j), the area of intersection of the polygon described by $\mathbf{P}$ and the pixel with which the image value I_{ij} is associated. If I_{ij} is viewed as a function of $\mathbf{P}$, then this function is highly nonlinear over the scale of the pixel spacing, because, as a vertex in $\mathbf{P}$ passes through the interior of the pixel (i, j), the value $I_{i,j}$ might go from zero to close to one while the vertex that is passing through only moves the width of the pixel (see Fig. 3). This nonlinearity is the major source of difficulty in solving the global optimization problem - finding the MAP solution (see Fig. 2).

The transformation $\mathbf{P} \rightarrow \mathbf{I}(\mathbf{P})$ described above is a special case of taking the projection of $\mathbf{P}$ onto the image with continuous coordinates, $I(x, y)$, blurring $I(x, y)$ with a function $G(x, y; \sigma)$, and then sampling it on a discrete, uniform lattice. The intermediate continuous image $I(x, y)$ has values $I(x, y) = 1$ if $(x, y) \in R_P$ and $I(x, y) = 0$ otherwise, where R_P is the region covered by the linear polygon $\mathbf{P}$. Thus, the transform $\mathbf{P} \rightarrow \mathbf{I}(\mathbf{P})$ is just a case of blurring $I(x, y)$ with a rectangular function that has the same width and height as one of the pixels and discrete sampling at the centers of the pixel boxes. The 2D Gaussian, $G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp(\frac{x^2+y^2}{\sigma^2})$ could be used as the blurring function instead.

The partial derivative $\frac{\partial I_{ij}}{\partial x_k}$ has two components, one from perturbing the line segment $P_k \rightarrow P_{k+1}$ and one from perturbing the line segment $P_{k-1} \rightarrow P_k$ (note: let $P_0 = P_N$ and $P_{N+1} = P_1$). Let's consider the component that comes from perturbing $P_k \rightarrow P_{k+1}$, and assume that a positive perturbation of x_k increases the area of $\mathbf{P}$ (see Fig. 4).

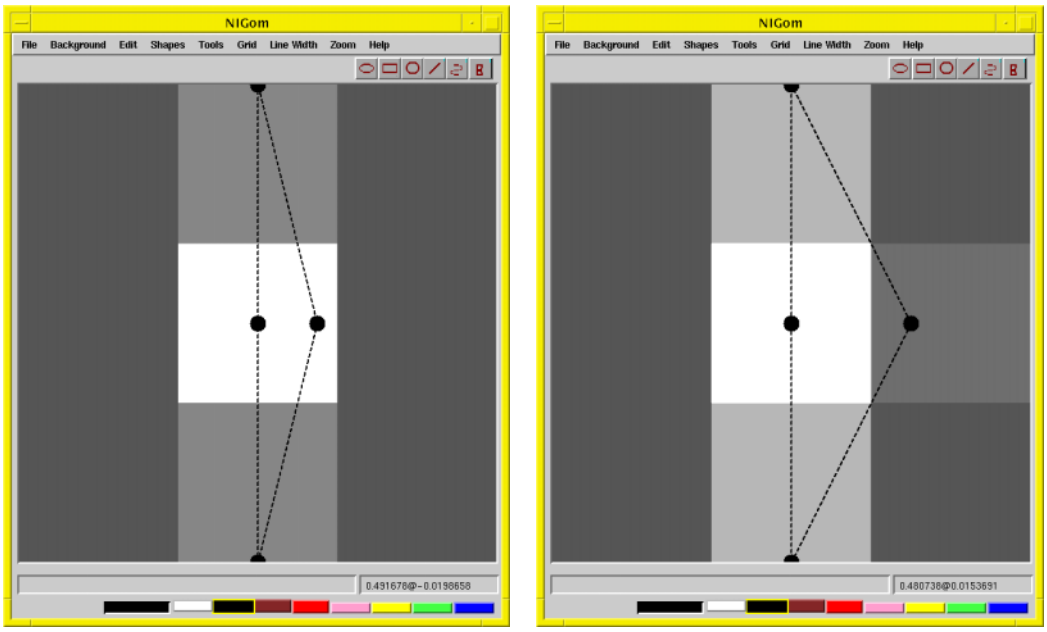

Figure 4. The numerical derivative of the image w.r.t. the x part of the middle vertex of the polygon for a 3x3 subsample of the image. On the left hand side the numerical derivative is only nonzero for the center column of 3 pixels, whereas on the right side, the numerical derivative is nonzero for four of the 9 pixels. The polygon shown is the 4-sided polygon $< P_k \rightarrow P_{k+1} \rightarrow P_k + \delta P_k \rightarrow P_{k-1} >$.

Then this contribution to $\frac{\partial I_{ij}}{\partial x_k}$ is the limit as δx tends to zero of $\frac{1}{\delta x}$ times the integral of $G(x - \frac{u_{i+1}+u_i}{2}, y - \frac{v_{j+1}+v_j}{2}; \sigma)$ over the area defined by the triangle $< P_k \rightarrow P_{k+1} \rightarrow P_k + \delta P_k >$, where $\delta P_k = (\delta x, 0)$. Note that the area of $< P_k \rightarrow P_{k+1} \rightarrow P_k + \delta P_k >$ is $\frac{1}{2}|y_{k+1} - y_k|\delta x$, so that, in the limit, the partial derivative is determined by a linearly-weighted integral of $G(x - \frac{u_{i+1}+u_i}{2}, y - \frac{v_{j+1}+v_j}{2}; \sigma)$ along the line segment $P_k \rightarrow P_{k+1}$. The contribution due to the perturbation of the line segment $P_{k-1} \rightarrow P_k$ is a weighted line integral of the same shifted blur function over the line segment $P_{k-1} \rightarrow P_k$, with a maximum weight that depends on $\frac{1}{2}|y_k - y_{k-1}|$. This derivation holds for any blur function at any pixel sampling point and for any vertex. Obviously, if the blur function has infinite extent, then the derivative of every image point with respect to the x and y components of every polygonal vertex may be nonzero. A similar derivation exists for $\frac{\partial I_{ij}}{\partial y_k}$, but the maximum value of the weight will depend on the differences of the x components of the vertices instead of the y components, as in the above derivation.

For a Gaussian blur function, if σ is large compared to the length of the line segments $P_{k-1} \rightarrow P_k$, then the weighted integral of $G(x - \frac{u_{i+1}+u_i}{2}, y - \frac{v_{j+1}+v_j}{2}; \sigma)$ over the line segment $P_{k-1} \rightarrow P_k$ will approximately be weighted "samples" of $G(x - \frac{u_{i+1}+u_i}{2}, y - \frac{v_{j+1}+v_j}{2}; \sigma)$ at $(x, y) = (x_{k-1}, y_{k-1})$ or $(x, y) = (x_k, y_k)$, depending on which vertex the derivative is being calculated with respect to, and with the weight determined by the total weighted integral value ($\frac{1}{2}|y_{k+1} - y_k|$ in the example above). Thus,

$$\frac{\partial I_{ij}}{\partial x_k} \approx \pm\tfrac{1}{2}(|y_k - y_{k+1}| + |y_k - y_{k-1}|)G(x_k - \frac{u_{i+1} + u_i}{2}, y_k - \frac{v_{j+1} + v_j}{2}; \sigma) \quad (3)$$

$$\frac{\partial I_{ij}}{\partial y_k} \approx \pm\tfrac{1}{2}(|x_k - x_{k+1}| + |x_k - x_{k-1}|)G(x_k - \frac{u_{i+1} + u_i}{2}, y_k - \frac{v_{j+1} + v_j}{2}; \sigma) \quad (4)$$

If σ is also large compared to the components of a perturbation δP_k of the k^{th} vertex of the polygon $\mathbf{P_o}$, then the image that results from the perturbed polygon $\mathbf{I}(\mathbf{P_o} + \delta\mathbf{P})$ is just $\mathbf{I}(\mathbf{P}) + \mathbf{I}(\mathbf{P'})$, where $\mathbf{P'}$ is the four-sided polygon $< P_k \rightarrow P_{k+1} \rightarrow P_k + \delta P_k \rightarrow P_{k-1} >$, and $\delta P_k = (\delta x, \delta y)$ is the perturbation in the k^{th} vertex. The area of $\mathbf{P'}$ is $\frac{1}{2}(|y_k - y_{k+1}| + |y_k - y_{k-1}|)\delta x + \frac{1}{2}(|x_k - x_{k+1}| + |x_k - x_{k-1}|)\delta y$ so that, if the size of $\mathbf{P'}$ is small compared to σ, a

Taylor series expansion of $\mathbf{I}(\mathbf{P})$ about the point $\mathbf{P_o}$ can be used to construct an accurate **linear** approximation to the nonlinear transform $\mathbf{P} \rightarrow \mathbf{I}(\mathbf{P})$ near $\mathbf{P_o}$. The error in such an approximation (the higher order derivatives of $\mathbf{I}(\mathbf{P})$ w.r.t. $\mathbf{P}$) can be bounded by the spatial derivatives of the blur function $G(x, y; \sigma)$. Thus, the Taylor series expansion is good for small perturbations to a single vertex of $\mathbf{P_o}$. It is trivial to extend the analysis to a perturbation of $\mathbf{P_o}$ that involves all of the vertices. In fact, the restriction that the length of the segments $P_k \rightarrow P_{k+1}$ be small in relation to σ can be relaxed. Note that the spatial derivatives of a rectangular blurring function are zero everywhere except at the edge of the blur, where even the first derivative is infinite, making it a suspicious choice for blurring function if the goal is to eliminate bothersome nonlinear discontinuities in the forward transform.

The multi-scale method we propose, then, requires an initial configuration of the geometric model that is close to the global optimum, and a measure of how large the error in the initial configuration is. A sequence of optimization steps is performed at increasing resolutions (decreasing σ), starting with a value of σ that forces the inital optimizaton problem to be nearly quadratic over a range of perturbations to the geometric model that are the size of the error in the initial configuration. At each step, the geometric model will be closer to the global minimum, and thus the region of plausible perturbations will shrink, allowing a smaller value for σ. For the case in which the local optimization problems are exactly quadratic, and under some conditions on the the Jacobian of the image w.r.t. the geometric model, a unique minimum in each sub-problem is guaranteed, and convergence to the unique global minimum is assured.

5. EXAMPLES

The multi-scale approach outlined above is applicable to many Bayesian estimation problems in which geometric models of images play a crucial role. The approach is an appealing alternative to stochastic relaxation strategies, which, although guaranteed to converge to the global minimum, do so at considerable cost in computation time.[6,24]

The derivation in Section 4 showed that multi-scale analysis can be used to linearize the nonlinear transform of a geometric model into an image, over perturbations to the geometry that are much smaller than the scale of the blurring function. This derivation was for the simplest type of geometric model – vertices connected by straight-line segments, which will now be referred to as a simple polygon. More complicated polygonal models are easily handled, however. A very large class of 2D curves can be approximated arbitrarily well by piecewise Bezier polygons.[25] The building block for these types of polygons are Bezier curve patches. A Bezier curve patch is a parameterized curve $(x(t), y(t))$, where $x(t)$ and $y(t)$ are described as weighted sums of Bernstein polynomials:

$$x(t) = \sum_{n=0}^{N} x_n B_n^N(t) \tag{5}$$

$$y(t) = \sum_{n=0}^{N} y_n B_n^N(t) \tag{6}$$

and

$$B_n^N(t) = \frac{N!}{(N-n)!n!} t^n (1-t)^{N-n} \tag{7}$$

The "control points", (x_n, y_n), allow for one Bezier patch to be linked to another by ensuring that the end control points of one patch, $(x_N^{(1)}, y_N^{(1)})$, are identical in value to the beginning control points of another, $(x_0^{(2)}, y_0^{(2)})$. Geometric continuity across patch boundaries can be accomodated easily with linear constraints on the control points of the adjacent patches. One nice attribute of piecewise Bezier polygons is that they can be easily and efficiently transformed into simple polygons by sampling uniformly in the underlying 1D parameter space, t. This transformation is linear in the control points $\mathbf{B}$. That is, if $\mathbf{B}$ is a vector containing all of the control points for all of the patches and $\mathbf{B} \rightarrow \mathbf{P}(\mathbf{B})$ is the transformation of a piecewise Bezier polygon into a simple polygon using a fine uniform sampling in t, then there is a matrix H such that $\mathbf{P} = H\mathbf{B}$. Note that the resultant simple polygon will not be sampled uniformly in arclength. A uniform arclength sampling (or any sampling strategy that depends on the control points) will, in general, be a nonlinear transformation of the control points, $\mathbf{B}$.

Since using a piecewise Bezier polygon only introduces a linear transformation into the data-flow diagram, any non-linearity of the final image model (as a function of the control points of the piecewise Bezier polygon) can come

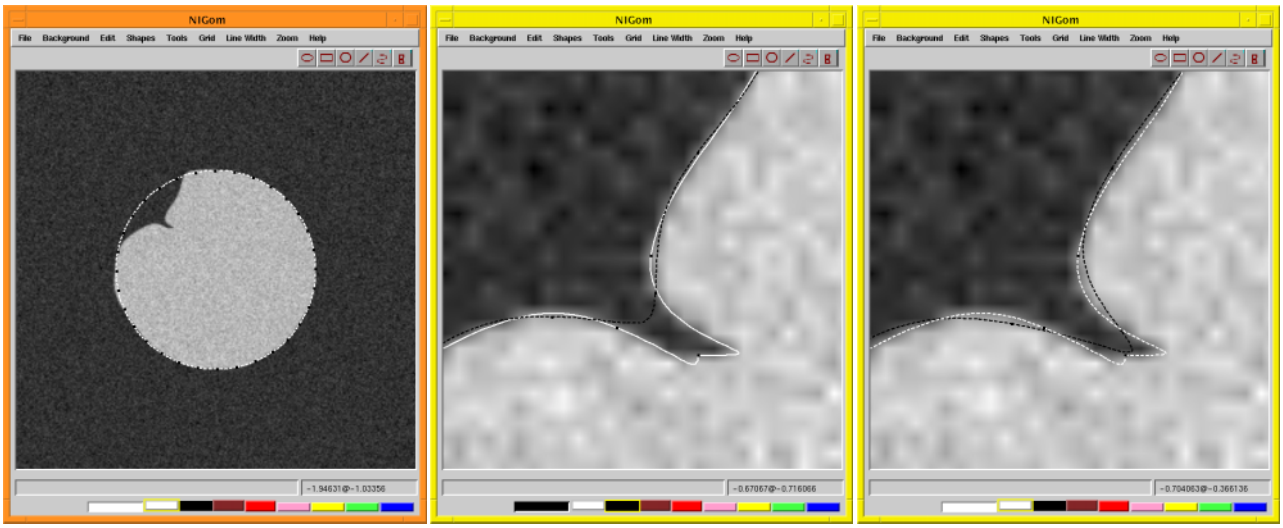

Figure 5. A simple Bayesian segmentation problem using a geometric model that is 32 piecewise cubic patches with C^2 continuity at every patch boundary. Image data and original configuration of geometric model shown on left. Local minimum attained by naive application of conjugate gradient to geometric model shown in middle (in black) with overlaying grayscale zoomed on interesting "defect" in image data. Local minimum is due to high value of a noisy image pixel in defect that stops model from penetrating in. Global minimum attained using multi-scale approach shown on right (in black). The first-derivative discontinuous geometric model used to generate the noise-free part of image data is shown in the middle and right (in white). Noise is additive Gaussian with σ equal to 0.1 times the interior value of the geometric model.

only from the non-linearity of the transformation of the simple polygon into an image, treated in Section 4. If perturbations to the piecewise Bezier's control points are restricted in size so that the resultant simple polygon's vertices are perturbed by only a small amount relative to the blurring function's σ, then the entire forward model $\mathbf{I}(\mathbf{P}(\mathbf{B}))$ is approximately linear in the perturbations to $\mathbf{B}$. Therefore, the quadratic nature of the sub-optimization problems will not be compromised and we should expect convergence to the global minimum. In Fig. 5, we see a comparison of the multi-scale strategy. using cubic B-splines with 2nd-order differential continuity across patch boundaries, versus a traditional strategy for a simple Bayesian segmentation problem. The approach outlined in this article is thus applicable to Bayesian segmentation using snake-like models of the geometry, and may be a desirable alternative to snakes when prior information about the image values inside and outside the boundary is available.

The segmentation problem displayed in Fig. 5 involves the simplest measurement model that one can have, since the model for the measurements of the true image is just the predicted image plus noise. In a real image segmentation problem, though, raw data will have been manipulated into "image data" (which we are calling the measurements for the segmentation problem) by directly inverting the measurement system, so that the segmentation algorithm operates on a previous estimator of the image values.

Other Bayesian estimation problems operate on data that are not so transparently related to the "measurements". For example, a common image restoration problem is de-blurring. In this case, the measurement system is usually modelled as a convolution of the image model $\mathbf{I}$ with some filter, denoted by the transform $\mathbf{I} \rightarrow \mathbf{H}(\mathbf{I})$. The predicted image, as a function of, say a piecewise Bezier geometric model, can be written as $\mathbf{H}(\mathbf{I}(\mathbf{P}(\mathbf{B})))$. Once again, if $\mathbf{I}(\mathbf{P})$ has been linearized over some size of perturbations to $\mathbf{P}$, and only perturbations to $\mathbf{B}$ that lead to smaller perturbations in $\mathbf{P}$ than those required by this linearization, we have a sub-optimization problem that is approximately quadratic in the geometric model $\mathbf{B}$. We should again expect global convergence using multi-scale analysis.

Another example of the use of multi-scale analysis is Bayesian image reconstruction. In this case, the measurements are related to the image model (which can be a slice of a 3D volumetric object model, e.g.) through line integrals. If the line integral through the image are computed at every angle and every displacement, the transform is called the Radon transform.[26] If line integrals at only a small number of angles is computed, then the problem of

reconstructing the image from these data is called limited-view tomography.[5,22,23] If the image is parameterized as circularly symmetric about some center point (so a 1D function $\rho(r)$ and the center of symmetry completely describe the image) and the set of line integrals at a complete set of displacements are computed, the transform is called the Abel transform.[27,28] All of these transforms are linear, and all have been implemented in the BIE. See these proceedings[5] for an application of the multi-scale approach to the limited-view tomography problem.

The real data that the BIE was developed to analyze are flash radiographs of dynamically evolving, cylindrically-symmetric, 3D objects.[28] The most important component of the measurement system is the "3D" Abel transform. That is, given a single-material object with a 3D volumetric density profile, $\rho(x, y, z) = \rho'(\sqrt{x^2 + y^2}, z)$, compute the Abel transform of $\rho'(r, z)$ for every z. This results in the intermediate image model, $\rho L(x, z)$, that is equal to line integrals of the object's density. A photon of energy E, produced by the radiographic machine, which travels from the source to the point (x, z) on the detector, will penetrate the object with probability $\exp^{-\frac{\mu}{\rho}(E)(\rho L)}$, where $\frac{\mu}{\rho}(E)$ is the material's mass attenuation coefficient at energy E. Since the machine actually produces a flux of photons at a wide variety of energies, the number of photons that arrive at detector point (x, z) will be a weighted integral over energy of the probability defined above, to produce another intermediate image model $b(x, z)$. Since the photons at the source do not originate from a single point, there will be a blur of $b(x, z)$ (to produce $c(x, z)$) to model this. When photons hit the detector, they will scatter before depositing energy. This process can be modelled by yet another blur of the image model $c(x, z)$, yielding image model $d(x, z)$. Finally, the detector may be preferentially sensitive to photons of certain energies and the response may saturate if too many photons impact the detector. These transforms can be modelled as nonlinear point transforms of the image $d(x, z)$. There are other effects that need to be modelled as part of the measurement process, e.g. registration, scattered radiation, intensity of the source as a function of angle, collimation, photon-beam divergence, etc. Fig. 6 shows the canvas of the BIE for a real data problem. This data analysis problem is a complex mixture of segmentation (determine the boundaries of the material), restoration (determine the density distribution with a resolution that is much better than that allowed by the source and detector blur), reconstruction (determine a density distribution from line integrals), and registration (errors in calibration of detector location, rotation and shift relative to the object and source).

6. CONCLUSIONS

We have outlined a multi-scale approach to global optimization in this article that is a general-purpose strategy for problems utilizing geometric models of image data. We have shown that the approach can be interpreted as linearizing a highly nonlinear forward model of the image data. Since linear models result in quadratic log likelihoods when the noise is additive Gaussian, the optimization problem at any single scale is close to quadratic if the adjustments to the geometric model are small in relation to the blur. This allows us to be confident that a unique solution exists at any single scale in an appropriate region around the configuration of the current initial geometry for that scale. Scale-space continuation is invoked to argue that the unique sequence converges to the unique global minimum, under mild conditions on the closeness of the initial guess, the size of the initial scale, and the rate at which the scale is decreased. Some user input is required at the start of the global optimization session to define an initial configuration that is close to the global minimum and to provide a measure of that closeness. The BIE facilitates this interaction with linked graphical and textual geometry editors. The approach proposed in this article is different from a similar approach, used in the context of image segmentation, called snakes. In our approach, it is the geometric model of the image that is blurred with successively fine resolution, rather than, as with snakes, the energy function derived from the image data. This difference allows us to use our approach as a general-purpose tool for solving Bayesian estimation problems that utilize explicit models of geometric content in image data, including segmentation, reconstruction, and restoration.

ACKNOWLEDGEMENTS

This work was supported by the U.S. Dept. of Energy under Contract W-7405-ENG-36.

REFERENCES

1. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models", *Inter. J. Comp. Vision* **1**, pp. 321-331, 1988.

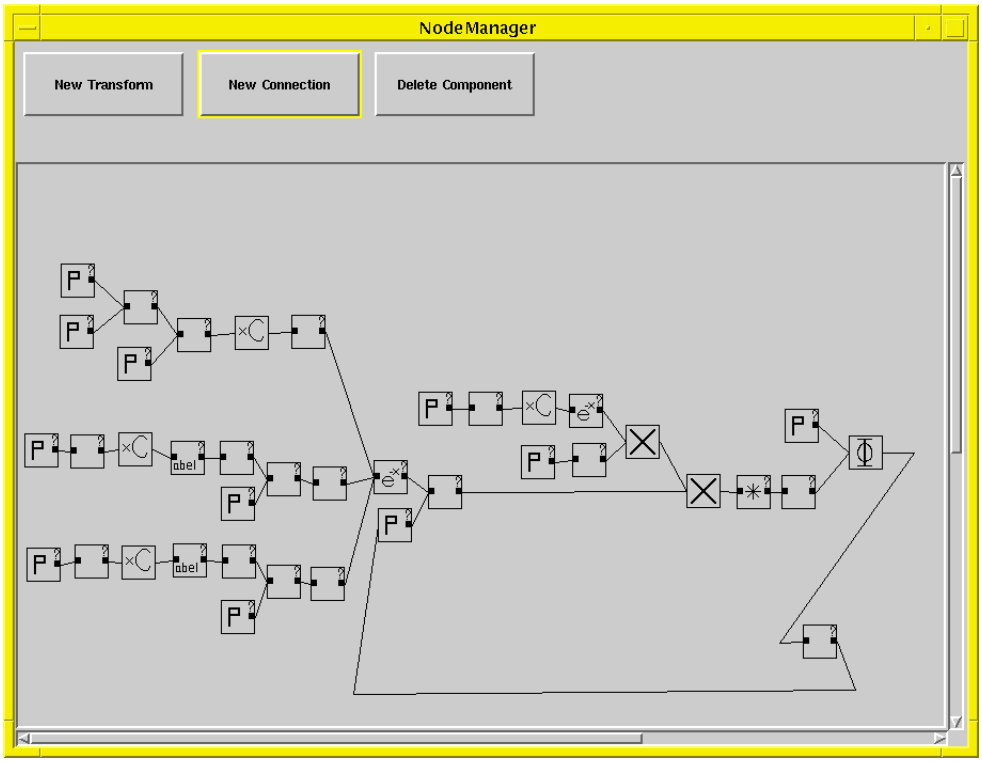

Figure 6. The canvas for the BIE when used for analysis of a real static radiograph of a cylindrically-symmetric object with 2 materials (lower left chains) and 2 collimators (upper left and upper middle). Also modelled are the intensity profile of the beam (upper middle), registration of the detector, radiographic magnification, system blur, and scatter field (all in the middle chain). An energy-dependent exponential attenuation law is used as discussed in the text.

2. U. Grenander, M. I. Miller, and G. E. Christensen, "Deformable anatomical data bases using global shape models: a position paper for the 1992 EIHB Workshop", *Proc. of the Cooperative Working Group on Whole Body 3-D Electronic Imaging of the Human Body*, pp. 826-830, 1992.
3. K. M. Hanson, "Bayesian reconstruction based on flexible prior models," *J. Opt. Soc. Amer.* **A10**, pp. 997–1004, 1993.
4. D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-8**, pp. 413-424, 1986.
5. K. M. Hanson, R. Bilisoly, and G. S. Cunningham, "Kinky tomographic reconstruction," *Proc. SPIE* **2710**, these proceedings, 1996.
6. D. Rueckert and P. Burger, "A multiscale approach to contour fitting for MR images," *Proc. SPIE* **2710**, these proceedings, 1996.
7. K. M. Hanson and G. S. Cunningham, "Exploring the reliability of Bayesian reconstructions," *Proc. SPIE* **2434**, pp. 416-423, 1995.
8. J. Besag, P. Green, D. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems, *Statistical Science* **10**, pp. 3-66, 1995.
9. S. Geman and D. Geman, "Stochastic relaxation, Gibb's distributions, and the Bayesian restoration of images", *IEEE Trans. Pattern Anal. Machine Intell.* **PAMI-6,** pp. 721-741, 1984.
10. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, London, 1995.
11. D. J. C. MacKay, *Neural Computation* **4**, pp. 415-447, 1992.
12. S. F. Gull, "Developments in maximum-entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., pp. 53–71, Kluwer Academic, 1989.

13. S. F. Gull, "Bayesian inductive inference and maximum entropy," in *Maximum Entropy and Bayesian Methods in Science and Engineering (Vol. 1)*, G. J. Erickson and C. R. Smith, eds., pp. 53–74, Kluwer Academic, 1989.
14. K. M. Hanson and G. S. Cunningham, "A computational approach to Bayesian inference," in *Interface: Computational Science and Statistics*, M. Meyer, ed., 1995 (to be published).
15. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An object-oriented optimization system," *Proc. IEEE Int. Conf. Image Processing* **III**, pp. 826–830, 1994.
16. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An interactive tool for Bayesian inference", *Review of Progress in Quantitative Nondestructive Evaluation* **21**, pp. 747-754, 1994.
17. G. S. Cunningham, K. M. Hanson, G. R. Jennings, Jr., and D. R. Wolf, "An object-oriented implementation of a graphical-programming system", *Proc. SPIE* **2167**, pp. 914-923, 1994.
18. A. Witkin, "Scale space filtering," *Proc. Eighth Int. Joint Conf. Artif. Intell.,* pp. 1019-1021, 1983.
19. A. Witkin, D. Terzopoulos, and M. Kass, "Signal matching through scale space," *Proc. Amer. Assoc. Artif. Intell.,* pp. 714-719, 1986.
20. J. A. Schnabel and S. R. Arridge, "Multiscale shape description of MR brain images using active contour models," *Proc. SPIE* **2710**, these proceedings, 1996.
21. B. K. Jang, "Multiscale shape analysis for computed radiographic images," *Proc. SPIE* **2710**, these proceedings, 1996.
22. K. M. Hanson, G. S. Cunningham, G. R. Jennings, Jr., and D. R. Wolf, "Tomographic reconstruction based on flexible geometric models," *Proc. IEEE Int. Conf. Image Processing* **II**, pp. 145–147, 1994.
23. K. M. Hanson, "Reconstruction based on flexible prior models," *Proc. SPIE* **1652**, pp. 183–191, 1992.
24. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science* **220**, pp. 671-680, 1983.
25. G. Farin, *Curves and Surfaces for CAGD - a practical guide,* third edition, Academic Press, San Diego, 1983.
26. A. Rosenfeld and A. C. Kak, *Digital Picture Processing* **1**, second edition, p. 355,Academic Press, Orlando, 1982.
27. N. H. Abel, "Résolution d'un Problème de Mecanique", *J. Reine u. Angew. Math.* **1**, pp. 153-157, 1826.
28. K. M. Hanson, "Tomographic Reconstruction of Axially Symmetric Objects from a Single Radiograph", *Proc. 16th Inter. Cong. on High Speed Photography and Photonics*, *SPIE* **491**, pp. 180-187, 1984.