

Edge-Preserving Tomographic Reconstruction with Nonlocal Regularization

Daniel F. Yu and Jeffrey A. Fessler*, *Senior Member, IEEE*

Abstract—Tomographic image reconstruction using statistical methods can provide more accurate system modeling, statistical models, and physical constraints than the conventional filtered backprojection (FBP) method. Because of the ill posedness of the reconstruction problem, a roughness penalty is often imposed on the solution to control noise. To avoid smoothing of edges, which are important image attributes, various edge-preserving regularization methods have been proposed. Most of these schemes rely on information from *local* neighborhoods to determine the presence of edges. In this paper, we propose a cost function that incorporates *nonlocal* boundary information into the regularization method. We use an alternating minimization algorithm with deterministic annealing to minimize the proposed cost function, jointly estimating region boundaries and object pixel values. We apply variational techniques implemented using level-sets methods to update the boundary estimates; then, using the most recent boundary estimate, we minimize a space-variant quadratic cost function to update the image estimate. For the positron emission tomography transmission reconstruction application, we compare the bias-variance tradeoff of this method with that of a “conventional” penalized-likelihood algorithm with local Huber roughness penalty.

Index Terms—Edge-preserving reconstruction, level sets, positron emission tomography (PET), region-based, transmission tomography.

I. INTRODUCTION

THE PROBLEM of reconstructing an unknown image f from a measurement vector y is usually ill posed [1], [2]. The direct model alone rarely determines a satisfactory solution. If we find the maximum-likelihood estimate of the image by maximizing the log-likelihood function $L(f; y)$, then the resulting image is very noisy. Thus, it is necessary to regularize the solution by imposing *a priori* assumptions. One simple regularization method supposes that images are globally smooth, and enforces a roughness penalty on the solution by adding a quadratic function to the negative log-likelihood. Such a “penalized” likelihood cost function has the following form:

$$J(f) = -L(f; y) + \beta V(f) \quad (1)$$

Manuscript received December 8, 1999; revised November 28, 2001. This work was supported in part by the National Science Foundation (NSF) under Grant BES-9982349 and in part by the National Institutes of Health (NIH) under Grant CA-60711. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was V. Johnson. *Asterisk indicates corresponding author.*

D. F. Yu was with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109-2122 USA. He is now with the Institute of Applied Mathematics Department, University of British Columbia, Vancouver, BC V6T 1Z1, Canada.

*J. A. Fessler is with the University of Michigan, 4240 EECS Bldg., Ann Arbor, MI 48109-2122 USA (e-mail: fessler@umich.edu).

Publisher Item Identifier S 0278-0062(02)02934-8.

where

$$V(f) = \int |\nabla f(x)|^2 dx \quad (2)$$

is a measure of image roughness.¹ The image estimate is obtained by

$$\hat{f} = \arg \min_f J(f)$$

where often the minimization with regard to f is restricted to nonnegative values. The cost function in (1) is often unsatisfactory, since many images are not globally smooth. They have region boundaries across which the image values can vary rapidly. The quadratic regularization in (2) causes edges to become blurred. In many images, small differences between neighboring pixels are often due to noise, while large differences are due to the presence of edges. This assumption has formed the basis for many edge-preserving regularization schemes proposed in the literature.

Most edge-preserving regularization methods, including many line-site models, e.g., [5]–[9], rely on information from a *local* neighborhood to determine the presence of edges, i.e., the penalty assigned to each pixel or clique of pixels depends solely on pixel values within a small *fixed* neighborhood. (We return to this point near the end of this section.) One such scheme is to replace the quadratic penalty function in (2) with a nonquadratic function ψ that increases less rapidly than the quadratic function for sufficiently large arguments, such as the Huber function [10], [11]

$$V(f) = \int \psi(|\nabla f(x)|) dx$$

or

$$V(f) = \int \sum_{i=1}^n \psi \left(\left| \frac{df}{dx_i} \right| \right) dx \quad (3)$$

where

$$\psi(t) = \begin{cases} t^2/2, & |t| \leq \delta \\ \delta|t| - \delta^2/2, & |t| > \delta. \end{cases}$$

This function increases linearly, instead of quadratically, for arguments larger than δ . Thus, the cost function penalizes large

¹Following the convention in the literature on partial differential equation (PDE)-based image analysis, we present a nondiscretized formulation in which f belongs to the class of differentiable functions with derivative that is absolutely integrable over some domain $\Omega \subset \mathbb{R}^n$. In the above integral, dx is a shorthand for $dx_1 \cdots dx_n$, where n is the dimension of the “image;” typically $n = 2$.

In practice, one implements a discretized formulation, for example $V(f) = \sum_j \sum_k w_{jk} (f_j - f_k)^2$, where w_{jk} is nonzero only for neighboring pixels. In problems with inhomogeneous noise (such as Poisson measurements), one may need to adjust the w_{jk} s to control resolution properties [3], [4].

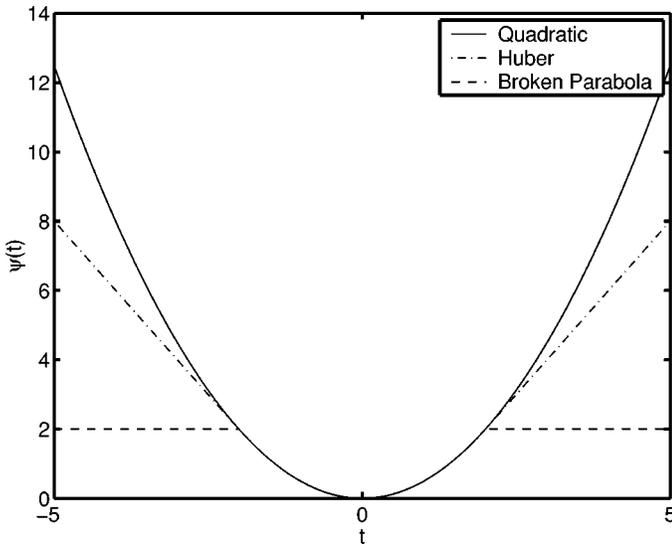


Fig. 1. A comparison of quadratic, Huber, and broken parabola penalty functions.

differences between neighboring pixels less severely than the quadratic penalty, while maintaining the same level of penalty for small differences. This property permits sharper edges in the reconstructed image.

The Huber penalty is a convex function of image pixel values. One could instead use a nonconvex penalty function, such as the “broken parabola” function (e.g., [12]–[16])

$$\psi(t) = \min\{t^2, \delta^2\}.$$

This function is nondifferentiable, thereby precluding gradient-based descent methods. Instead, deterministic annealing algorithms are usually applied, e.g., [16], where one sequentially minimizes a series of cost functions approaching the original cost in the limit. This method also uses local information only. Fig. 1 compares the quadratic, Huber, and broken parabola function.

Both of the above edge-preserving methods modify the local penalty function ψ so as to penalize large differences between neighboring pixels less than the quadratic penalty does. These methods implicitly use local information to “detect” the presence of an edge locally. This local view is formalized by considering a hierarchical Bayesian model in which a Gauss–Markov prior for pixels is conditioned on a noninteracting line-site model [17]–[20]. Even when line-site interactions are included in such models to encourage boundary continuity, typically only small cliques are used [5], [21]–[23] so such approaches are still inherently local. In the context of blurred image restoration, comparatively large line-site neighborhood sizes that match the size of the point-spread function (PSF) of the imaging system have been proposed [7], [8]. (How to apply that principle in the tomography problem of interest here is unclear since each measurement sees long strips traversing the entire object.) None of these line-site models address *global* connectivity or continuity and, thus, are inherently local. One of the few previous methods to capture global properties is a region-based Bayesian prior that has been applied successfully in tomography [6], [9]. That method uses discrete region identifiers (motivated by image segmentation

problems) and assigns costs that prohibit disconnected regions, encourage regularly shaped regions, and discourage having too many regions. Some of these costs involve the entire image and are, therefore, global. The Bayesian formalism permits the exploration of estimate uncertainty, but using discrete region labels is challenging for computing point estimates. That method and our proposed approach share the property that the number of regions need not be specified *a priori*. Our approach is boundary based rather than region based; boundaries are continuous-valued, so simple gradient-descent methods are available for computing point estimates (at local minima of the cost function).

This paper describes a method for including *nonlocal* information, specifically, boundary information, into the regularization method. The proposed penalty adjusts how much we penalize differences between pixels based on the distance of the pixels to the nearest boundary curve. This approach involves regions only implicitly. However, since both the estimates for the boundary curves and the distance of a pixel to the boundary curves depend on the entire image, our penalty is nonlocal. We hope to achieve better results than purely local penalties under certain cases. In the specific case of emission computed tomography, accurate attenuation correction is usually necessary for a quantitative emission reconstruction [24]. Accurate attenuation correction requires an accurate map of attenuation coefficients. A positron emission tomography (PET) attenuation map consists of a small number of regions, i.e., lungs, spine, body tissue, etc. The attenuation coefficients within each region are fairly uniform, but they vary a great deal between neighboring regions and the transition between regions can be fairly rapid, e.g., across a few pixels. A regularization method that incorporates this additional prior information, such as the one we propose, should be able to outperform a purely *local* regularization method. Section II describes our new algorithm; Section III compares the proposed algorithm to a “conventional” statistical algorithm and to filtered backprojection (FBP); Section IV summarizes our results and outlines possible future work.

II. METHOD

A. New Cost Function

Our development of the new cost function was motivated by the specific application of PET and SPECT transmission tomography, but its use is not restricted to attenuation map reconstructions. We assume that the actual object to be reconstructed is everywhere differentiable. We also assume that the object consists of regions that are piecewise smooth (meaning almost uniform) everywhere except near the region boundaries where the object intensity changes rapidly (but differentially) to values in its neighboring region(s). Thus, an edge-preserving penalty function should penalize local gradients that are within each region more than local gradients that are very close to the boundary curves. Furthermore, we assume that the boundary separating the regions consists of smooth curves. Fig. 2 shows an example object (representing a thorax attenuation map at 511 KeV) and one of its profiles. This object is piecewise smooth, but not piecewise constant, due to variations in lung density.

Like (1), the cost function we propose also consists of a data-fit term and a penalty term. However, our penalty considers not only the image values but also the characteristics of region

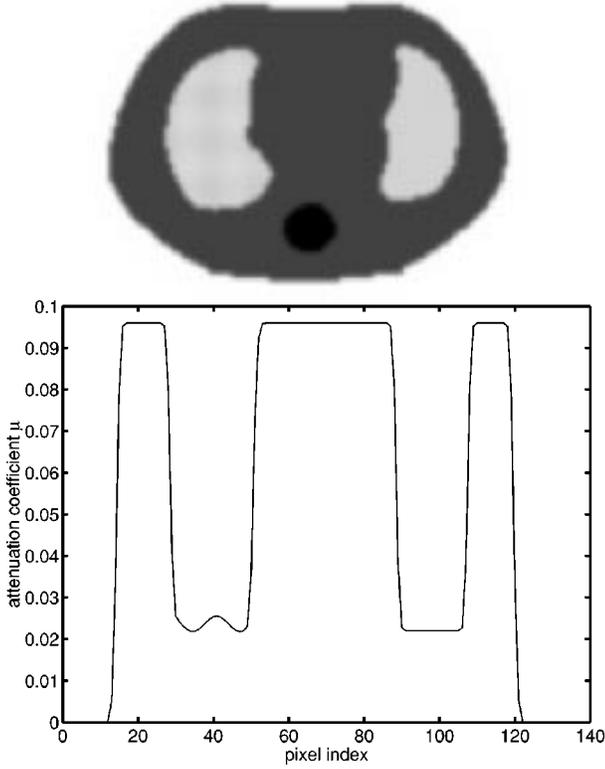


Fig. 2. Nonpiecewise-constant phantom and central horizontal profile (row 65).

boundaries within the image. We present our model in two dimensions, but extension to three dimensions is straightforward [25], [26]. Let f denote the object, \mathcal{G} denote a family of regular closed curves (i.e., no self-intersections, corners, or cusps), $\Gamma \in \mathcal{G}$ denote the boundary curve(s), and Ω denote the domain of the image. We do not require Γ to have a single connected component, i.e., Γ can have several connected components, one being the boundary curve between the lung and the soft tissues, while another one being the boundary curve between the body and air, etc. The number of boundaries need not be known *a priori*.

We propose the following cost function of the object f and the boundary curves Γ [27]

$$J(f, \Gamma) = -L(f; y) + V(f, \Gamma) \quad (4)$$

$$V(f, \Gamma) = \beta J_2(f, \Gamma) + \mu \int_{\Gamma} ds \quad (5)$$

$$J_2(f, \Gamma) = \int_{\Omega} h(d(x, \Gamma)) |\nabla f(x)|^2 dx. \quad (6)$$

The first term $-L(f; y)$ is the negative log-likelihood that measures the “faithfulness” of the reconstructed object to the measured data. The term $\int_{\Gamma} ds$ penalizes the length of the boundary curves, encouraging smooth boundary curves, and discouraging large numbers of boundary curves. The term $J_2(f, \Gamma)$ penalizes local gradients inside each region more than local gradients close to the boundary curves; $d: \Omega \times \mathcal{G} \mapsto \mathbb{R}$ is the signed distance from x to Γ , i.e., we require d to have opposite signs on the opposite sides of a segment of Γ , so d is differentiable across the boundary curve. We also require d to be continuous inside Ω . There is no intuitive way of defining

“outside” or “inside” when curves enclose other curves, but we can define the outward normal \vec{N} to be the direction in which d is increasing. The roughness weighting function $h: \mathbb{R} \mapsto [0, 1]$ maps small arguments to values near zero and larger arguments to values near unity [see (18)]. For simplicity, we use only h s that belong to $C^\infty(\mathbb{R})$, i.e., differentiable arbitrarily many times on the entire real line.

The purpose of the J_2 term is to penalize image roughness, like (3). However, in our J_2 , how much the local gradient at a specific location is penalized is weighted by the distance of this location to the boundary curves, and since this distance depends on the entire curve, the proposed penalty is “nonlocal.” Fig. 3 shows an example h . Fig. 3(a) shows a one-dimensional (1-D) object. Fig. 3(b) shows the signed distance to the boundary points; in this case, the boundary points are at $-4, 0, 2$, and 4 ; we chose the sign of the distance so that d is continuous. Fig. 3(c) illustrates the type of h function we may want to use; the value of $h(d(\cdot))$ is unity well inside each region, but smoothly decreases near the boundary curves. This approach allows larger gradients in the reconstructed object close to the boundary curves. In two dimensions (or three dimensions), the value of $h(d(\cdot))$ at every point is determined by the distance between the point and the boundary curves (or surfaces in three dimensions).

As in all edge-preserving regularization methods, one must choose carefully the weighting parameters, β , μ , and the function h to avoid over-smoothing of the reconstructed image or the boundary curves. The curve length term in (5), which acts to keep the boundary curves smooth, will favor shorter curves over longer curves although the region roughness penalty J_2 should keep this force in check. Nevertheless, the associated parameter μ should always be small to avoid excess shrinking of the boundary curves.

B. Alternating Minimization Scheme

We use alternating minimization to jointly minimize the cost function given in (4) over f and Γ . We first hold f constant and minimize J with regard to Γ . Then, using the most recent estimate of Γ , we minimize J with regard to f . We alternate between these two steps until convergence.

1) *Boundary Update:* When f is fixed, the second and third terms of J depend on Γ . We must minimize the following cost function:

$$J_f(\Gamma) = \beta J_2(f, \Gamma) + \mu \int_{\Gamma} ds \quad (7)$$

$$\Gamma^{n+1} = \arg \min_{\Gamma} J_{f^n}(\Gamma) \quad (8)$$

where J_2 was defined in (6) and J_{f^n} denotes the cost function J with f fixed at f^n . As in standard PDE-based image analysis, we perform steepest descent with respect to Γ [28]. Each point $v = (x_1, x_2)$ on the boundary curve Γ evolves according to the following differential equation:

$$\frac{dv}{dt} = - \frac{\delta J_f(\Gamma)}{\delta v} \quad (9)$$

where the right-hand side is the negative functional derivative of the cost function. Finding the functional derivatives of J_2 analytically is nontrivial, so the natural starting point is to evaluate the functional derivatives numerically. (We show that this leads to an analytical formula as well.) The functional derivative of J_2

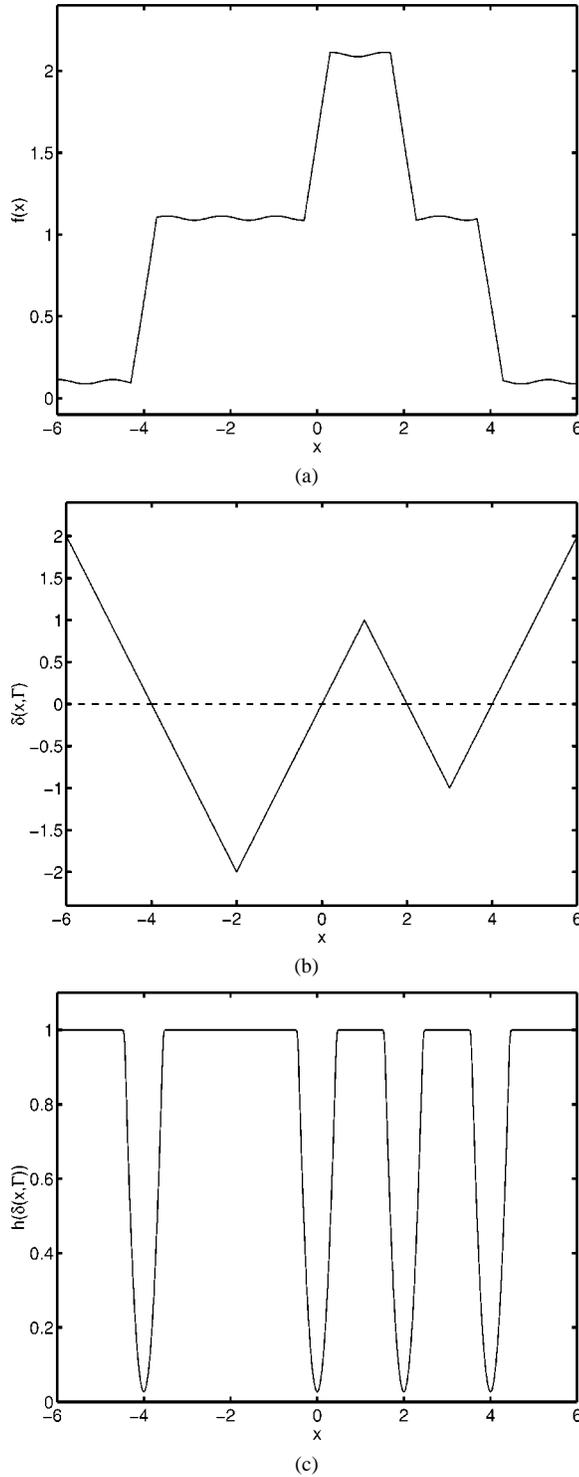


Fig. 3. (a) Example of a 1-D object. (b) Signed distance to the boundary. (c) One possible h function.

must point in the normal direction of the curve, since movement in the tangential direction would not change the curve. We can use a scheme similar to the central difference method to evaluate local derivatives. (Central differences are usually accurate to a higher order than one-sided differences.) Let $\gamma: \mathcal{S} \mapsto \mathbb{R}^2$ be a parameterization of Γ , where \mathcal{S} is a subset of \mathbb{R} . For a given point p_0 on the curves Γ , we define a function $z: \mathbb{R} \mapsto \mathbb{R}$ that is zero except in the neighborhood of p_0 , so that $\gamma + z\vec{N}$ differs from Γ only in the normal direction. (We can imagine some force being

exerted on the curve; this force is nonzero only in the neighborhood of p_0 ; briefly exerting this force in the outward normal direction of the curve at p_0 causes a small perturbation of the curve at p_0 in the normal direction). Using this idea, we approximate the functional derivative of $J_2(\Gamma)$ at $v = p_0$ as follows:

$$\frac{\delta J_2}{\delta v} \approx \frac{\widehat{\delta J_2}}{\delta v} \triangleq \frac{1}{2} \frac{J_2(\gamma + z\vec{N}) - J_2(\gamma - z\vec{N})}{\Delta\sigma} \quad (10)$$

where \triangleq denotes “defined to be” and $\Delta\sigma$ is the area lying between the curve Γ and the perturbed curve $\gamma + z\vec{N}$ [29].

For the second term in $J_f(\Gamma)$, the direction in which the curve length decreases most rapidly is when [30]

$$\frac{\partial \gamma}{\partial t} = -\kappa \vec{N} \quad (11)$$

i.e., the speed of the evolution at any point is the curvature of the boundary curve at that point, and the curve evolves in the inward normal direction. For implementation, we use the level sets method [30]–[32] because of its simplicity. In the level sets method, instead of evolving the curve Γ itself, we embed Γ as the zero level set of a smooth function $\Phi: \mathbb{R}^2 \mapsto \mathbb{R}$ such that $\Gamma = \{(x_1, x_2) \in \mathbb{R}^2: \Phi(x_1, x_2) = 0\}$, and evolve Φ so that Γ evolves according to (9). As discussed in [31], this method has several advantages over directly evolving Γ , e.g., periodically resampling the curve becomes unnecessary, and, more importantly for us, topological changes of the curve occur seamlessly. We use the latter advantage to allow merging or splitting of boundary curves. Combining (9)–(11), the curves evolve according to the following:

$$\frac{dv}{dt} = -\left(\mu\kappa + \beta \frac{\widehat{\delta J_2}}{\delta v}\right) \vec{N}. \quad (12)$$

Evolving the curve via (12) yields a curve estimate $\hat{\Gamma}^{n+1}$ that approximately minimizes J_{f^n} ; we call this step the “boundary estimation” step. (See Section II-D for details about initialization.)

2) *Image Update:* For the second stage of the minimization, we hold Γ fixed at its previous estimate Γ^n and minimize with regard to f . When Γ is held fixed, the relevant terms in the cost function (4) are the following:

$$J_{\Gamma}(f) = -L(f; y) + \beta J_2(f, \Gamma) \quad (13)$$

where J_2 was defined in (6). We minimize $J_{\Gamma}(f)$ with regard to f as follows:

$$f^{n+1} = \arg \min_f J_{\Gamma^n}(f). \quad (14)$$

When updating the boundary curves using (8), the h function in J_2 pushes the boundary curves toward image locations where the gradient is large; when updating the object f using (14), the h function imposes a space-varying weighting of the penalty on local gradients. This weighting depends on the signed distance from each pixel to the nearest estimated boundary curve. Every term in (13) is quadratic in f , except possibly the log-likelihood term, which involves logarithms in the case of Poisson measurements. Thus, (14) is a standard penalized-likelihood image reconstruction problem, and we can minimize J_{Γ} over f using methods such as the conjugate gradient method (if quadratic) [33], [34] or the paraboloid surrogates/coordinate descent method (if not) [35].

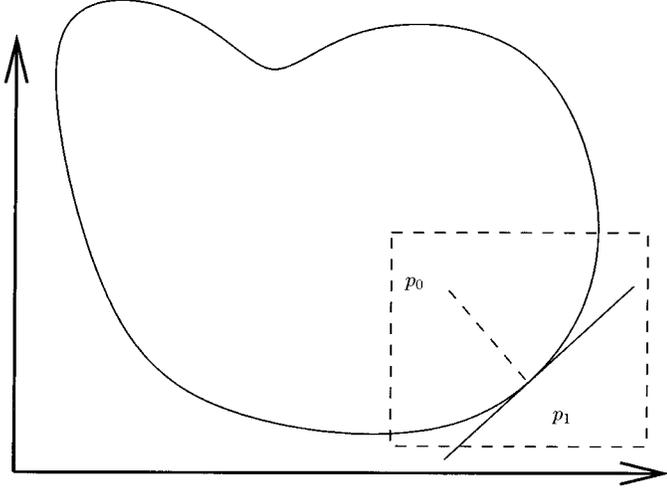


Fig. 4. p_1 is the closest point to p_0 on the curve.

We iteratively alternate between the two steps (8) and (14). Both these two steps will, under ideal circumstances,² monotonically decrease the cost as defined in (4). In addition, the cost is bounded below, so the algorithm will presumably converge toward a local minimum.

C. Discretization and Implementation

We discretize the image f using the usual square grid. For simplicity, we discretize the level sets, which embed the boundary curves, using the same square sampling grid as the image. However, the sample spacings of the boundary curves (when extracted from the zero contours of the level set) may be finer than the image pixel spacing. The boundary curves evolve according to (12). The first term causes image-independent curve smoothing; its level-sets implementation is described in [31]. In addition to the smoothing term, the evolution of the boundary curves is influenced by the functional derivative of J_2 .

To implement (12), we observe that for a point p_1 on the boundary curve to be the closest to a point p_0 in the image, the line connecting p_0 and p_1 must be perpendicular to the tangent line of the boundary curve at p_1 (see Fig. 4). Thus, if we make a small enough perturbation of the boundary curve in a neighborhood of p_1 , the only points in the image that are possibly affected by this change of the boundary curve (in terms of their distances to the curve) will lie in a narrow band perpendicular to the boundary curve (see Fig. 5). Since this band can be made arbitrarily narrow by making the boundary curve perturbation small enough, we can make the approximation that the image values remain constant in the lateral direction of the band and evaluate the functional derivative of J_2 using equally spaced points on the line perpendicular to the boundary curve at p_1 ; we use bi-cubic interpolation from the neighboring points [33] where ∇f is not available. At every time step of the evolution of the boundary curve, we also keep a record of which point on the boundary curve is closest to each image point. When evaluating the functional derivative of J_2 at a point p_1 , we make the approximation that a very small movement in the curve near

²Under realistic circumstances, where dt is taken to be finite, the minimization of (7) according to a discretized version of (9) may not be exactly monotonic. Such effects are inevitable when continuous formulations are discretized.

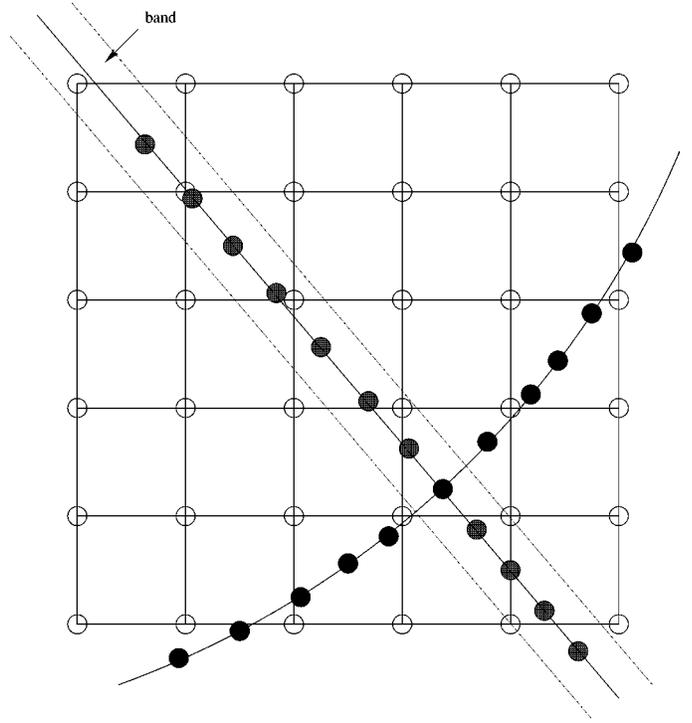


Fig. 5. White dots denote image points; black dots denote boundary points; J_2 is evaluated on points represented by shaded dots.

p_1 will not cause any image point that was not closest to p_1 previously to become closest to p_1 after the movement of the boundary curve. Hence we only need to evaluate the change in J_2 on those points that are already closest to p_1 . For a $n \times n$ image containing a boundary curve with m samples, evaluating the functional derivative of J_2 is an $O(mn)$ operation.

The above ideas suggest the following functional derivative of J_2 (see the Appendix for an explanation):

$$\frac{\delta J_2}{\delta v} = \left(\int_l h'(r) |\nabla f(l(r))|^2 I_{(l(r), \Gamma)}(r) dr \right) \vec{N} \quad (15)$$

where $l(r)$ denotes the line perpendicular to the boundary curve Γ at the point v , parameterized by r which increases in the direction of \vec{N} , h' denotes the derivative of h with regard to r , and $I_{(l(r), \Gamma)}$ is an indicator function

$$I_{(l(r), \Gamma)}(r) = \begin{cases} 1, & v \text{ is the closest point in } \Gamma \text{ to } l(r) \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

D. Initialization

To form an initial image estimate f^0 , we perform conventional penalized-likelihood image reconstruction using a local penalty such as the space-invariant quadratic penalty as described in (1) and (2). Initialization of the boundary estimate Γ^0 requires greater care. Performing steepest descent with respect to (7) may not push the estimated curve toward the true boundary if the initial curve is too far away from that boundary.

The force exerted by J_2 in (7) is nearly zero in smooth regions, and is significant only close to the actual boundary curves where local gradients are large. Fig. 6 illustrates this property in one dimension. Let f' denote the derivative of f in the x direction; let Γ_o denote the ‘‘old’’ boundary point at 0.3 and Γ_n denote the ‘‘new’’ boundary point at 0.35. Moving the

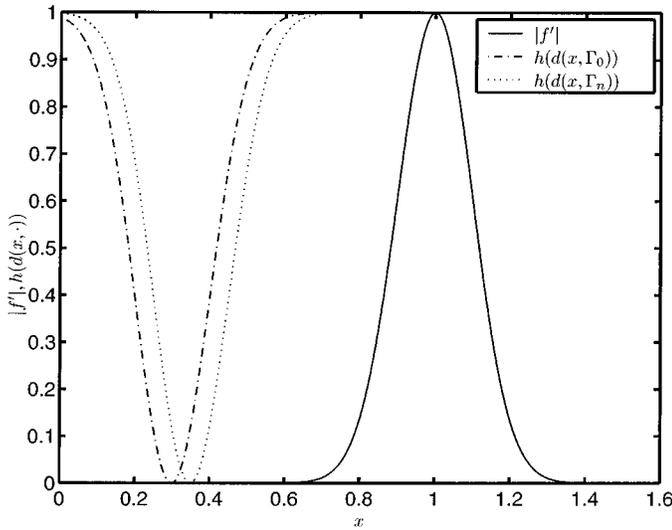


Fig. 6. Illustration that J_2 only changes slightly when the “curve” is far from the true boundary point.

boundary point Γ from 0.3 to 0.35 would change $h(d(\cdot, \Gamma))$ from $h(d(\cdot, \Gamma_0))$ to $h(d(\cdot, \Gamma_n))$, i.e., the “valley” of $h(d(\cdot, \Gamma))$ is moved from 0.3 to 0.35, but the change in the roughness penalty, i.e., $\int (h(d(x, \Gamma_n)) - h(d(x, \Gamma_0))) |f'| dx$, would be very small. Thus, evolution via (12) alone would require a fairly close initialization to the actual boundary curves. We circumvent this problem by using an initialization procedure for the boundary curves employing another force (from a global measure) that ensures that the boundary curves move even when the initial boundary curves are far from the actual boundary. During the early iterations, we add a third penalty J_3 (within-region homogeneity) to J_f as follows:

$$\tilde{J}_f(\Gamma) = \beta J_2(f, \Gamma) + \mu \int_{\Gamma} ds + \sum_{m=1}^{M(\Gamma)} \alpha J_3(f, \mathcal{R}_m(\Gamma))$$

$$J_3(f, \mathcal{R}_m) = \int_{\mathcal{R}_m} \left| f(x) - \frac{\int_{\mathcal{R}_m} f(x') dx'}{\int_{\mathcal{R}_m} dx'} \right|^2 dx \quad (17)$$

where $\mathcal{R}_m(\Gamma) \subset \Omega$ denotes the m th region defined by Γ , and $M(\Gamma)$ denotes the number of regions defined by Γ . This J_3 penalizes the difference between each pixel value and the average value of its region. This global measure exerts a force on the curve no matter how far the boundary estimate is from the image gradients. The evolution of the curve, as determined by J_3 , is essentially a competition between bordering regions. Each pixel on the curve borders two regions; each of these two regions exerts a force trying to pull the pixel inside; the boundary curves will evolve toward whichever region exerts a stronger force, as determined by (17). We gradually reduce α to zero. Eventually, we rely on J_2 alone to move the curve to a local minimum³ of $J_f(\Gamma)$.

E. Deterministic Annealing

To form an initial estimate of the image f prior to applying boundary estimation step, we perform penalized likelihood

³In practice, we run a fixed number of iterations determined sufficient for the boundary curves to converge to a local minimum. One could also stop the curve evolution when the maximum force exerted by J_2 falls below a preset threshold.

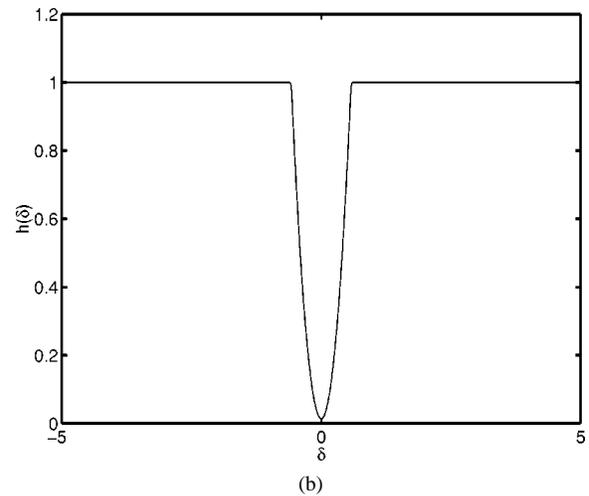
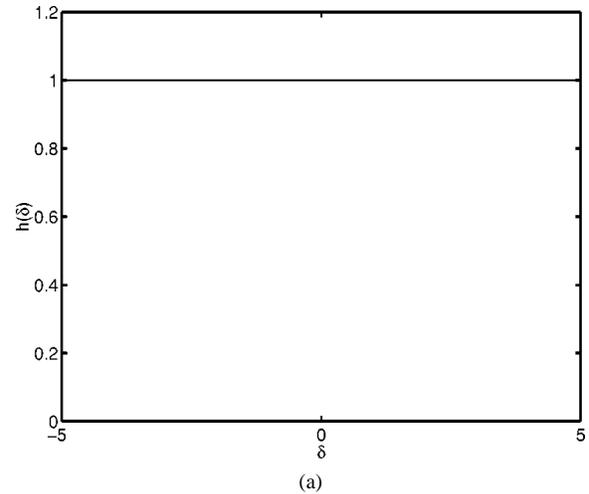


Fig. 7. (a) Implicit h function used in local regularization. (b) Edge-preserving h function as a function of signed distance to the boundary curves.

tomographic reconstruction using a standard space-invariant penalty. There is no h function as given in (6) in such a reconstruction, but we can think of h as simply being a constant, say unity [Fig. 7(a)], i.e., h is independent of the boundary curves. But for the reconstructed image to have sharp boundaries, we must assign small weights to differences in pixel pairs close to the boundary curves, e.g., as shown in Fig. 7(b). To avoid getting stuck in a poor local minimum, we must change the shape of h function gradually from the initial constant function to the desired h function, i.e., we employ deterministic annealing.

Instead of going from the implicit constant function to the desired h function in one step, we take several steps. Suppose that from empirical experience with a given category of images with similar noise levels (e.g., 3-min PET transmission scans of the thorax), we have found that the boundary curves we obtain from our initial image are within, say, five pixels. Then we assign small weights (via the h function) to all pixel pairs within a distance of five or six pixels to the detected boundary curves, and assign large weights (unity) to all other pixel pairs; thus, neighboring pixels that are more than five or six pixels away from the detected boundary curves will be coupled, while the boundary curves are allowed to evolve within those pixels between which the weights are very small. We gradually evolve the h function from the constant function toward the final desired h function,

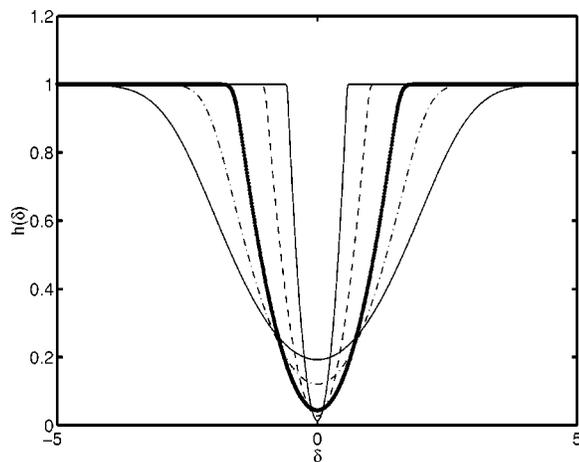


Fig. 8. The evolution of h function, where δ has units of pixels.

as shown in Fig. 8. Hopefully, the final boundary curves will eventually be a very good local minimum. The functions used in Fig. 8 are [16]

$$h(t) = -\log \frac{e^{-b\lambda t^2} + e^{-b}}{2} \quad (18)$$

with $(b, \lambda) = (2.5, 1/6), (5, 1/4), (15, 2/5), (25, 1),$ and $(50, 3)$. Evolving h^{initial} to h^{desired} usually involves four to five “cycles”; one cycle consists of two stages, i.e., the “image reconstruction” stage, and the “boundary estimate” stage.⁴ Fig. 9 shows a flow chart of the proposed algorithm.

This deterministic annealing procedure is somewhat related to graduated nonconvexity methods, e.g., [16]. However, there are also significant differences between how the “broken parabolas” are used here versus in line-site models. Here, the argument of h is the *distance* between a pixel and the nearest boundary curve point; through h , this distance controls the strength of the (quadratic) penalty between neighboring pixels [see (6)]. In contrast, in (noninteracting) line-site models, the broken parabola function is itself the penalty assigned to the difference between pixel *grayscale values* [18], [20], and there is no explicit concept of “distance” or “boundary curve.”

III. STATISTICAL RESULTS

In this section, we compare the proposed algorithm to the edge-preserving reconstruction method described in [35] which is based on local regularization, hereafter referred to as the “Huber method,” in terms of bias-variance tradeoffs. We simulated PET transmission scans of a digital phantom that resembles the human thorax (at 511 keV). The body has attenuation coefficient 0.096/cm. Within the body are two “lungs”; the left lung has nonuniform attenuation coefficients; the right lung has uniform attenuation coefficients; both lungs have average attenuation coefficient 0.022/cm. The “spine” has uniform attenuation coefficient 0.14/cm. Fig. 2 shows the phantom and a central profile. The image has 128×128 square pixels, each of width 0.42 cm. The sinogram consisted of 192 radial samples and 160 angular samples, similar to the CTI

⁴The h^n s need not be different for every cycle; one may h fixed for a few cycles.

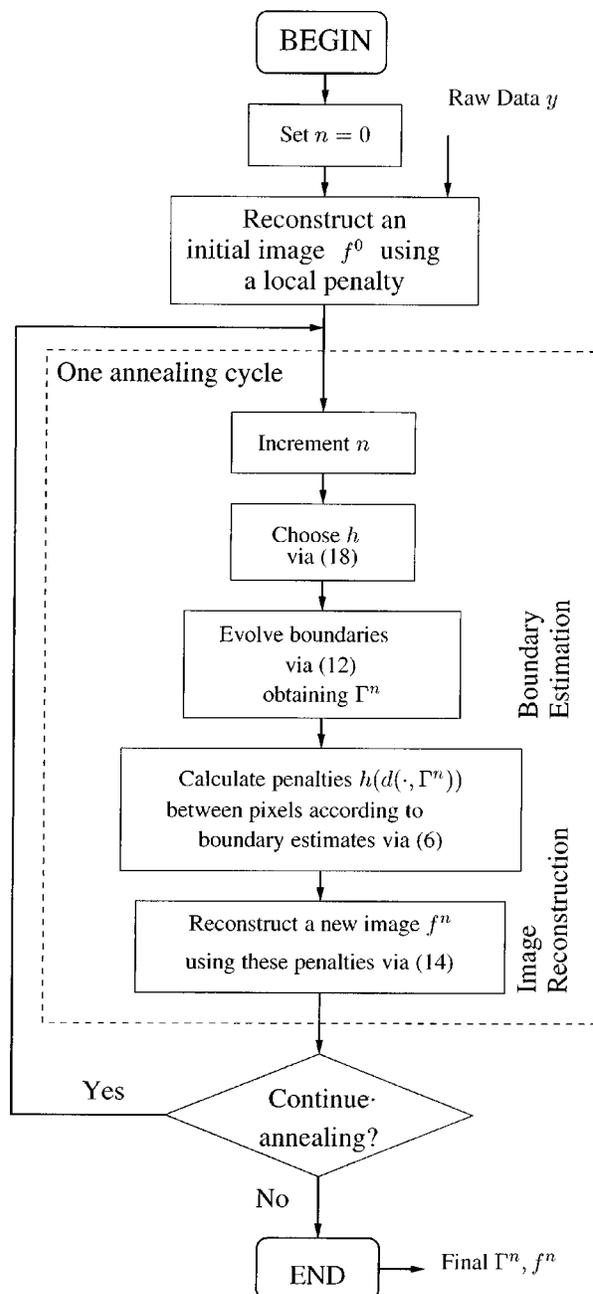


Fig. 9. Algorithm flowchart.

ECAT921 PET scanner. For simplicity we simulated rays with equal spacing of 0.3375 cm. Random coincidences account for about 5% of the recorded counts. We performed two studies, one with 1 M counts, the other with 300 K counts (comparable with 10- and 3-min scans, respectively). Our initial image was reconstructed using conventional space-invariant quadratic penalty over first-order neighbors for the proposed method. The boundaries were initialized manually⁵ as shown in Fig. 10.

We analyzed three regions of interest (ROIs) in the reconstructed attenuation maps: the (true) left lung (region 1), the (true) right lung (region 2), and a 5×5 region (region 3) near

⁵An automatic procedure could be easily developed to obtain better initial boundaries, which would reduce computation time.

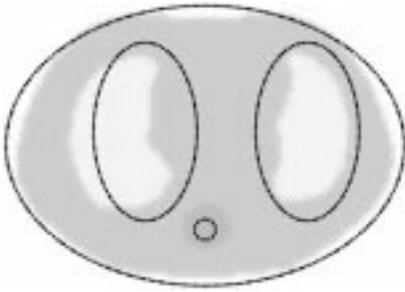


Fig. 10. Initial contours for simulation study.



Fig. 11. The three ROIs; region 1: left lung; region 2: right lung; and region 3: the 5×5 square region.

the heart, as shown in Fig. 11. We performed 50 realizations with pseudo-random Poisson noise. For the i th realization, we computed three ROI means, m_1^i , m_2^i , and m_3^i , via

$$m_k^i = \frac{\sum_{j \in \mathcal{R}_k} \hat{\mu}_j^i}{\sum_{j \in \mathcal{R}_k} 1}, \quad i = 1, \dots, 50 \quad (19)$$

where \mathcal{R}_k is the set of pixels in the k th region and $\hat{\mu}_j^i$ denotes the estimated attenuation coefficient of the j th pixel in the i th realization. Then we computed the sample means ($\bar{m}_k = (1/50) \sum_{i=1}^{50} m_k^i$) and the sample standard deviations ($\sigma_{m_k} = \sqrt{(1/49) \sum_{i=1}^{50} (m_k^i - \bar{m}_k)^2}$) of the three ROI averages, $k = 1 \dots 3$. For the Huber penalty, we plotted the bias versus the standard deviation of the ROI values as a function of the regularization parameter β , for four δ : 0.002/cm, 0.004/cm, 0.008/cm, and 0.02/cm. (Results for $\delta = 0.04$ /cm were nearly identical to those for $\delta = 0.02$ /cm, so are not shown.) The difference in attenuation values between the lung region and the soft tissue region is about 0.08/cm, while the difference between the spine and the soft tissue region is about 0.04/cm, which is also the minimum contrast between neighboring regions; equating δ to one-tenth of this minimum is often effective [36]. For the proposed penalty, we manually selected β and h to cover a range of bias-variance tradeoffs.

For the high count case, Fig. 12 shows one realization from the FBP, Huber (with $\delta = 0.004$ /cm), and proposed recon-

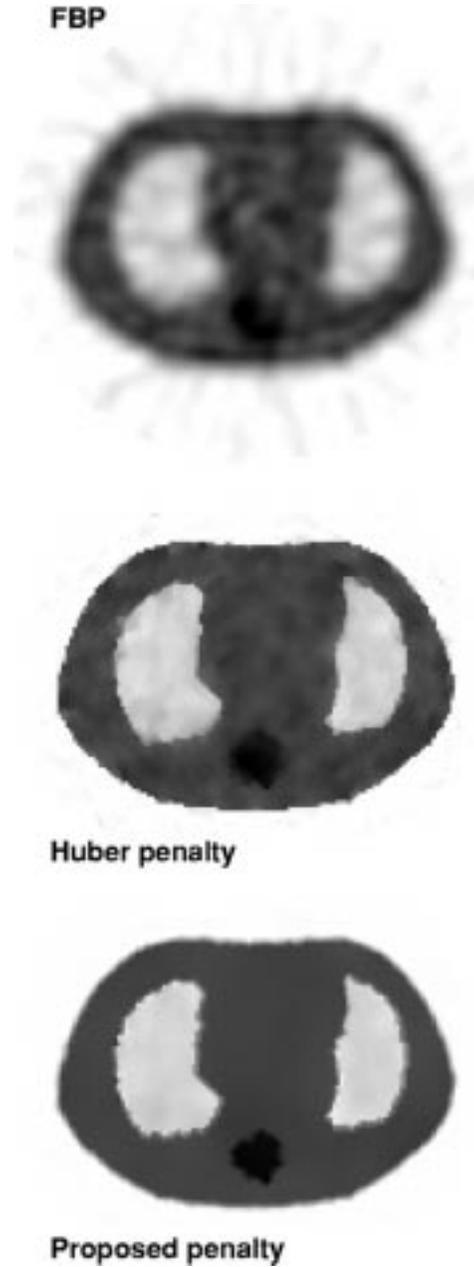


Fig. 12. 1M-count transmission reconstruction.

struction methods. Fig. 13 compares profiles (row 65) from the Huber and proposed reconstruction methods, shown at similar bias levels. The proposed method yields less variance than the Huber method at this bias level. Fig. 14 confirms this initial observation quantitatively by plotting⁶ the bias against the variance of the ROIs defined above. Among Huber penalties, $\delta = 0.02$ /cm produced the best bias-variance curve for regions one and two, while for region three, $\delta = 0.002$ /cm produced the best bias-variance curve. For an image with multiple contrasts like the one used here, it is very difficult to optimize δ , since

⁶Error bars in the bias direction are too small to be shown, and the “bounce back” in the bias-variance curves for the Huber penalty is due to the nonnegativity constraint we place on our estimates; heuristically, when the smoothing parameter is small, the reconstructed image is noisy and many pixels have negative values, but the nonnegativity constraint makes them 0 causing a large positive bias.

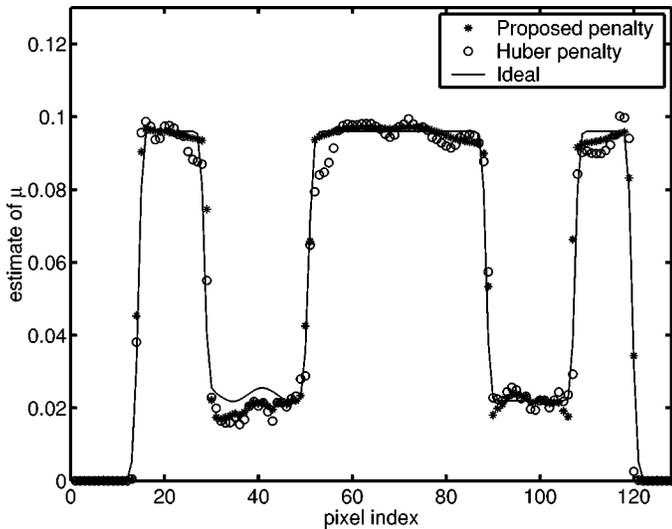


Fig. 13. 1M-count transmission reconstruction profiles: comparison of Huber penalty and proposed penalty.

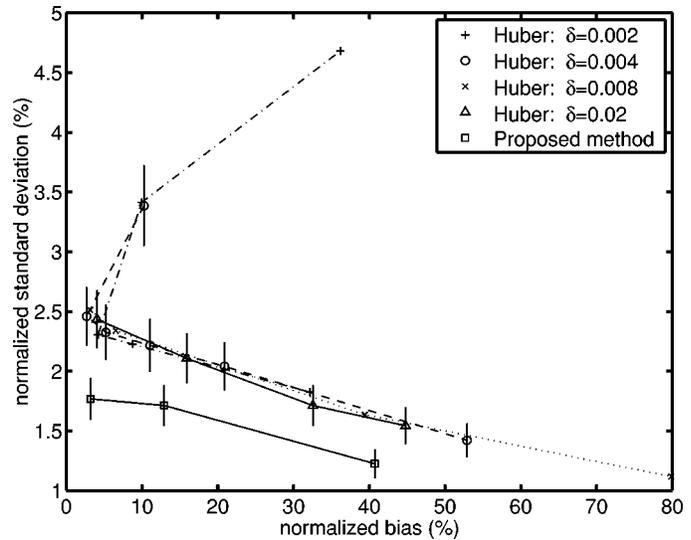
no single δ can produce the best bias-variance curves for all the ROIs. This is a drawback of Huber-like penalty functions.

To investigate how the proposed nonlocal regularization performs when the transmission map is used for attenuation correction of emission reconstructions (for details, see [37]), we simulated noiseless PET emission scans of the digital phantom shown in Fig. 15. The relative radioactivity concentration of the lungs, spine, heart, and body were 12, 9, 40, and 22, respectively. The emission projections included the effects of nonuniform attenuation corresponding to the attenuation map in Fig. 2. We reconstructed emission images using FBP with attenuation correction based on two sets of transmission maps: one set regularized by the proposed nonlocal penalty, and the other set by the Huber penalty.

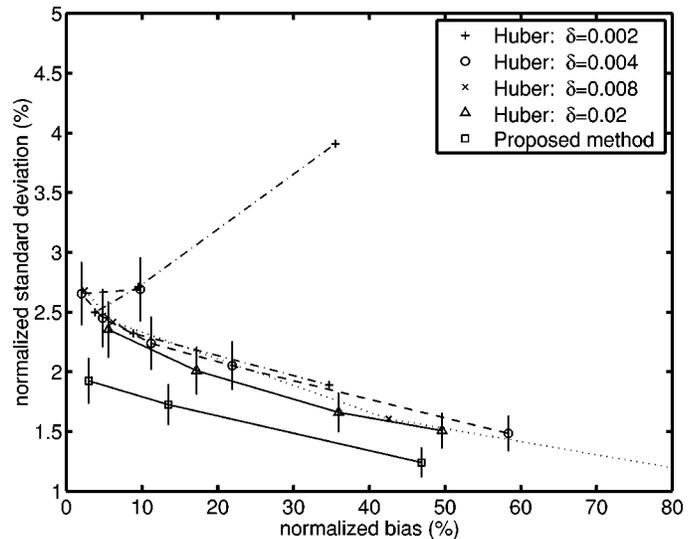
Fig. 16 shows emission reconstructions using transmission maps reconstructed with the Huber penalty and the proposed penalty. To compare the effect of the two different penalties on the emission reconstruction, we selected two groups of reconstructed transmission maps, one using the Huber penalty, and the other using the proposed penalty. Then [38] we smoothed the projection of the emission phantom (applied with ideal attenuation, i.e., the attenuation of each line of response is obtained from the true attenuation coefficients) and the transmission map, so that the resolution of the final emission reconstructions using Huber penalty and the proposed penalty matched each other. We use the following simple method to determine the resolution of a set of reconstructions. Given the ideal image μ^{true} and the average reconstructed image $\check{\mu}$, the resolution of $\check{\mu}$ is

$$\arg \min_{\sigma} \sum_{j \in \mathcal{M}} |[G_{\sigma} \mu^{\text{true}}]_j - \check{\mu}_j|^2 \quad (20)$$

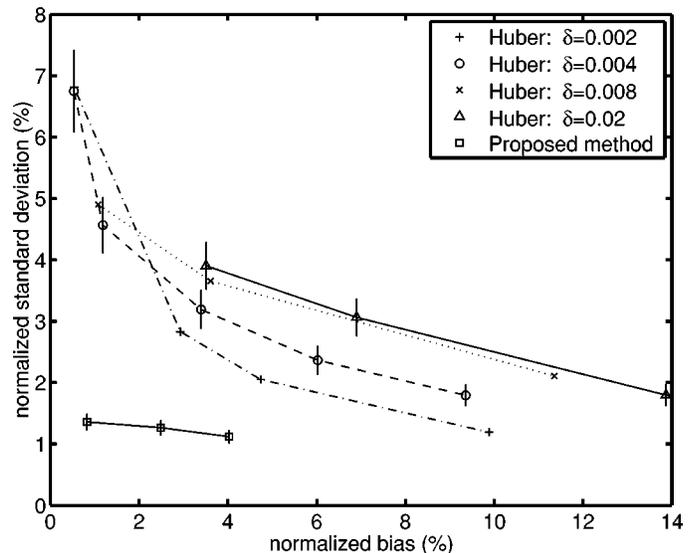
where G_{σ} represents a Gaussian smoothing filter with FWHM σ , and \mathcal{M} denotes a mask used during reconstruction. Table I



(a)



(b)



(c)

Fig. 14. ROI bias-variance plots for 1M-count transmission reconstruction.

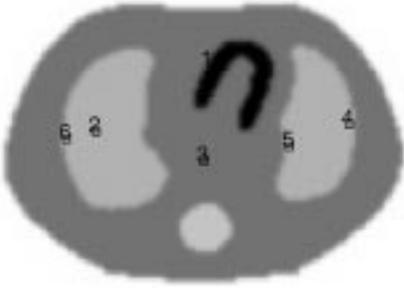


Fig. 15. Emission phantom.

shows the normalized⁷ standard deviation (in units of % for the mean of the six regions of 3×3 pixels as defined in Fig. 15, with a fixed spatial resolution of 4.2 pixels FWHM. The noise is significantly reduced in the interior ROIs, and is statistically comparable for the boundary ROIs.

We also calculated the empirical standard deviation of each pixel for both the proposed method and the Huber method, at a fixed spatial resolution of 4.2 pixels FWHM. For each pixel j in the image we computed

$$\sigma_j \triangleq \sqrt{\frac{1}{49} \sum_{i=1}^{50} \left(\hat{\lambda}_j^i - \frac{1}{50} \sum_{k=1}^{50} \hat{\lambda}_j^k \right)^2} \quad (21)$$

where $\hat{\lambda}_j^i$ denotes the estimated emission intensity of the j th pixel in the i th realization. Fig. 17 shows a histogram of the ratios of these standard deviations, $\{\sigma_j^{\text{proposed}}/\sigma_j^{\text{Huber}}, j \in \mathcal{M}\}$. For 85.8% of the pixels, the proposed method produced lower standard deviations than the Huber method. The median reduction in the standard deviations was 47.1%.

We performed similar studies for the low count case. Fig. 18 shows one realization from the FBP, Huber (with $\delta = 0.004/\text{cm}$), and proposed transmission reconstruction methods. (The streaks in the FBP reconstruction in this case are caused by the fact that some rays recorded zero counts). Fig. 19 compares two profiles (row 65) from the Huber and proposed reconstruction methods; these profiles came from images of similar bias. Fig. 20 shows the bias-variance tradeoffs for the three regions as illustrated in Fig. 11.

Fig. 21 shows the emission reconstructions from noiseless emission data corrected using transmission maps regularized by the proposed penalty and the Huber penalty. The heart region is much more uniform in reconstructed images using the proposed penalty than using the Huber penalty. Fig. 22 shows the histogram of $\sigma_{\text{proposed}}/\sigma_{\text{Huber}}$ for all pixels within the image. For 81.7% of pixels, the proposed penalty produced lower standard deviations than the Huber penalty. The median reduction in the standard deviations was 34.6%. Table II shows the normalized standard deviation (in units of %) for the mean of the six regions of 3×3 pixels as defined in Fig. 15, with a fixed spatial resolution of 6.4 pixels FWHM.

Finally, we present an anecdotal example of reconstruction of a more complex transmission phantom, shown in Fig. 23(a). There are two masses inside the right lung. Fig. 23(c) shows

⁷Normalized by the mean of each region.

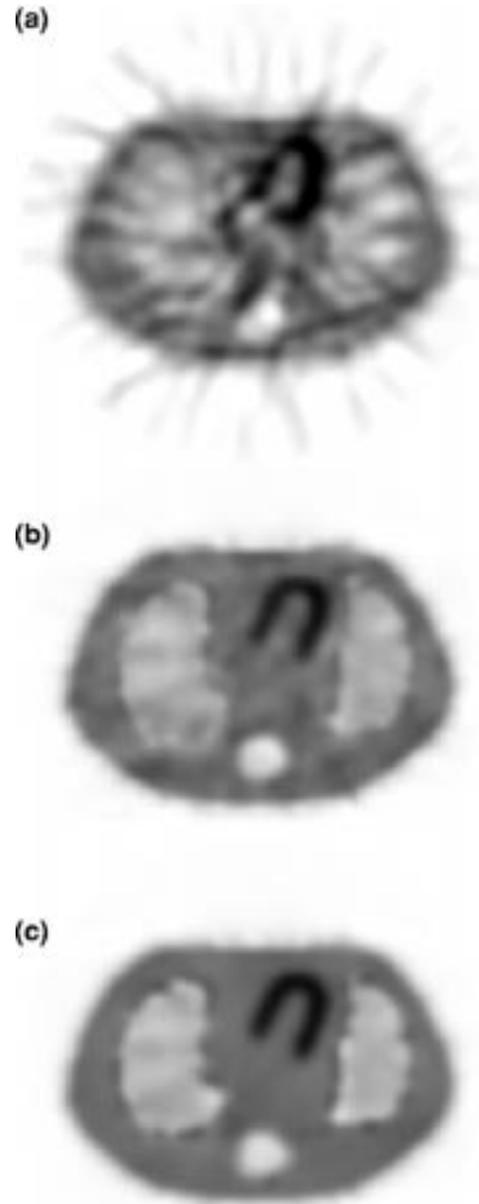


Fig. 16. Emission FBP reconstruction using attenuation correction based on attenuation maps reconstructed from IM-count transmission scans by (a) FBP, (b) Huber penalty, and (c) proposed penalty.

the boundary curve initialization. Fig. 23(d) and (e) show the final boundary curves $\hat{\Gamma}$ and final image \hat{f} estimated by the proposed algorithm, respectively. Fig. 23(b) shows the reconstruction using the Huber penalty. Even though the initial curve for the right lung had only one connected component, the algorithm is capable of topological changes and automatically separated into two connected components, one forming the boundary between the mass inside the right lung and the right lung, the other forming the boundary between the right lung and soft tissue.

IV. DISCUSSION AND CONCLUSION

We have presented a new regularization method for tomographic image reconstruction based on a nonlocal penalty func-

TABLE I
 NORMALIZED STANDARD DEVIATION (%) FOR THE SIX EMISSION ROIS USING 1M-COUNT TRANSMISSION RECONSTRUCTIONS

Region	<i>Interior</i>			<i>Boundary</i>		
	1	2	3	4	5	6
Huber penalty	4.8 ± 0.5	11.0 ± 1.1	6.7 ± 0.7	14.8 ± 1.5	16.4 ± 1.6	15.9 ± 1.6
Proposed penalty	1.1 ± 0.1	8.1 ± 0.8	1.9 ± 0.2	14.4 ± 1.4	17.1 ± 1.7	18.1 ± 1.8

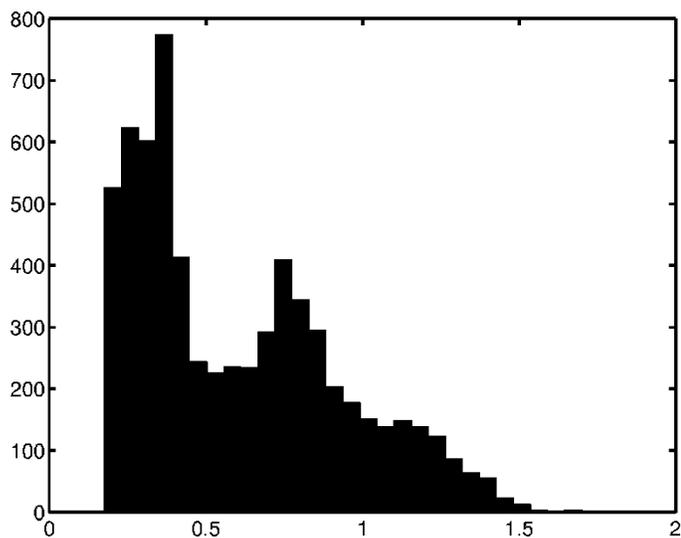


Fig. 17. Histogram of $\sigma_{\text{proposed}}/\sigma_{\text{Huber}}$ for emission reconstruction using 1M-count transmission scans.

tion. In conventional local regularization methods, including most line-site models, the roughness penalty assigned to differences between neighboring pairs of pixels (or to cliques of pixels) depends solely on pixel values within a small *fixed* local neighborhood.

Like line-site models, our regularization method is shift-variant: the roughness penalty strength is reduced near object edges. Unlike line-site models, however, in the proposed approach the penalty function for any pair of neighboring pixels depends on the nearest point to the entire boundary curve, which makes our penalty nonlocal. The region-based Bayesian prior of [6], [9] is also nonlocal. Since we only use *closed* boundary curves, our penalty is implicitly region based; in fact, we use a region-based penalty J_3 during the early stages of our reconstruction.

Since our algorithm is a descent method, the iterates approach a local minimum that depends on the initialization. For example, in the reconstruction of the phantom in Fig. 23, if the initial curve for the right lung lies completely *outside* that lung, then our algorithm will not “find” the tumor inside. Unless additional curves were introduced somehow during later annealing stages, the boundary curve between the tumor and the right lung would not be “found,” losing the benefit of nonlocal penalty near the tumor. Our algorithm was able to split the initial curve shown in Fig. 23(c) for the right lung into the two curves shown in Fig. 23(d), one separating the lung and the soft tissue, the other separating the tumor and the lung, because the initial curve intersected both boundaries. Both boundary curves would also have been “found” if the initial curve for the right lung were completely inside that lung.

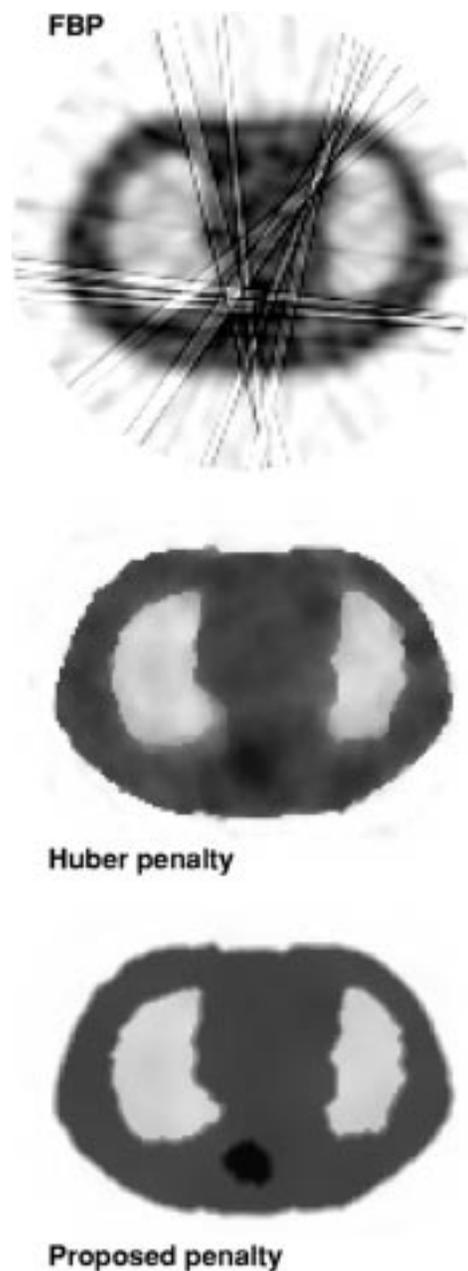


Fig. 18. 300 K-count transmission reconstruction.

As described above, a weakness of our algorithm is its inability to move past existing boundaries. Fundamentally, this weakness is due to the fact that the curves move in the direction along which the cost function decreases the fastest, i.e., a greedy strategy. This is a traditional difficulty with PDE formulations. On the other hand, the PDE formulation (with level sets) allows the curves to merge or split without direct intervention.

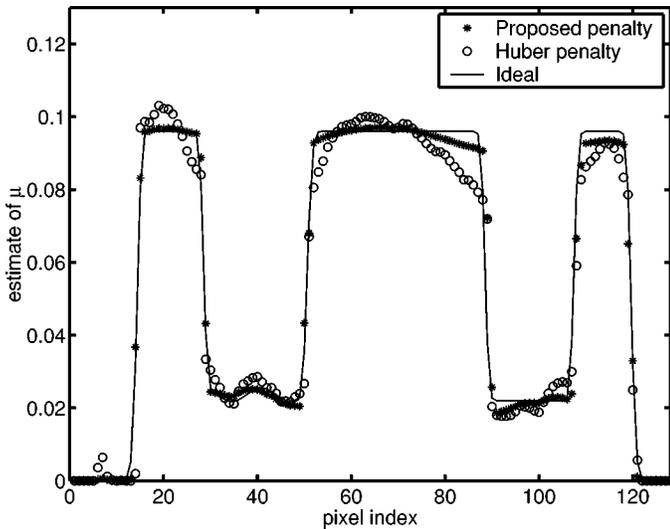


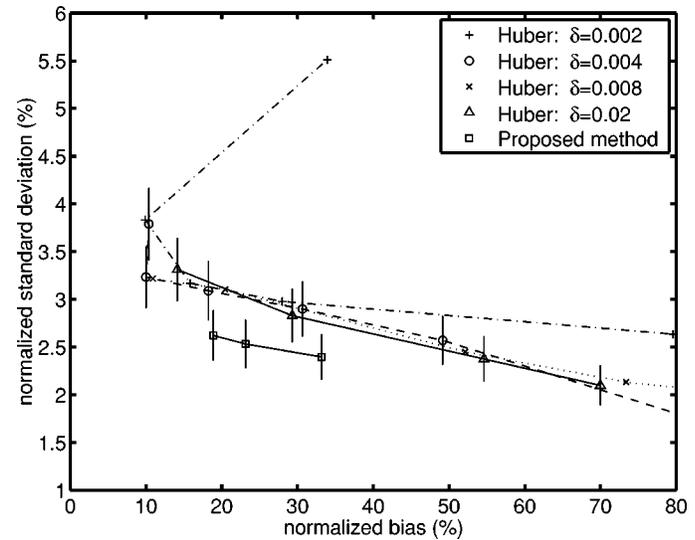
Fig. 19. 300 K-count transmission reconstruction profiles: comparison of Huber penalty and proposed penalty.

Hence, we do not need to initialize with the “correct” number of curves, provided the initial curves are suitably placed as described above.

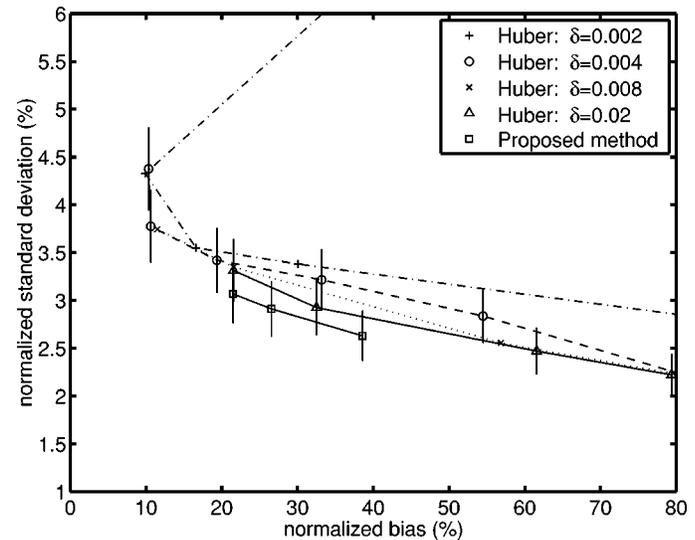
One could partially overcome this limitation by initializing with many small curves (e.g., circles) that would likely intersect most object edges. This strategy could lead to better local minima, although it would require more computation in the early iterations.

The proposed nonlocal penalty produces transmission reconstructions with better ROI bias-variance tradeoffs than a local Huber penalty. When these transmission reconstructions are applied to noiseless emission data, the nonlocal penalty used for transmission reconstruction produces emission images with smaller variances (for a fixed spatial resolution) for most (80%–85%) pixels in the image; the median standard deviation in the image is reduced by 35% to 50% relative to the Huber method.

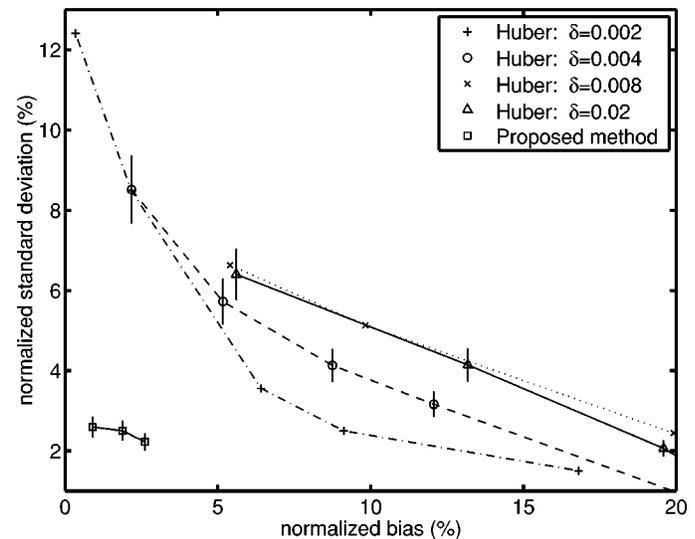
However, reconstruction using the proposed penalty is more time consuming than using conventional local penalties. Each cycle of the annealing process requires one “stage” of image reconstruction. If one runs five annealing cycles, the time devoted to updating the image is up to five times that of the local penalties, although for “reconstruction” stage of the second to fifth annealing cycles, we can use fewer iterations than in the first annealing cycle. The “boundary estimate” stage is also most time-consuming during the first cycle of the annealing process. Since we deliberately used initial boundaries that were far away from the true boundary (to demonstrate the robustness of our boundary estimation algorithm), it took about 60 time steps for the two lung boundary curves to converge (the body and spine boundary curves converge much faster). Each time step of curve evolution for each of the two lungs takes about 60% of the time needed for a single iteration of Huber reconstruction. The total time needed by the “boundary estimate” stages in all annealing cycles greatly depends on initialization of curves, size of the time steps, etc. In the present implementation, the boundary estimate stage of the first annealing cycle takes about twice the time of an “image reconstruction” stage. Subsequent boundary estimations were much faster since we evolved from the curve



(a)



(b)



(c)

Fig. 20. Bias-variance plots for 300 K-count transmission reconstructions.

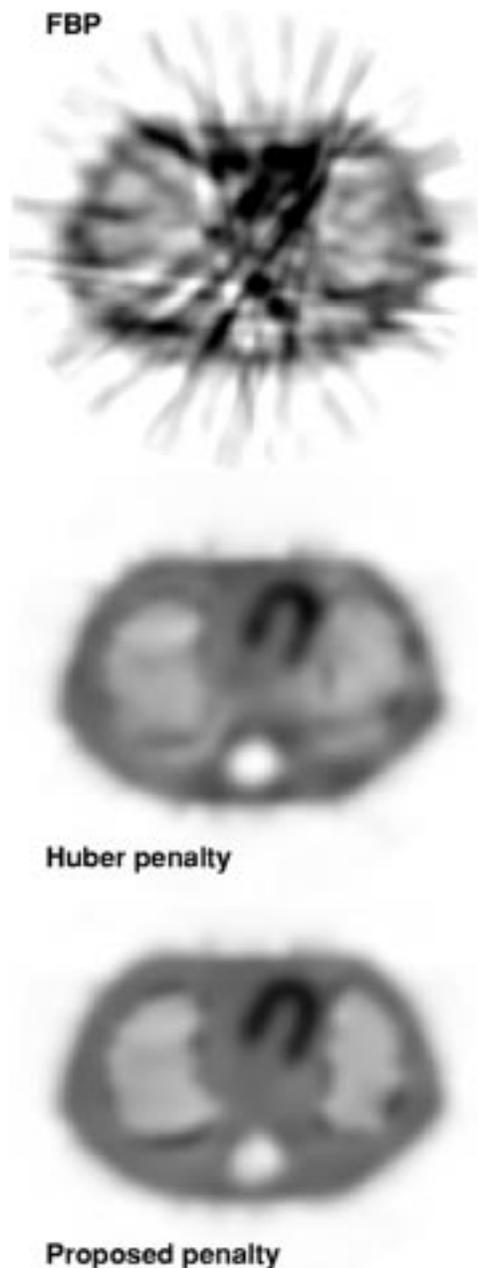


Fig. 21. Emission FBP reconstruction using attenuation correction based on attenuation maps reconstructed from 300 K-count transmission scans.

estimated by the previous cycle in the annealing process. In total, computing the boundary estimates takes about the time of three to four reconstructions.⁸

A related but fundamentally different approach to the problem of reconstructing attenuation maps using region information has also been proposed [39] based on the following functional of the boundary curves C :

$$J(C) = \int_0^\lambda W(C(p))|C'(p)| dp$$

⁸In the present implementation, the “boundary estimate” stage is written in Matlab code, as compared to relatively optimized C code for the image reconstruction stage; hence significant improvement can be made by converting the code for “boundary estimate” to C.

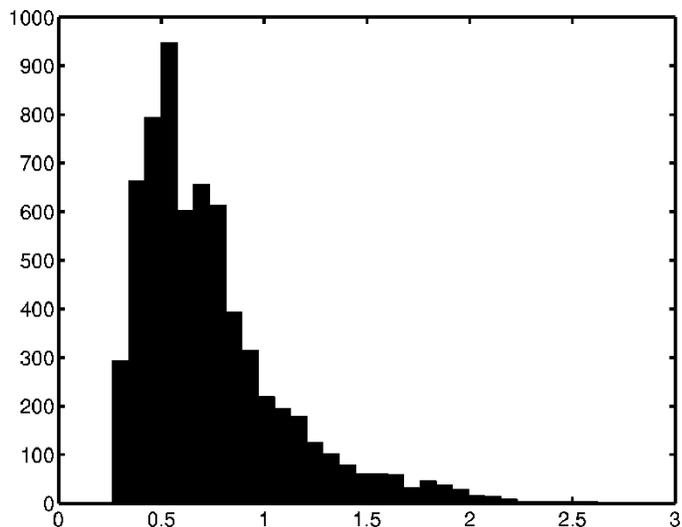


Fig. 22. Histogram of $\sigma_{\text{proposed}}/\sigma_{\text{Huber}}$ for emission reconstruction using 300 K-count transmission scans.

where p parameterizes C and W is a data consistency term. This functional was minimized by evolving a PDE to obtain a “segmentation” of the attenuation map without actually “reconstructing” the attenuation map [39]. That approach assumed that the attenuation coefficient inside each region is constant, whereas our approach allows for nonuniform regions as may arise in clinical situations.

We have performed a preliminary investigation applying the proposed nonlocal penalty to three-dimensional (3-D) image reconstruction [25], [26]. We have not systematically investigated the bias-variance properties for the 3-D reconstructions, but the interslice information should give our penalty a further advantage over conventional local regularization methods.

APPENDIX

The signed distance function d , as defined following (6), may not be differentiable (in the functional sense) in Γ for the x s that are closest to two or more points on the curve Γ , and herein lies the main difficulty in proving (15) rigorously. But we can modify d slightly to produce a functional that is differentiable in Γ for all x . One way is to define $d_q: (\Omega - \Gamma) \times \mathcal{G} \mapsto \mathbb{R}$ by

$$d_q(x, \Gamma) = \pm \left(\int_{\mathcal{S}} \frac{1}{\|\gamma(s) - x\|^q} ds \right)^{-1/q}$$

where γ parameterizes the curve Γ by arc length $s \in \mathcal{S} \subset \mathbb{R}$, the integer q is even, and d_q takes the + or – sign depending on whether x is “inside” or “outside” Γ . Note that

$$\begin{aligned} \left(\int_{\mathcal{S}} \frac{1}{\|\gamma(s) - x\|^q} ds \right)^{-1/q} &\rightarrow \left[\max_{p \in \Gamma} \frac{1}{\|p - x\|} \right]^{-1} \\ &= \min_{p \in \Gamma} \|p - x\| = d(x, \Gamma) \end{aligned}$$

as $q \rightarrow \infty$. Unlike d , the above d_q has the nice property of being smooth in x and γ on its domain. Replacing d by d_q in our definition of J_2 in (6), i.e., defining

$$\bar{J}_2(f, \Gamma) = \int_{\Omega - \Gamma} h(d_q(x, \Gamma)) |\nabla f(x)|^2 dx$$

TABLE II
NORMALIZED STANDARD DEVIATION (%) FOR THE SIX EMISSION ROIS USING 300 K-COUNT TRANSMISSION RECONSTRUCTIONS

Region	Interior			Boundary		
	1	2	3	4	5	6
Huber penalty	5.0 ± 0.5	14.1 ± 1.4	9.0 ± 0.9	14.0 ± 1.4	15.7 ± 1.6	15.3 ± 1.5
Proposed penalty	1.7 ± 0.2	8.5 ± 0.9	3.9 ± 0.4	13.4 ± 1.3	18.0 ± 1.8	14.9 ± 1.5

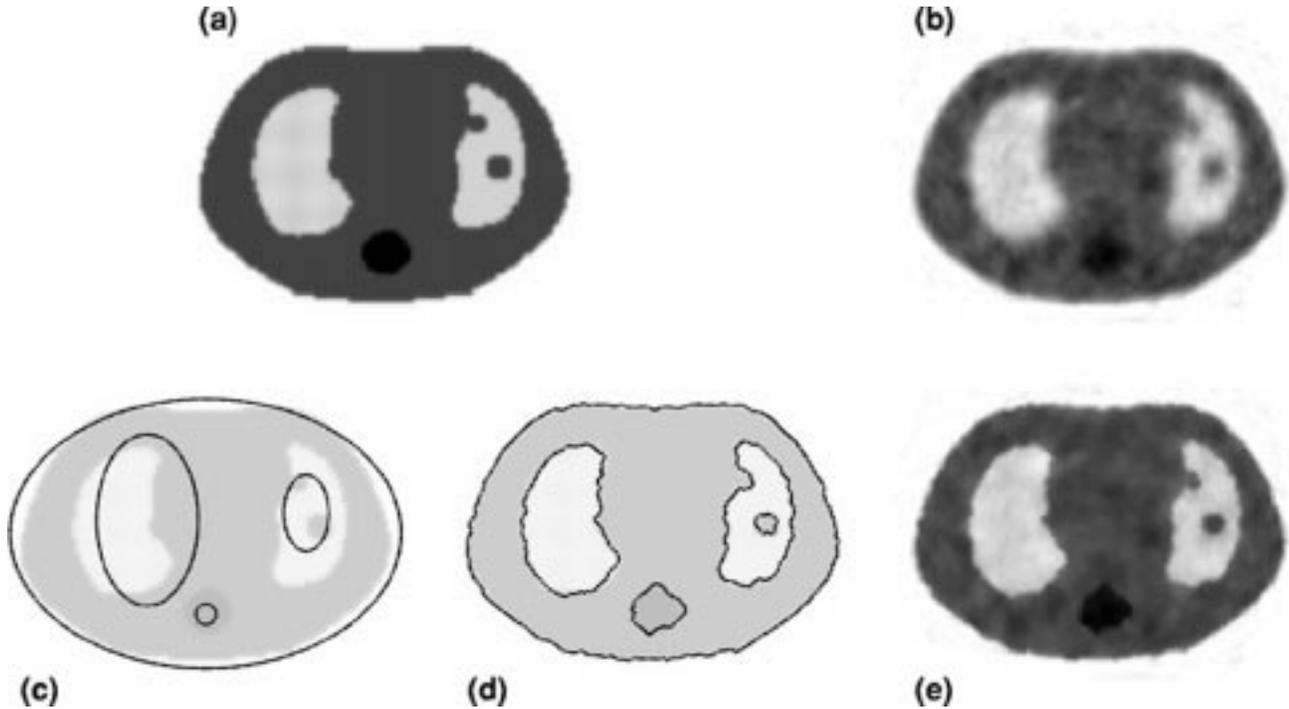


Fig. 23. (a) Phantom with lung masses. (b) Huber-penalty reconstruction. (c) Initial curves. (d) Final estimated curves. (e) Final reconstructed image using the proposed algorithm.

leads to the following functional derivative:

$$\begin{aligned} \frac{\delta \bar{J}_2}{\delta v} &= \int_{\Omega - \Gamma} \frac{\delta}{\delta v} h(d_q(x, \Gamma)) |\nabla f(x)|^2 dx \\ &= \int_{\Omega - \Gamma} h'(d_q(x, \Gamma)) \frac{\delta}{\delta v} d_q(x, \Gamma) |\nabla f(x)|^2 dx. \end{aligned}$$

In the first equality, the interchange of differentiation and integration is legitimate because h and d_q are smooth, and we use the chain rule in the second equality. Since $(\delta/\delta v) d_q(x, \Gamma)$ approaches $\delta_0 I_{(r), \Gamma}(r) \vec{N}$ as $q \rightarrow \infty$, where $I_{(r), \Gamma}(r)$ is defined in (16) and δ_0 denotes the Dirac delta, the functional derivative of J_2 given in (15) is, in some sense, a “weak limit” of $(\delta \bar{J}_2/\delta v)$.

Implementation-wise, there is no difference between d and d_q provided q is large enough that the difference between d and d_q is below machine precision.

ACKNOWLEDGMENT

The authors would like to thank E. P. Ficaro for providing the digital phantom. D. F. Yu acknowledges M. Miller, J. Prince, and C. Xu for suggestions inspiring the Appendix.

REFERENCES

- [1] M. Bertero, C. De Mol, and E. R. Pike, “Linear inverse problems with discrete data—I: General formulation and singular system analysis,” *Inverse Prob.*, vol. 1, no. 4, pp. 301–330, Nov. 1985.
- [2] M. Bertero, T. Poggio, and V. Torre, “Ill posed problems in early vision,” *Proc. IEEE*, vol. 76, pp. 869–889, Aug. 1988.
- [3] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs,” *IEEE Trans. Image Processing*, vol. 5, pp. 1346–1358, Sept. 1996.
- [4] J. W. Stayman and J. A. Fessler, “Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction,” *IEEE Trans. Med. Imag.*, vol. 19, pp. 601–615, June 2000.
- [5] H. Derin and H. Elliott, “Modeling and segmentation of noisy and textured images using Gibbs random fields,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 39–55, Jan. 1987.
- [6] V. E. Johnson, “A model for segmentation and analysis of noisy images,” *J. Amer. Statist. Assn.*, vol. 89, no. 425, pp. 230–241, Mar. 1994.
- [7] V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen, “Image restoration using Gibbs priors: Boundary modeling, treatment of blurring, and selection of hyperparameter,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 413–425, May 1991.
- [8] V. Johnson, “A framework for incorporating structural prior information into the estimation of medical images,” in *Information Processing in Medical Imaging*, H. H. Barrett and A. F. Gmitro, Eds. Berlin, Germany: Springer-Verlag, 1993, pp. 307–321.
- [9] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman, “Bayesian reconstruction and use of anatomical a priori information for emission tomography,” *IEEE Trans. Med. Imag.*, vol. 15, pp. 673–686, Oct. 1996.

- [10] J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," *Proc. SPIE*, vol. 3071, pp. 184–194, 1997.
- [11] —, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, June 1994.
- [12] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
- [13] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 2–12, Jan. 1989.
- [14] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and I. G. Zubal, "Bayesian reconstruction for emission tomography via deterministic annealing," in *Lecture Notes in Computer Science*, H. H. Barrett and A. F. Gmitro, Eds. Berlin, Germany: Springer-Verlag, 1993, vol. 687, Information Processing in Medical Imaging, pp. 322–338.
- [15] M. Lee, A. Rangarajan, G. I. Zubal, and G. Gindi, "A continuation method for emission tomography," *IEEE Trans. Nucl. Sci.*, vol. 40, pp. 2049–2058, Dec. 1993.
- [16] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imag.*, vol. 14, pp. 669–680, Dec. 1995.
- [17] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRFs: Surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 401–412, May 1991.
- [18] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, Mar. 1992.
- [19] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, July 1995.
- [20] M. J. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, July 1996.
- [21] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.
- [22] B. W. Silverman, C. Jennison, J. Stander, and T. C. Brown, "The specification of edge penalties for regular and irregular pixel images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1017–1024, Oct. 1990.
- [23] F. C. Jeng and J. W. Woods, "Compound Gauss–Markov random fields for image estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 683–697, Mar. 1991.
- [24] S. C. Huang, E. J. Hoffman, M. E. Phelps, and D. E. Kuhl, "Quantitation in positron emission computed tomography: 2 Effects of inaccurate attenuation correction," *J. Comput. Assist. Tomogr.*, vol. 3, no. 6, pp. 804–814, Dec. 1979.
- [25] D. F. Yu, "Statistical methods for transmission image reconstruction with nonlocal edge-preserving regularization," Ph.D. dissertation, Univ. Michigan, Ann Arbor, 2000.
- [26] D. F. Yu and J. A. Fessler, "Three-dimensional nonlocal edge-preserving regularization for PET transmission reconstruction," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 2, 2000, pp. 1566–1570.
- [27] —, "Edge-preserving tomographic reconstruction with nonlocal regularization," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 1998, pp. 29–33.
- [28] J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences*. Cambridge, MA: Cambridge Univ. Press, 1996.
- [29] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, NJ: Prentice-Hall, 1963. Translation by R. A. Silverman.
- [30] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans. Med. Imag.*, vol. 16, pp. 199–210, Apr. 1997.
- [31] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 158–176, Feb. 1995.
- [32] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Conformal curvature flows: From phase transitions to active vision," *Arch. Rational Mech. Anal.*, vol. 134, no. 3, pp. 275–301, 1996.
- [33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, 2 ed. New York: Cambridge Univ. Press, 1992.
- [34] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Processing*, vol. 8, pp. 688–699, May 1999.
- [35] H. Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, pp. 801–814, Sept. 1999.
- [36] E. Ümücuoğlu, R. M. Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Biol.*, vol. 41, no. 9, pp. 1777–1807, Sept. 1996.
- [37] M. Yavuz and J. A. Fessler, "Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans," *IEEE Trans. Med. Imag.*, vol. 18, pp. 665–674, Aug. 1999.
- [38] A. Chatzioannou and M. Dahlbom, "Detailed investigation of transmission and emission data smoothing protocols and their effects on emission images," *IEEE Trans. Nucl. Sci.*, vol. 43, pp. 290–294, Feb. 1996.
- [39] E. Debreuve, M. Barlaud, G. Aubert, and J. Darcourt, "Attenuation map segmentation without reconstruction using a level set method in nuclear medicine imaging," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 1998, pp. 34–38.