Deformable Template Models for Emission Tomography

Yali Amit and Kevin M. Manbeck

Abstract—**The reconstruction of emission tomography data is an ill-posed inverse problem and, as such, requires some form of regularization. Previous efforts to regularize the restoration process have incorporated rather general assumptions about the isotope distribution within a patient's body. Here we present a theoretical and algorithmic framework in which the notion of a deformable template can be used to identify and quantify brain tumors in pediatric patients. Patient data and computer simulation experiments are presented which illustrate the performance of the deformable template approach to single photon emission computed tomography (SPECT).**

I. Introduction

SINGLE photon emission computed tomography (SPECT) and positron emission tomography (PET) are powerful medical imaging techniques that are underutilized in the clinical setting. The SPECT and PET modalities are plagued by generally low-quality images, with low spatial resolution. This is unfortunate, because SPECT and PET are able to measure *metabolic activity,* which is currently impossible to measure with more popular modalities such as computed tomography (CT) and magnetic resonance imaging (MRI). Furthermore, the development of highly discriminating monoclonal antibodies for use in SPECT promises to allow clinicians to accurately locate both normal and pathological specific tissue types.

Both PET and SPECT data are badly degraded by various sources of random noise. As a result, statistical approaches are well-suited to the problem of emission tomography image reconstruction. In particular, previous work has demonstrated a potential utility of applying low-level *prior knowledge,* in a Bayesian framework (see [20], [7], [13], [16]) to the problem of image reconstruction. Here we propose higher-level mathematical constraints that can be applied to the problem of reconstructing SPECT brain tumor data in the Bayesian framework.

A. Soft Modeling, Template Modeling, and Hard Modeling

Typically, image reconstruction methods attempt to produce an image which captures the original brightness pattern of a scene without degrading effects. They incorporate generic assumptions about the images under study, including, for example, constraints on the intensity range, allowed Fourier frequencies, degree of local smoothness, or entropy of allowed images (for a survey and extended bibliography, see Geman [4]). Following Dupuis, Geman, Horowitz, and Reynolds [17], these approaches are referred to as "soft modeling."

In recent years, attempts have been made to incorporate higher-level knowledge into image restoration methodologies. The basic idea is to assume that each image is generated by some element from a set of objects, and each such element is obtained as some transformation of a reference object, called the template. The transformations are defined so as to preserve certain topological features of the template which characterize the set of objects.

A prior distribution is then constructed on the set of transformations, thus indirectly inducing a prior distribution on the set of objects and on the images they generate. In this context, the prior knowledge consists not only of the prior distribution but also of the typical fixed (nonrandom) template representing the type of objects being observed. A likelihood function is introduced which describes the distribution on the space of images given an object from the class, or equivalently given the transformation which generated that object. Finally, Bayes formula yields a posterior distribution on the *space of transformations.* In contrast to the "soft model" approach, where the posterior is given directly on the space of images, the "template model" approach specifies the posterior on the space of template transformations, which indirectly induces a posterior distribution on the space of images.

The restoration procedure consists of finding the transformation which minimizes the posterior. This transformation is then applied to the template in order to generate the restored image. In addition to the restoration of the image, the estimated transformation provides a *structural interpretation* of the image in terms of the template. Important landmarks in the template are mapped into the corresponding landmarks in the image, and can thus be automatically located (see Chow, Grenander, and Keenan [22], Ripley [19], Amit, Grenander, and Piccioni [2]).

In theory, soft model and template model approaches involve a very large number of parameters and can be considered nonparametric. However, the type of information introduced in the template approach greatly reduces the support of the posterior distribution and lends itself to very low-dimensional approximations.

The template approach has the drawback that it requires the user to introduce a template. In that sense it is more limited.

Manuscript received November 7, 1991; revised June 11, 1992. The work of Y. Amit was supported by the Army Research Office under Contract DAAL03-86-K-0171 to the Center for Intelligent Control Systems and by The University of Chicago Block Fund. The work of K. M. Manbeck was supported by the National Science Foundation under Grants ISI-9060519 and DMS-8813699.

Y. Amit is with the Department of Statistics, University of Chicago, Chicago, IL 60637.

K. M. Manbeck is with the Division of Applied Mathematics, Brown University, Providence, RI 02912.

IEEE Log Number 9208169.

However, it should be emphasized that the same procedure can be applied to different problems which involve the extraction of similar types of topological features. The only change would be in the template used. For example, the same procedure developed in [22] for identifying boundaries of hands using an outline of a typical hand is used in Knoerr [11] to identify outlines of leaves using a typical leaf outline.

A third approach applicable in some image restoration tasks is to incorporate highly specific, application-dependent constraints. For example, Dupuis *et al.* [17] note that by modeling relevant physical phenomena in some astrophysical imaging problems, it is possible to formulate mathematical constraints which allow the true intensity map to be described by a parametric scene model. This approach is referred to as "hard modeling."

Each approach has its own merits, and the approach chosen for a particular image restoration task should depend on the specific goals of the task. The primary strength of the soft modeling approach is its versatility. For example, the smoothing prior proposed by Geman and McClure [7], [8] for tomography reconstruction has been successfully applied to many disparate areas of image restoration including infrared image enhancement [5], image deblurring [6], movie restoration [9], and Hubble telescope image enhancement [14].

The primary strength of the hard modeling approach is in the ability to directly estimate certain parameters which might be of scientific importance without actually restoring the image. In the case of soft model image restoration, any parameter estimates of a global scene or assessments regarding the shape or structure of an object within it will necessarily be based on measurements of the restored image. While the soft model restoration technique may produce an image that satisfies some pixel-by-pixel optimality criterion, the optimality of the parameter estimates themselves is unknown. On the other hand, the hard model approach estimates the optimal parameters directly from the data. Furthermore, the hard modeling approach may be applicable at noise levels at which the generic models break down. Generally speaking, when more scientific knowledge is incorporated into these models, the resulting parameter estimates are more reliable.

The deformable template model lies somewhere between the two. It does not require hard scientific modeling. General assumptions regarding the shape of the object in the image are easily incorporated. For example, when trying to identify some object in a scene, the general assumptions could be that the outline of the object is a closed curve and that different likelihood functions describe the interior and exterior of the object. Even simple assumptions of this form can greatly assist in identifying the required object within the image (see Grenander and Miller [10]). More restrictive templates would have an initial shape or structure that is preserved under the transformations (see [2] or [22]). The template model approach seeks to estimate the optimal transformation directly from the data.

B. Planar Deformations

This paper investigates the feasibility of employing a deformable template approach similar to that in [2] to the restoration of SPECT images of brain tumors.

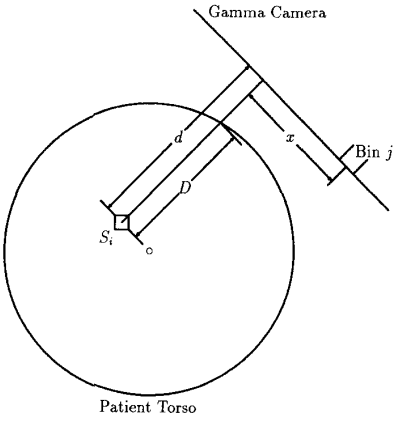

Fig. 1. Detection density, due to a source at site S_i, is modeled as a function of: 1) distance d from the source to the camera; 2) distance D from the source to the boundary of the object; 3) distance x from the projection of S_i.

The radiopharmaceutical used in the patient studies was Thallium-201. This agent concentrates in the interstitial fluid of the scalp and in the brain tumor, if one is present. Thallium-201 does not collect in either the brain or the cerebrospinal fluid because it respects the blood–brain barrier. Consequently, a cross-sectional image of a brain tumor patient should contain an active region corresponding to the tumor which is surrounded by an inactive background corresponding to the brain. An active annulus, corresponding to the scalp, should surround the brain.

Consequently, the template used to describe the isotope intensity in the presence of the tumor is a bivariate Gaussian "hump," surrounded by an ellipse describing the isotope concentration in the scalp (see Fig. 1).

In this model, the template is transformed by composing it with *planar mappings*. Initially, affine mappings are used. Subsequently, more general mappings, as in [1], are introduced, parameterized via their expansion in some orthonormal basis. In both cases, the posterior is derived from the likelihood of the data given a transformation which is applied to the template, and the prior distribution on the space of transformations. The posterior is then maximized using the simplex method for the affine mappings and using a gradient descent method for the more general planar mappings.

The choice of template does not represent any scientific modeling of the structure of the tumor, rather it expresses the assumption that the tumor will be simply connected, each level curve is expected to have one component only, and level curves of lower value contain those of higher value. Finally, the tumor is expected to have a local support somewhere within the image. The rotational symmetry avoids any bias regarding the shape of the tumor. These are very general assumptions regarding the topography of the tumor intensity function. Increased knowledge about this topography would allow us to design more appropriate templates, which could be introduced directly into the existing algorithm.

In fact, the introduction of realistic scientific models for the structure of the tumors in question, in terms of a small number

of relevant parameters, could enable the use of a "hard modeling" approach to this problem. This would of course supply the physician with the parameters of clinical interest; but, as mentioned above, the resulting algorithm would be dedicated to tumor parameter estimation. The algorithm described in this paper solves a more general problem. It can serve for identifying other types of objects in tomography data *as well as* in other types of degradation settings which involve some form of blurring.

The experiments described in this paper are inspired by SPECT data of pediatric brain tumor patients. We assess the feasibility of applying deformable template models to the SPECT reconstruction process by carefully studying computer simulated brain tumor data. The brain phantom data are realistic in the sense that they are generated from a stochastic model which simulates all significant sources of noise in SPECT images. Once the feasibility of the deformable template approach is established, the technique is applied to patient data. We believe that the experiments presented here indicate that in those cases where a template can be defined, reconstruction using the proposed deformable template algorithm will yield more precise results than the more generic "soft modeling" techniques.

Section II presents the forward model used to describe the stochastic mechanism which generates the observed data. This model is then used to generate the synthetic data from a given isotope concentration map. Section III presents the deformable template approach to SPECT reconstruction. Section IV presents some experimental results; and Section V has some concluding remarks.

II. The Forward Model

A. Mathematical Formulation

We have used the following standard Poisson model for the observed data. Let X be the isotope concentration map and let Y be the observed data. The transverse patient slice X is typically discretized into an $N \times N$ array of pixels (usually 64×64). X_i is the isotope concentration at pixel i and physically represents the *mean* of a Poisson process. The gamma camera stores data in J bins (typically 64×64). Let Y_j be the number of photon counts in bin j.

The physics of SPECT imaging determines that the observed data $(Y_j)_{j=1}^{J}$ is a collection of independent Poisson random variables with mean

$$E[Y_j] = \sum_{i=1}^{N^2} \mathcal{A}_{ji} X_i = (\mathcal{A}X)_j \tag{1}$$

where $\mathcal{A}_{ji}$ is proportional to the probability that a photon released at pixel site i is detected in camera bin j. These probabilities are determined by attenuation through the three basic types of material: bone, soft tissue, and lung; by Compton scattering effects which depend on the distance between pixel i and bin j; and by a blurring caused by the "collimator effect" of the gamma camera. The matrix $\mathcal{A}$ is known as the *modified Radon transform* (MRT).

Given the isotope intensity distribution $X = \{X_i\}_{i=1}^{N^2}$, the Poisson likelihood model for Y is then

$$P(Y|X) = \prod_{j=1}^{J} \frac{\left(\sum_{i=1}^{N^2} \mathcal{A}_{ji} X_i\right)^{Y_j}}{Y_j!} \cdot \exp\left(-\sum_{i=1}^{N^2} \mathcal{A}_{ji} X_i\right). \tag{2}$$

This model, although physically precise, gives rise to serious computational difficulties when the actual estimation of X is carried out. Significant computational savings can be achieved as follows. The Poisson distribution at each bin has mean $(\mathcal{A}X)_j$, and can be approximated as a Gaussian distribution with both mean and variance given by $(\mathcal{A}X)_j$. This would give rise to the following likelihood model:

$$P(Y|X) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi(\mathcal{A}X)_j}} \cdot \exp\left\{-\frac{((\mathcal{A}X)_j - Y_j)^2}{2(\mathcal{A}X)_j}\right\}.$$

This new model would not in itself provide for less computation because of the presence of the unknowns in the denominator. However, if the denominator, namely, the variance at each bin, is replaced by the value of the data at that bin, i.e., Y_j, maximizing the likelihood will reduce to solving the *weighted* least squares problem

$$\min_X \sum_{j=1}^{J} \frac{[(\mathcal{A}X)_j - Y_j]^2}{2Y_j} \tag{3}$$

which clearly leads to a drastic decrease in computations.

Observe that the data Y_j are unbiased estimators of the intensity $(\mathcal{A}X)_j$ of the Poisson process at bin j. Moreover, in experiments carried out with the soft model restoration techniques, there has been no significant difference between using the true Poisson model or the weighted least square model (see [15]). In terms of the deformable template models, the calculations would have been virtually impossible without this modification. It should be noted that careful studies in [21] have shown that replacing the Poisson model with a simple *unweighted* least square problem, which is equivalent to solving the system $\mathcal{A}X = Y$, leads to serious variance problems.

B. The Modified Radon Transform

The MRT $\mathcal{A}_{ji}$ is modeled through a *detection density*, whose integral over the width of the jth detection bin, B_j, is proportional to $\mathcal{A}_{ji}$. The underlying assumption here is that the medium is more or less a uniform water medium, which is a reasonable assumption for SPECT of the brain.

Referring to Fig. 1, the detection density α is modeled as a function of: 1) the (perpendicular) distance d, from pixel site S_i to the camera containing bin B_j; 2) the distance D, this same perpendicular line passes through the object being

imaged, in going from location S_i to the camera containing bin B_j; and 3) the distance x, along the camera surface, measured from the perpendicular projection of S_i. Thus, $\alpha = \alpha(D, d, x)$ and

$$\mathcal{A}_{ji} = k \int_{B_j} \alpha(D, d, x) dx. \tag{4}$$

The proportionality constant k is independent of i and j. It is a scale parameter, and its value depends on how we choose physical units such as the time unit and the scale for the observed projection counts Y.

There are two contributions to the detection density: a *direct* contribution from detected photons that have not been scattered; and a *scatter* contribution from photons that have undergone Compton scatter prior to being detected. Thus,

$$\alpha(D, d, x) = \alpha_{\text{direct}}(D, d, x) + \alpha_{\text{scatter}}(D, d, x). \tag{5}$$

It is convenient to factor each of the three detection densities into the product of a depth-dependent attenuation term and a residual line spread function, where, by definition, the former depends only on D and d, and the latter integrates to one with respect to x:

$$\begin{aligned}
\alpha(D, d, x) &= A(D, d)\zeta(x; D, d) \\
\alpha_{\text{direct}}(D, d, x) &= A_{\text{direct}}(D, d)\zeta_{\text{direct}}(x; D, d) \\
\alpha_{\text{scatter}}(D, d, x) &= A_{\text{scatter}}(D, d)\zeta_{\text{scatter}}(x; D, d) \\
\int_x \zeta(x; D, d) &= \int_x \zeta_{\text{direct}}(x; D, d) \\
&= \int_x \zeta_{\text{scatter}}(x; D, d) = 1.
\end{aligned}$$

Modeling the MRT amounts to estimating the four functions $A_{\text{direct}}, A_{\text{scatter}}, \zeta_{\text{direct}}$, and ζ_{scatter} from experimental data. See [20] and [13] for a complete description of how these four functions can be measured with a simple calibration experiment.

The following MRT model was estimated from a calibration experiment, in uniform water medium, performed at Children's Hospital in Boston, MA, using the isotope Thallium (Tl). The parameters of this model are machine and isotope dependent

$$\mathcal{A}_{ji} = k \int_{B_j} \alpha(D, d, x) dx$$

where

$$\begin{aligned}
\alpha(D, d, x) &= e^{-0.19D} \frac{1}{\sqrt{2\pi\sigma^2_{\text{direct}}(d)}} e^{-x^2/2\sigma^2_{\text{direct}}(d)} \\
&\quad + (1.11e^{-0.095D} - e^{-0.19D}) \\
&\quad \frac{1}{\sqrt{2\pi\sigma^2_{\text{scatter}}(D)}} e^{-x^2/2\sigma^2_{\text{scatter}}(D)} \\
\sigma_{\text{direct}}(d) &= 0.0049d + 1.49 \\
\sigma_{\text{scatter}}(D) &= 0.037D + 2.53.
\end{aligned}$$

$d, D, \sigma_{\text{direct}}$, and σ_{scatter} are all measured in centimeter units.

III. Deformable Template Models

A. Initial Parameter Estimates

In describing the deformable template models, we will assume that both the isotope intensity previously denoted by X and the projection data Y are defined on the continuum, i.e., on the unit square I^2. The isotope intensity will typically be described by a piecewise smooth function $X(x), x \in I^2$, and the projection data by a function $Y(y), y \in I^2$. The MRT is also assumed to be defined on the continuum, $\mathcal{A} : I^2 \times I^2 \to \mathcal{R}$. *Note:* equation (1) is rewritten $E[Y(y)] = \int X(x)\mathcal{A}(x, y)dx$.

The isotope intensity within and immediately surrounding the tumor is modeled as a piecewise smooth function which has high values in a central region and low values in the area surrounding it. The prototype function used in our experiments is the standard bivariate Gaussian $F_t(x) = \exp\left[-\frac{1}{\sigma^2}(x_1^2 + x_2^2)\right]$. The isotope intensity in the scalp is modeled as a function $F_s(x)$ which has high values in a thin ellipsoid annulus corresponding to the scalp and zero everywhere else. The whole template thus has the form $F_t + F_s$. Observe that these two functions have basically disjoint supports.

Once a template is chosen, the function $X(x)$ describing the isotope intensity is assumed to be a deformation of the template obtained through composition with various planar mappings together with a magnitude adjustment. In other words,

$$X(x) = e^{\alpha} F_s(\Phi_s(x)) + e^{\mu} F_t(\Phi_t(x))$$

where Φ_t, Φ_s are mappings of the plane into itself. The function X itself needs only to be considered on the unit square which corresponds to the domain of the image.

If the mappings Φ_s and Φ_t are restricted to the set of affine mappings, the resulting function X will be a bivariate Gaussian with certain covariance and mean, surrounded by an ellipse with certain center and major and minor axes. In our experiments, Φ_s is always an affine map. Estimating Φ_s corresponds to estimating the center and axes.

More general types of smooth planar mappings, which modify shape, are applied only to the "hump" F_t. These transformations do preserve the unimodal character of the "hump." We will defer a more detailed description of these general types of deformations to Section III-B, and describe now the procedure for identifying the parameters of the affine mappings.

Estimating the Ellipse: The ellipse is given by the equation

$$\left(\frac{x_1 - x_{01}}{A}\right)^2 + \left(\frac{x_2 - x_{02}}{B}\right)^2 = 1.$$

Let the camera face that images the slice under consideration at the kth angle define the line $x_2 = m_k x_1 + b_k$.

Now project this ellipse onto the camera at angle k. A line perpendicular to the camera has slope $-1/m_k$, so the two lines tangent to the ellipse that are also perpendicular to the camera

pass through (x_1^+, x_2^+) and (x_1^-, x_2^-)

$$x_1^{\pm} = x_{01} \pm \frac{A}{\sqrt{A^2 + m_k^2 B^2}}$$
$$x_2^{\pm} = x_{02} \pm \frac{Bm_k}{\sqrt{A^2 + m_k^2 B^2}}.$$

The lines tangent to the ellipse and perpendicular to the camera have the form

$$x_2 = \frac{-1}{m_k} x_1 + \left(x_2^{\pm} + \frac{x_1^{\pm}}{m_k} \right).$$

These two lines intersect the camera at the two extremes of the projected ellipse. Thus, the outer edge of the ellipse projects onto the points

$$x_1 = \frac{m_k b_k - (m_k x_2^{\pm} + x_1^{\pm})}{m_k^2 + 1} \tag{6}$$

$$x_2 = \frac{2m_k^2 b_k - (m_k^2 x_2^{\pm} + m_k x_1^{\pm}) + b_k}{m_k^2 + 1}. \tag{7}$$

Estimating the parameters of the ellipse requires the identification of the left and right boundaries of the projected ellipse. Although this could doubtless be accomplished automatically, in the interest of expediency we perform this step by hand. Now it is possible to use a Gauss–Seidel type of iterative scheme to identify the parameters of the ellipse (x_{01}, x_{02}, A, and B) that minimize the squared distance between the observed body projection boundaries and the theoretical projection boundaries in (6) and (7).

Let $\hat{F}_s(x)$ be the indicator function of an annulus, one pixel wide, along the estimated ellipse. The intensity e^{μ} of the isotope in the scalp is calculated by minimizing

$$I_s(\mu) = \int \frac{1}{2Y(y)} |e^{\mu} \hat{Y}_s(y) - Y(y)|^2 dy$$

where Y_s is the function describing the projection $\int \mathcal{A}(x, y) \cdot \hat{F}_s(x) dx$ of $\hat{F}_s$. From this point on, the parameters of the ellipse and μ remain unchanged.

Affine Transformations of the "Hump": We introduce seven parameters $\alpha, \beta_1, \beta_2, \rho, \theta, \gamma_1, \gamma_2$. The two parameters β_1, β_2 correspond to a shift. The next represents a uniform scaling of the argument x by the factor e^{ρ}. The last three parameters describe a skewing matrix $\mathcal{S}$ as follows:

$$\mathcal{S}_{\gamma\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{\gamma_1} & 0 \\ 0 & e^{-\gamma_2} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
$$= \begin{pmatrix} e^{\gamma_1}\cos^2\theta + e^{-\gamma_2}\sin^2\theta & (e^{\gamma_1} - e^{-\gamma_2})\cos\theta\sin\theta \\ (e^{\gamma_1} - e^{-\gamma_2})\cos\theta\sin\theta & e^{\gamma_1}\sin^2\theta + e^{-\gamma_2}\cos^2\theta \end{pmatrix}$$

where θ describes the angle of the axes along which the skewing is done, e^{γ_1} describes the scaling along one axis, and $e^{-\gamma_2}$ is the scaling along the second. Thus, given values for the parameters, the suggested isotope intensity function is

$$X(x) = e^{\alpha} F_t(e^{\rho} \mathcal{S}_{\gamma\theta}(x - \beta)).$$

The parameters α and ρ are used as exponents in order to ensure a uniform interpretation of parameter values. Note that setting all parameters to zero yields the identity map, which leaves the template unchanged. Deviations of the same modulus to the left or to the right of the origin correspond to shrinking or magnifying by the same factor.

At this point, we estimate $\alpha, \beta, \rho, \theta, \gamma$ so as to maximize the likelihood of the observed data. Using the Gaussian approximation descrbied at the end of Section II-A, we need to minimize the following functional:

$$I(\alpha, \beta, \gamma, \rho, \theta) = \int \frac{1}{2Y(y)} \cdot \left| \int \left[e^{\mu} \hat{F}_s(x) + e^{\alpha} F_t(e^{\rho} \mathcal{S}_{\gamma\theta}(x - \beta)) \right] \cdot \mathcal{A}(x, y) dx - Y(y) \right|^2 dy.$$

Note that since the estimated ellipse is fixed, this functional can be rewritten as

$$I_t(\alpha, \beta, \gamma, \rho, \theta) = \int \frac{1}{2Y(y)} \left| \int e^{\alpha} F_t(e^{\rho} \mathcal{S}_{\gamma\theta}(x - \beta)) \mathcal{A}(x, y) dx - Y_t(y) \right|^2 dy \tag{8}$$

where $Y_t(y) = Y(y) - e^{\mu} \hat{Y}_s(y)$. Observe that only the data in the numerator are changed. The functional I_t is nonlinear in the parameters and depends on the structure of the template function F. All we can hope for is to find a local minimum close to the initial point.

Initially we preform a minimization of I_t in the shift parameter β only. If $\hat{\beta}$ is the point where the minimum is obtained, the new template is set to $F^1(x) = F_t(x - \hat{\beta})$. Subsequently, the magnitude parameter α is calculated together with the scaling parameter ρ which determines the variance of the Gaussian hump. Again, the template is updated as $F^2(x) = e^{\hat{\alpha}} F^1(e^{\hat{\rho}}(x - \hat{\beta}) + \hat{\beta}) = e^{\hat{\alpha}} F_t(e^{\hat{\rho}}(x - \hat{\beta}))$.

Finally, the skewing parameters are estimated to obtain a first estimate of the tumor intensity of the form

$$\hat{F}_t(x) = F^2(\mathcal{S}_{\hat{\gamma}\hat{\theta}} x) = e^{\hat{\alpha}} F_t(e^{\hat{\rho}} \mathcal{S}_{\hat{\gamma}\hat{\theta}}(x - \hat{\beta})).$$

One could of course plug the scaling parameter into the matrix $\mathcal{S}$ and minimize over all four parameters $(\rho, \theta, \gamma_1, \gamma_2)$. Experiments have shown that this does not perform as well as when the two minimization procedures are separated. When all three are done at once, the only parameter that really changes is the scale.

The final tumor intensity $\hat{F}_t$ is a bivariate Gaussian with mean given by $\hat{\beta}$ inverse covariance matrix $\mathcal{S}^2_{\hat{\gamma}\hat{\theta}}$. It is possible to do additional sweeps through all the parameters.

We have used the simplex algorithm provided in [18] for these minimizations. While it has proved sufficient for testing the validity of this approach, it is definitely not the most efficient method.

At this point, we have estimates $(\hat{\beta}, \hat{\rho}, \hat{\theta}, \hat{\gamma})$ of the location, the size, and the orientation of the tumor. Finer details regarding the shape of the tumor can now be obtained in a more general setup as described in the next subsection.

B. Estimation of Planar Deformations

Restoration via planar deformations requires choosing an initial template F. This could be the initial template F_t of the previous section; it could be the restoration $\hat{F}_t$ of the initial parameter estimates applied to F_t; or it could be the result of a "soft modeling" restoration.

The template is then transformed through composition with planar deformations which, in effect, describe the deformations of the level curves of F. In other words, the restoration will have the form $X(x) = F(\phi(x)) = F(x + U(x))$, where $\phi(x)$ is the planar mapping and $U(x) = \phi(x) - x$ is the displacement field. Thus, if η and ω are the level curves corresponding to the same value of F and X, respectively, then $\omega = \phi^{-1}(\eta)$. Moreover, no new levels can be created by this type of transformation of the template. Heuristically, one can say that the entire configuration of level curves of F is preserved. Thus, taking the Gaussian hump as a template means that we expect the intensity function of the tumor to consist of simply connected level curves. Each level curve is expected to have one component only, and level curves of lower value contain those of higher value. Additional information regarding the topography of the tumor intensity function would lead us to more specific templates.

The likelihood term parameterized by the field U has the form

$$I(U) = \int \frac{1}{2Y(y)} \left| \int F(x + U(x)) \cdot \mathcal{A}(x, y)dx - Y_t(y) \right|^2 dy. \tag{9}$$

Recall that Y_t is obtained from the data by removing the estimate of the scalp intensity.

Since U is an infinite-dimensional parameter, it must be approximated to allow for actual calculations. Let $\psi_k(x), k = 1, \cdots, \infty$ be any orthonormal basis of $L_2(I^2)$. Write $U_L(x) = \sum_{k=1}^{L}(u_{1,k}, u_{2,k})\psi_k(x)$ and define

$$I_L(u_1, u_2) = \int \frac{1}{2Y(y)} \left| \int F(x + U_L(x)) \cdot \mathcal{A}(x, y)dx - Y_t(y) \right|^2 dy.$$

Now the minimization is in the $2L$ parameters u_1, u_2, which are the coefficients of U_L with respect to the chosen basis.

It is well known that in the estimation of infinite-dimensional parameters, it is necessary to introduce some smoothing mechanism to constrain the solution to lie within an acceptable subspace of the parameter space. In the Bayesian framework, a prior distribution is defined on the space of parameters so that mass is concentrated around the acceptable subspace. The posterior distribution conditioned on the data is then minimized. Assuming independent Gaussian prior distributions on each of the coefficients with mean zero and variance λ_k, the posterior is proportional to

$$\hat{I}_L = \sum_{k=1}^{L}(u_{1,k}^2 + u_{2,k}^2)/\lambda_k + I_L(u_1, u_2).$$

This is equivalent to introducing a quadrtic smoothing term.

The basis chosen for this algorithm was a basis of compactly supported wavelets (see Daubechies [3]). There were several reasons for this choice. First, the wavelet basis is constructed in such a way that the elements with low indexes have a large support and therefore allow for large-scale global adjustments in the field U. The elements with higher indexes have smaller support and hence allow for local adjustments. The second reason is that the wavelet transforms in the compact case are very fast [$o(n)$] (see [3] and [12]). Each time U_L is calculated from (u_1, u_2), such a transform is being computed. The third reason is that there is a very precise relationship between the rate of decay of the coefficients in the expansion of a function and the degree of smoothness of that function. This can be exploited in order to set the variances or smoothing weights λ_k. For a more detailed discussion of this issue, see [1].

In previous applications of planar deformations for image matching, the posterior distribution or smoothed likelihood was minimized through a gradient descent algorithm. The gradient had an analytical expression which was easily computed using forward and inverse transforms (see [1] and [2]). In the present setting, it is also possible to write an expression for the gradient.

$$\begin{aligned}\frac{\partial \hat{I}(u_1, u_2)}{\partial u_{i,k}} &= u_{i,k}/\lambda_k \\ &\quad + \int \left\{ \frac{1}{Y(y)} \left[\int F(x + U_L(x)) \mathcal{A}(x, y)dx - Y_t(y) \right] \right. \\ &\quad \times \int \frac{\partial F}{\partial x_i}(x + U_L(x)) \\ &\quad \left. \mathcal{A}(x, y)\psi_k(x)dx \right\} dy\end{aligned}$$

where $U_L(x) = \sum_{k=1}^{L}(u_{1,k}, u_{2,k})\psi_k(x)$.

Changing the order of integration and integrating first in y, we set

$$B(x, z) = \int \mathcal{A}(x, y)\mathcal{A}(z, y)/Y(y)dy$$

and

$$C(z) = \int \mathcal{A}(z, y)Y_t(y)/Y(y)dy.$$

Both $B(x, z)$ and $C(z)$ can be calculated off line for a given data set. The gradient then has the form

$$\begin{aligned}\frac{\partial \hat{I}(u_1, u_2)}{\partial u_{i,k}} &= u_{i,k}/\lambda_k \\ &\quad + \int \int B(x, z)F(x + U_L(x)) \cdot \frac{\partial F}{\partial z_i}(z + U_L(z))\psi_k(z)dxdz \\ &\quad - \int C(z)\frac{\partial F}{\partial z_i}(z + U_L(z))\psi_k(z)dz \\ &= u_{i,k}/\lambda_k + \int_{I^2} [BF(z) - C(z)] \\ &\quad \cdot \frac{\partial F}{\partial z_i}(z + U_L(z))\psi_k(z)dz\end{aligned} \tag{10}$$

where $BF(z) = \int_{I^2} B(x,z)F(x + U_L(x))dx$. Observe that calculating U_L is a backward wavelet transform, whereas calculating the last integral in the equation above is a forward transform of $[BF(z) - C(z)]\frac{\partial F}{\partial z_i}(z + U_L(z))$. Both transforms can be done with great efficiency using the algorithm described in Mallat [12].

The restoration algorithm computes a discrete time version of

$$\frac{du_{i,k}(t)}{dt} = -\frac{\partial \hat{I}(u_1, u_2)}{\partial u_{i,k}}$$

for $i = 1, 2$ and $k = 1, \cdots, L$. For fixed L, the steps of the algorithm are as follows.

1) Set $u_1(0) = 0, u_2(0) = 0$. Set time step dt.

2) Calculate the field U_L from $(u_1(t), u_2(t))$—inverse wavelet transform.

3) Calculate $F(x + U_L(x))$ for each point on the lattice.

4) Calculate $BF(z) = \int B(x,z)F(x + U_L(x))dx$ for each point on the lattice.

5) Calculate $\frac{\partial F}{\partial z_i}(z + U_L(z)), i = 1, 2$ for each point on the lattice.

6) Calculate WBF_i the wavelet transform of $[BF(z) - C(z)]\frac{\partial F}{\partial z_i}(z + U_L(z)), i = 1, 2$.

7) Update the wavelet coefficients

$$u_{i,k}(t+1) = u_{i,k}(t) - dt \cdot [(u_{i,k}(t)/\lambda_k + WBF_{i,k}],$$

for $i = 1, 2$ and $k = 1, \cdots, L$.

The derivatives of F are calculated using a symmetric difference scheme and stored in an array at the start of the program. The integration in step 4 is done by quadrature over a lattice.

IV. Experiments

We carried out two sets of experiments with simulated data. For each set of experiments, a tumor intensity function $\tilde{Y}_t$ was generated (see Fig. 2). The projection was calculated using the matrix $\mathcal{A}$ which was estimated for the SPECT machine at the Children's Hospital, as described in Section II-B. An ellipse corresponding to the intensity in the scalp was generated (see Section III-A), and its projection $\tilde{Y}_s$ was calculated with the same $\mathcal{A}$. The intensities of both the tumor and the ring were initially set on a scale between 0 and 1. The data Y which served as input to the algorithm was Poisson distributed with mean

$$E[Y] = S \cdot \tilde{Y}_s + T \cdot \tilde{Y}_t$$

where S and T are constants which control the quality of the data. Larger values for the constants S and T resulted in higher signal-to-noise ratios and hence cleaner data. Each set of experiments involved running three reconstruction algorithms for a sequence of decreasing values of S and T to determine when various methods break down. The three reconstruction methods used in these experiments are the ICM algorithm (see [20] and [13]), the parametric estimator (Section III-A), and the deformable template matcher (Section III-B). These synthesized data sets are displayed in 3a and 4a, together with the corresponding values of S and T.

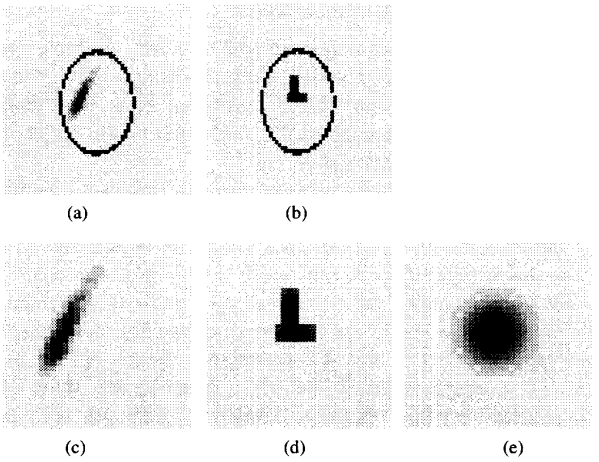

Fig. 2. (a) Simulated tumor A with scalp. (b) Simulated tumor B with scalp. (c) Simulated tumor A, enlarged. (d) Simulated tumor B, enlarged. (e) Gaussian "hump" used as a deformable template. Note: dark represents largest pixel intensity; light represents smallest pixel intensity.

TABLE I
Parameter Estimates and L_2 Distance to True Tumor

	Parameter estimates					L_2 distance		
Experiment	$\hat{T}$	$\hat{S}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	Soft	Para.	Def. Tem.
Simulation A								
$T=84, S=21$; fig 3	51.5	23.1	0.039	0.183	0.771	0.012	0.0064	0.0041
$T=21, S=5$; fig 4	9.4	5.4	0.039	0.176	0.862	0.012	0.0086	0.0075
$T=10, S=4$; fig 5	4.2	4.0	0.035	0.153	0.913	0.014	0.014	0.0087
$T=4, S=4$; fig 6	1.4	4.0	0.011	0.031	0.966	0.029	0.025	0.024
Simulation B								
$T=84, S=21$; fig 7	74.8	20.4	0.084	0.038	0.742	0.0081	0.0084	0.0059
$T=41, S=10$; fig 8	32.9	9.4	0.067	0.031	0.762	0.010	0.010	0.0071
$T=21, S=5$; fig 9	12.8	4.8	0.089	0.037	0.858	0.010	0.011	0.0073
$T=10, S=4$; fig 10	7.7	3.8	0.077	0.037	0.873	0.012	0.011	0.017
Patient A; fig 11	3.2	5.4	-0.007	-0.021	0.904	True tumor shape in not known.		
Patient B; fig 12	3.3	3.5	-0.004	-0.032	0.926	True tumor shape in not known.		

The following parameter estimates were obtained from the parametric reconstruction method: $\hat{T}$ = tumor amplitude; $\hat{S}$ = scalp amplitude; $\hat{\beta}_1$ = shift in horizontal direction; $\hat{\beta}_2$ = shift in vertical direction; $\hat{\rho}$ = scale. (The other parameters $\gamma_1, \gamma_2, \theta$ introduced in Section 4.1 were estimated, but their values are not useful for comparative purposes in these experiments.) The L_2 distance is a measure of the effectiveness of each of the three methods.

The numerical results of applying the various algorithms to the simulated data are presented in Table I. An L_2 measure of distance used to determine the relative performance of each of the three reconstruction methods is presented. The distance is

$$L_2 = \sum_i \left(\frac{X_i}{\sum_j X_j} - \frac{\hat{X}_i}{\sum_j \hat{X}_j} \right)^2$$

where X is the true tumor intensity and $\hat{X}$ is the reconstructed tumor intensity. Clearly, as the noise level increases, the reconstructions deteriorate.

The images in Figs. 3–6 show some actual restorations. The first two use synthetic data, the second two use data sets obtained from the Children's Hospital. The signal-to-noise ratio in the real data sets is very low and corresponds roughly to the worst cases presented in the synthesized experiments listed in Table I. At this level of noise, it is difficult to assess

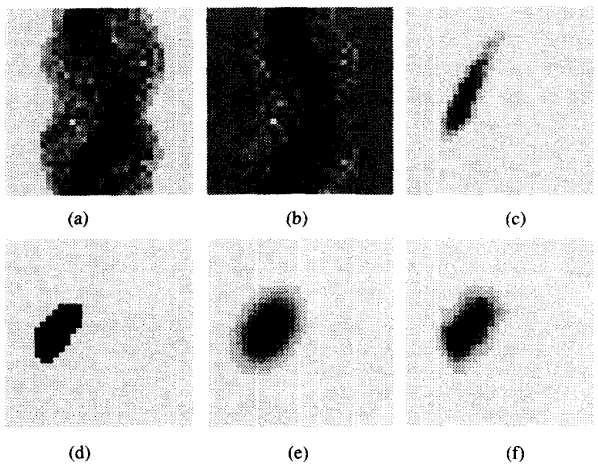

Fig. 3. Reconstruction experiments for simulated tumor A, with $T = 21$ (tumor intensity) and $S = 5$ (scalp intensity). (a) Sinogram, 7179 total counts. (b) Sinogram after removing projections of estimated scalp. (c) True tumor shape. (d) Soft model reconstruction. (e) Parametric reconstruction. (f) Deformable template reconstruction. Note: dark represents largest pixel intensity; light represents smallest pixel intensity.

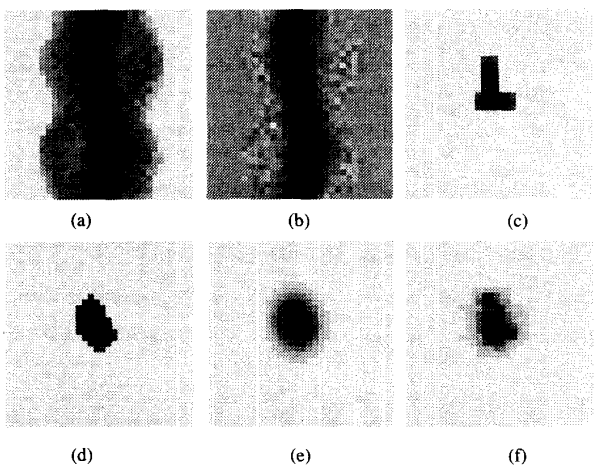

Fig. 4. Reconstruction experiments for simulated tumor B, with $T = 84$ (tumor intensity) and $S = 21$ (scalp intensity). (a) Sinogram, 30502 total counts. (b) Sinogram after removing projections of estimated scalp. (c) True tumor shape. (d) Soft model reconstruction. (e) Parametric reconstruction. (f) Deformable template reconstruction. Note: dark represents largest pixel intensity; light represents smallest pixel intensity.

the various techniques, especially since the actual shape of the tumors is not known.

The parameter values for the prior distribution used with the ICM algorithm were $\delta = 0.2$ (20 percent of the dynamic range) and $\beta = 0.01$ (a relatively mild level of smoothness). The ICM restorations are shown in the (d) panels.

The first step in both the parametric and planar deformation reconstruction procedures was to estimate the projection data generated by the scalp (i.e., $S \cdot \tilde{Y}_x$) and subtract it from the data *in the numerator* of the likelihood term, as in (8). As indicated in Section III-A, the functional I_t was minimized using the simplex algorithm provided in [18]. Although not very efficient, this method is very easy to implement.

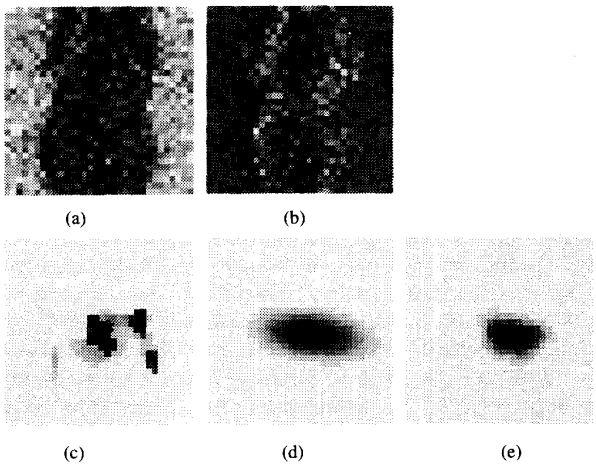

Fig. 5. Reconstruction of patient study A. (a) Sinogram, 6536 total counts. (b) Sinogram after removing projections of estimated scalp. (c) Soft model reconstruction. (d) Parametric reconstruction. (e) Deformable template reconstruction. Note: dark represents largest pixel intensity; light represents smallest pixel intensity.

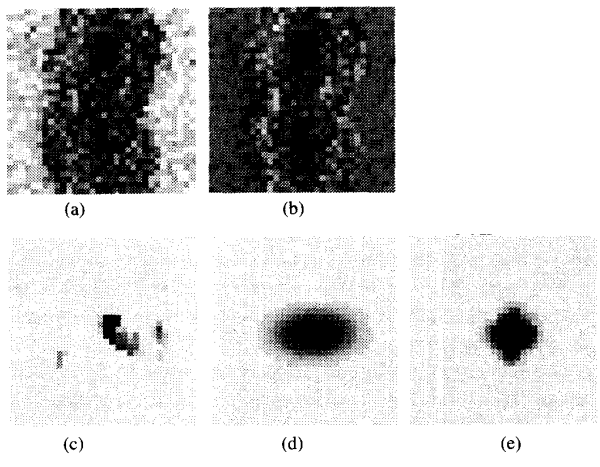

Fig. 6. Reconstruction of patient study B. (a) Sinogram, 4493 total counts. (b) Sinogram after removing projections of estimated scalp. (c) Soft model reconstruction. (d) Parametric reconstruction. (e) Deformable template reconstruction. Note: dark represents largest pixel intensity; light represents smallest pixel intensity.

The template used in the parametric reconstructions presented here is the Gaussian "hump" centered at the middle of the unit square, i.e., at the point (0.5,0.5), see Fig. 2(e). Since the support of the tumor is not expected to lie outside the square [0.25,0.75] $\times$ [0.25,0.75], the integrals in (8) need to be calculated only in that domain. The matrix $\mathcal{A}$ described in Section II-B is calculated on a 64 $\times$ 64 lattice for the x variable and a 64 $\times$ 64 lattice for the y varialbe; however, restricting the tumor to the inner square allows us to use only the inner 32 $\times$ 32 lattice for the x variable. Moreover, since $\mathcal{A}$ is smooth, we sampled it at every second point in both variables to obtain a reduced 16 $\times$ 16 lattice in the x variable and a 32 $\times$ 32 lattice in the y variable.

Clearly, some information is lost in this sampling procedure, especially in the case of a high signal-to-noise ratio. However,

for the sake of experimentation with these methods, the sampled $\mathcal{A}$ proved sufficient and allowed us to run the algorithm at a reasonable memory size!

The (e) panels of Figs. 3–6 present the parametric reconstructions of the data simulated for these experiments. The estimated parameters are listed in Table I. Each such restoration takes 1 min on the SUN SPARC IPX.

The reconstructions using planar deformations were also done using the Gaussian template. The only input used from the parametric reconstructions was the amplitude parameter $e^{\alpha} = T$. For each such reconstruction, it was necessary to precompute the B and C arrays described in Section III-B. Each such calculation takes 2 min on a SPARC 2. The size of the B array is 16×16 in the x and y variables, producing a data set of 16^4 floating-point entries. Observe that using the full $\mathcal{A}$ matrix without the sampling described above would produce an array B of 32^4 floating-point entries, i.e., a 4 Mbyte size array. Since each iteration in the algorithm requires the calculation of BF [see (10)] using the full $\mathcal{A}$ would be at least 16 times slower.

The (f) panels of Figs. 3–6 present the reconstructions using the deformable template method. Based on the images generated from the experiments with synthetic data, it is evident that the planar deformation reconstruction method outperforms ICM and the parametric model reconstruction methods.

V. Conclusion

In a SPECT imaging session, effects of Poisson noise, photon attenuation and scatter, and the collimator effect combine in such a way that the reconstruction problem for SPECT is ill-posed. Because of the nature of the problem, it is not surprising that some form of regularization must be used when reconstructing isotope intensity maps.

Several regularization techniques are used in SPECT reconstruction algorithms. In the case of filtered backprojection, low-pass filters are applied either before or after (or both) backprojecting the data. In the maximum likelihood approach, the EM algorithm is prematurely halted or is smoothed using a sequence of sieves. In the case of Bayesian and penalized maximum likelihood approaches, images which lack local smoothness are assigned a penalty term.

The deformable template models described here incorporate smoothing by the very nature of the template. In addition, they incorporate assumptions regarding various topological features of the intensity function.

The experiments presented here indicate that in cases where a template can be defined, reconstructing using the proposed deformable template algorithm will yield more precise results than the general-purpose "soft modeling" algorithms. In addition, this method provides a mapping between the template and the reconstructed image. This mapping provides an automatic method of locating any landmark in the reconstructed image by simply identifying that landmark in the original template image. Finally, the parametric algorithm provides direct estimates for certain parameters of clinical interest.

References

[1] Y. Amit, "A non-linear variational problem for image matching," SIAM, *J. Sci. Comp.*, to appear. (Tech. Rep., Div. Appl. Math., Brown Univ., 1991.)

[2] Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *J. Amer. Stat. Assoc.*, vol. 86, pp. 376–387, 1991.

[3] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. XLI, pp. 909–996, 1988.

[4] D. Geman, "Random fields and inverse problems in imaging," *Lect. Notes Math.*, p. 1427, 1990.

[5] D. Geman, S. Geman, and D. E. McClure, "Improved FLIR image segmentation by iterative techniques: Algorithms for real-time processing, Tech. Rep., Mathematical Technologies Inc., 1989.

[6] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," IEEE PAMI, vol. 14, no. 3, 1992. (Tech. Rep., Dep. Math. Stat., Univ. Mass. Amherst, 1990.)

[7] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," in *1985 Proc. Stat. Comput. Section, Amer. Stat. Assoc.*, 1985, pp. 12–18.

[8] ——, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. 52, pp. 5–21, 1987.

[9] S. Geman, D. E. McClure, and D. Geman, "A nonlinear filter for film restoration and other problems in image processing," Tech. Rep., Div. Appl. Math., Brown Univ., 1990.

[10] U. Grenander and M. Miller, "Jump-diffusion processes for abductions and recognition of biological shapes, Tech. Rep., Electron. Signals Syst. Res. Lab., Washington Univ., 1991.

[11] A. P. Knoerr, "Global models of natural boundries—Theory and applications," Ph.D. dissertation, Div. Appl. Math., Brown Univ., 1988.

[12] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-11, pp. 674–693, 1989.

[13] K. M. Manbeck, "Bayesian statistical methods applied to physical phantom and patient data," Ph.D. dissertation, Div. Appl. Math., Brown Univ., 1990.

[14] ——, "Hubble telescope image restoration by statistical methods," Tech. Rep., Div. Appl. Math., Brown Univ., 1990.

[15] ——, "On Gaussian approximations to the Poisson distribution in image processing," Tech. Rep. Pattern Theory, 157, Div. Appl. Math., Brown Univ., 1992.

[16] M. I. Miller and B. Roysam, "Bayesian image reconstruction for emission tomography incorporating Good's roughness prior on massively parallel processors," *Proc. Nat. Acad. Sci.*, vol. 88, pp. 3223–3227, 1991.

[17] P. Dupuis, D. Geman, J. Horowitz, and G. Reynolds, "Statistical inference on the shape of circumstellar disks from hst observations," Tech. Rep., Dep. Math. Stat., Univ. Mass., Amherst, 1991.

[18] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge University Press, 1988.

[19] B. D. Ripley, "Recognizing organisms from their shapes—A case study in image analysis," Tech. Rep., Univ. Strathclyde, Glasgo, 1990.

[20] S. Geman, K. M. Manbeck, and D. E. McClure, "A comprehensive statistical model for single photon emission tomography," in *Markov Random Fields: Theory and Applications,* R. Chellappa and A. Jain, eds. Boston: Academic, 1993.

[21] D. L. Snyder, D. G. Politte, and M. I. Miller, "A case study in statistical image processing: Positron-emission tomography," in *Spatial Statistics and Imaging,* A. Possolo, ed. IMS, Lecture Notes, vol. 20, 1990, pp. 369–381.

[22] U. Grenander, Y. Chow, and D. M. Keenan, *Hands: A Pattern Theoretical Study of Biological Shape.* New York: Springer-Verlag, 1991.