Maximum Entropy and Bayesian Methods, in press, 1996

MECHANICAL MODELS AS PRIORS IN BAYESIAN TOMOGRAPHIC RECONSTRUCTION

ANAND RANGARAJAN
Departments of Diagnostic Radiology and Computer Science
Yale University
New Haven, CT 06520-8042[§]

AND

SOO-JIN LEE AND GENE GINDI
Departments of Radiology and Electrical Engineering
SUNY at Stony Brook
Stony Brook, NY 11794-8460[¶] [‖]

Abstract. We introduce a new prior—the *weak plate*—to Bayesian tomographic reconstruction. The weak plate captures the piecewise ramplike spatial structure evident in primate autoradiograph source distributions. The weak plate is a part of a family of "mechanical" models—weak membrane (1st order), weak plate (2nd order), and weak quadric (3rd order)—in which a class of smoothness constraints derived from properties of ideal physical materials are used as models in the associated reconstruction problem. Since "weak" priors generate local minima in MAP estimation, we have designed novel Generalized Expectation–Maximization deterministic annealing algorithms to alleviate this problem. Our simulation studies qualitatively demonstrate the improvements over the weak membrane and maximum likelihood reconstructions.

Key words: tomographic reconstruction, weak plate, piecewise smooth, deterministic annealing, expectation–maximization, free energy, saddle point.

1. Introduction

Maximum likelihood (ML) tomographic reconstruction [1] is a very illustrative example of the "overfitting" problem in statistics. Given emission projection data—which can roughly be visualized as line integrals of a 2-D source distribution taken at different angles—the problem is to reconstruct the underlying source distribution (image). Unfortunately, the desired image resolution dictates that we esti-

§Email: anand@noodle.med.yale.edu
¶Email: sjlee@clio.rad.sunysb.edu, gindi@clio.rad.sunysb.edu
‖This work was supported by NIH R01-NS32879.

mate N parameters from approximately the same amount of data leading to the aforementioned overfitting problem [2]. Bayesian tomographic reconstruction is a natural way of specifying an *effective* number of parameters that is much less than the original number. From the Bayesian viewpoint, this leads quite naturally to the question: what is a good prior?

Priors can be seen as means of incorporating actual, known information regarding the local spatial character of the source. Several such prior distributions have been proposed in the literature [3–5]. In these proposals, the source was assumed to be piecewise smooth. Our own early work incorporated one such prior, a "weak membrane" (WM) model, to capture notions of piecewise smoothness [5,6]. Most of the simulation experiments were done with the mathematical phantoms which were themselves piecewise constant. Reconstruction algorithms, including our own, that use this kind of piecewise constant assumption as prior perform better, according to most reasonable metrics, than those that make no such assumption, when tested on such phantoms. However, the self-consistent loop of piecewise constant prior and piecewise constant phantom leads one to question whether these good results generalize to a realistic clinical setting where the underlying (patient) source distribution may not be piecewise constant.

Here we argue for more expressive priors able to model more complicated forms than a piecewise constant source. Our work is motivated by the observations of piecewise ramplike spatial structure in primate autoradiographs [7]. The essential feature of our new prior model, the *weak plate* (WP) is that it favors piecewise *linear* ramplike regions. The weak plate is a part of a family of "mechanical" models—weak membrane (1st order), weak plate (2nd order), and weak quadric (3rd order)—in which a class of smoothness constraints derived from properties of ideal physical materials are used as models in the associated reconstruction problem [8]. Smoothness priors often correspond to a notion of fitting a member of a class of smooth functions to the data; "weak" constraints refer to the allowed inclusion of spatial discontinuities in the fit. Thus a weak membrane is a low order (1st derivative) spline allowed to have breaks (0th-order discontinuities) and a weak plate is the familiar (2nd-order) thin plate spline that is allowed to have breaks and creases (discontinuities in 0th and 1st derivatives).

2. Bayesian Tomographic Reconstruction

We model our problem on a 2-D discrete lattice indexed by i, j. Lowercase bold quantities denote 2-D vector fields and the corresponding lowercase, italicized quantities denote the elements of the vector field. Similarly, uppercase bold quantities denote 2-D random fields. Thus, $\Pr(\mathbf{F} = \mathbf{f})$ denotes the probability that the random field $\mathbf{F}$ takes the value $\mathbf{f}$ and $f_{i,j}$ denotes an element of $\mathbf{f}$ at location (i, j).

To enforce the notion of *piecewise* smoothness, Geman and Geman [9] introduced an unobservable line process $\mathbf{l}$ into the image model to preserve the discontinuities in the reconstructions, and versions of the line processes have been proposed for medical imaging in [4–6]. We begin by formulating our reconstruction problem from Bayes' Theorem with the aid of the line processes:

$$\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) = \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})}{\Pr(\mathbf{G}=\mathbf{g})},$$

where $\mathbf{f}$, $\mathbf{l}$, and $\mathbf{g}$ are the source intensities, line processes, and projection data, respectively, and $\mathbf{F}$, $\mathbf{L}$, $\mathbf{G}$ are the associated 2-D random fields. We require the maximum *a posteriori* (MAP) estimate of both the source field and the discontinuity field $(\hat{\mathbf{f}},\hat{\mathbf{l}})$: $(\hat{\mathbf{f}},\hat{\mathbf{l}}) = \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})$.

We model both the likelihood and the prior as Gibbs distributions [9]. A Gibbs distribution is specified by an energy function E from which the partition function Z can be obtained. Since the number of detected counts in emission tomography (ET) is independently Poisson distributed, we model the likelihood as

$$\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) = \prod_{t\theta} \frac{(\sum_{ij}\mathcal{H}_{t\theta,ij}f_{i,j})^{g_{t\theta}}\exp(-\sum_{ij}\mathcal{H}_{t\theta,ij}f_{i,j})}{g_{t\theta}!}. \qquad (1)$$

In (1), $g_{t\theta}$ is the number of detected counts in the detector bin indexed by t at angle θ, and $\mathcal{H}_{t\theta,ij}$ is the probability that a photon emitted from source location (i,j) hits detector bin t at angle θ. The major physical factors can be adequately modeled as linear effects and summarized by the system matrix $\mathcal{H}_{t\theta,ij}$.

The likelihood energy in (1) is

$$E_D(\mathbf{f}) = -\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) = \sum_{t\theta}\left[-g_{t\theta}\log\sum_{ij}\mathcal{H}_{t\theta,ij}f_{i,j}\right] + \sum_{t\theta,ij}\mathcal{H}_{t\theta,ij}f_{i,j},$$

where the term $\log g_{t\theta}!$ was dropped since it does not involve $\mathbf{f}$. The prior is defined over the source intensities and line processes. The corresponding prior energy is $E_P(\mathbf{f},\mathbf{l}) = -\log\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})$.

2.1. THE WEAK PLATE AND THE WEAK MEMBRANE PRIORS

The energy function corresponding to the WM prior is a function of the source $\mathbf{f}$, and the line processes $\mathbf{l}$:

$$E_P^M(\mathbf{f},\mathbf{l}) = \lambda\sum_{ij}\left[f_v^2(i,j)(1-l_{i,j}^h) + f_h^2(i,j)(1-l_{i,j}^v)\right] + \alpha\sum_{ij}\left(l_{i,j}^h + l_{i,j}^v\right). \qquad (2)$$

In (2), f_v and f_h are the first order derivatives along the vertical and the horizontal directions, respectively, and are defined as $f_v(i,j) \stackrel{\text{def}}{=} f_{i+1,j} - f_{i,j}$ and $f_h(i,j) \stackrel{\text{def}}{=} f_{i,j+1} - f_{i,j}$. The binary variables l^v and l^h are vertical and horizontal line processes, respectively, and λ and α are positive hyperparameters. Note that if $\mathbf{l}$ is set to zero everywhere, the prior reduces to a familiar smoothness regularizing term penalizing the first squared derivatives. The last term penalizes the creation of the discontinuities, charging an amount α at each such site. Therefore, the two terms in (2) encourage smoothness except where discontinuities ($l_{i,j} = 1$) occur.

However, due to its nature in favoring piecewise constant reconstructions, the WM prior has the unfortunate effect of turning a ramp into a single step or stepped terraces depending on the parameter settings. This is the “gradient limit” effect [10], which is a fundamental limitation of the WM (and of many first order Markov models [11] as well). If the gradient of a ramp exceeds the limit, a discontinuity appears in its reconstruction. If the gradient is much greater than the limit, a solution with multiple breaks might have the smallest energy. The gradient limit is inversely proportional to the scale λ. This disadvantage of the WM leads to significant errors in tomographic reconstructions as shown in our simulations.

While the WM prior favors the piecewise constant reconstructions, the WP prior favors piecewise linear reconstructions. Our corresponding energy function for the WP is [10]:

$$E_P^P(\mathbf{f},\mathbf{l}) = \lambda \sum_{ij} \left\{ \left[f_{vv}^2(i,j) + 2f_{hv}^2(i,j) + f_{hh}^2(i,j) \right] (1 - l_{i,j}) \right\} + \alpha \sum_{ij} l_{i,j}, \quad (3)$$

where our discrete approximations of the partial second derivatives are $f_{vv}(i,j) \stackrel{\text{def}}{=} f_{i+1,j} - 2f_{i,j} + f_{i-1,j}$, $f_{hh}(i,j) \stackrel{\text{def}}{=} f_{i,j+1} - 2f_{i,j} + f_{i,j-1}$, $f_{hv}(i,j) \stackrel{\text{def}}{=} f_{i,j} - f_{i+1,j} - f_{i,j+1} + f_{i+1,j+1}$. Note that this form has the line processes $l_{i,j}$, associated with the six nodes, $f_{i-1,j}$, $f_{i,j-1}$, $f_{i,j}$, $f_{i,j+1}$, $f_{i+1,j}$, and $f_{i+1,j+1}$. (An advantage of mechanical models written in a functional form as in (2) and (3) is that Markov models with appropriate neighborhoods are automatically generated upon discretization.)

If $\mathbf{l}$ is set to zero everywhere, the prior reduces to the thin plate spline smoothing functional However, unlike the WM which encourages only the formation of piecewise constant regions, the WP encourages smoothness even in ramplike regions without incurring a penalty. That is, a “crease”, a discontinuity in the first derivative, will turn on the line process. Thus the discontinuities in the WP correspond to discontinuities in the source gradient in addition to those in the source itself.

The posterior energy functions can be written as

$$E^P(\mathbf{f},\mathbf{l}) = E_D(\mathbf{f}) + E_P^P(\mathbf{f},\mathbf{l}),\ E^M(\mathbf{f},\mathbf{l}) = E_D(\mathbf{f}) + E_P^M(\mathbf{f},\mathbf{l}) \quad (4)$$

where the WM and WP energy functions are defined in (2) and in (3) respectively.

2.2. MAP ESTIMATION VIA DETERMINISTIC ANNEALING

Minimization of the WP and WM energy functions is difficult due to the presence of the binary valued line processes $\mathbf{l}$. In this section, we present a deterministic annealing method to minimize the above energy functions. While deterministic annealing is not guaranteed to find the global minimum, it tends to avoid poor local minima.

Deterministic annealing methods begin by minimizing the *free energy* at high temperature $(\frac{1}{\beta})$ and then tracking the minimum through the variation of the temperature parameter. At high temperatures, the free energy is nearly convex and easily minimized. The free energy $F(\beta)$ is equal to $-\frac{1}{\beta}\log Z(\beta)$ where Z is the partition function. Since the partition function integral is intractable, we employ the well known *saddle-point* approximation from statistical physics. This is briefly summarized here—for more details, see [12].

$$\begin{aligned}
Z(\beta) &= \sum_{\{\mathbf{f},\mathbf{l}\}} \exp[-\beta E(\mathbf{f},\mathbf{l})] = \sum_{\{\mathbf{f},\mathbf{l}\}} \prod_{ij} \int_R dz_{i,j}\, \delta(l_{i,j} - z_{i,j}) \exp[-\beta E(\mathbf{f},\mathbf{z})] \\
&= \sum_{\{\mathbf{f},\mathbf{l}\}} \int_R dz_{i,j} \int_I du_{i,j} \exp \sum_{ij} u_{i,j}(l_{i,j} - z_{i,j}) \exp[-\beta E(\mathbf{f},\mathbf{z})] \\
&= \sum_{\{\mathbf{f}\}} \int_R dz_{i,j} \int_I du_{i,j} \exp \sum_{ij} [-u_{i,j} z_{i.j} + \log(1 + \exp u_{i,j})] \exp[-\beta E(\mathbf{f},\mathbf{z})]
\end{aligned}$$

where $\mathbf{z}$ is the *analog* line process; $\mathbf{z} \in [0,1]^N$ [13]. The original MAP combinatorial optimization problem on the binary valued line processes $\mathbf{l}$ has been converted into a nonlinear optimization problem on the continuous valued analog line processes $\mathbf{z}$ through the application of deterministic annealing [14]. Approximating the integral by the value of the integrand at its saddle points, we get

$$\begin{aligned}
F(\beta) &\approx -\frac{1}{\beta} \log Z(\beta, \hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{u}}) = \min_{\mathbf{f},\mathbf{z}} \max_{\mathbf{u}} E(\mathbf{f},\mathbf{z}) + \frac{1}{\beta} \sum_{ij} [u_{i,j} z_{i,j} - \log(1 + \exp u_{i,j})] \\
&= \min_{\mathbf{f},\mathbf{z}} E(\mathbf{f},\mathbf{z}) + \frac{1}{\beta} \sum_{ij} [z_{ij} \log z_{ij} + (1 - z_{ij}) \log(1 - z_{ij})] . \qquad (5)
\end{aligned}$$

From (5), it can be seen that at high temperature (low β), the entropy term $(\sum_{ij} [z_{ij} \log z_{ij} + (1 - z_{ij}) \log(1 - z_{ij})])$ convexifies the free energy. At very low temperatures (high β), the entropy term drops out leaving us with the original MAP energy function.

One last detail needs to be resolved before we can summarize the deterministic annealing (DA) algorithm. The global connectivity of the likelihood term precludes the use of standard descent methods. The Expectation–Maximization (EM) algorithm (with the familiar complete/incomplete data formulation) is usually pressed into service to alleviate this problem. This approach can be extended to the MAP estimation problem as well. Below, we take a somewhat idiosyncratic approach in presenting the EM algorithm as coordinate descent on a new likelihood objective function [15]:

$$\begin{aligned}
E_D(\mathbf{c}, \mu, \mathbf{f}) &= \sum_{t\theta,ij} [\mathcal{H}_{t\theta,ij} f_{i,j} - c_{t\theta,ij} \log \mathcal{H}_{t\theta,ij} f_{i,j}] \\
&+ \sum_{t\theta} \mu_{t\theta} (\sum_{ij} c_{t\theta,ij} - g_{t\theta}) + \sum_{t\theta,ij} c_{t\theta,ij} (\log c_{t\theta,ij} - 1) \qquad (6)
\end{aligned}$$

where $\mathbf{c}$ is associated with the complete data and μ is a Lagrange parameter satisfying the constraint $\sum_{ij} c_{t\theta,ij} = g_{t\theta}$. The EM algorithm results when (6) is minimized in two stages—first w.r.t. $(\mathbf{c}, \mu)$ and then w.r.t. $\mathbf{f}$. The same approach can be extended to cover the MAP case with either the WP or the WM energy functions:

$$E(\mathbf{c}, \mu, \mathbf{f}, \mathbf{z}) = E_D(\mathbf{c}, \mu, \mathbf{f}) + E_P(\mathbf{f}, \mathbf{z}) + \frac{1}{\beta} \sum_{ij} [z_{ij} \log z_{ij} + (1 - z_{ij}) \log(1 - z_{ij})] .$$

We apply a coordinate descent method to minimize the above energy function at each temperature setting. Step sizes necessary for gradient descent methods are eschewed. We first minimize the energy function w.r.t. $c_{t\theta,ij}$ (while satisfying the complete/incomplete data constraint using a simple normalization) followed by succesive minimization w.r.t. $\mathbf{f}$ and $\mathbf{z}$. After suitable convergence criteria are met, β is increased and the procedure repeated.

The update equations for $\mathbf{c}$, $\mathbf{f}$ and $\mathbf{z}$ in the WP case are:

$$\frac{\partial E}{\partial \mathbf{c}} = 0 \Rightarrow c_{t\theta,ij} = g_{t\theta} \frac{\mathcal{H}_{t\theta,ij} f_{i,j}}{\sum_{kl} \mathcal{H}_{t\theta,kl} f_{k,l}} \tag{7}$$

$$\frac{\partial E}{\partial f_{i,j}} = 0 \Rightarrow f_{i,j} = \frac{-(\sum_{t\theta} \mathcal{H}_{t\theta;ij} - 2\lambda X_3) + \sqrt{(\sum_{t\theta} \mathcal{H}_{t\theta;ij} - 2\lambda X_3)^2 + 8\lambda X_2 X_1}}{4\lambda X_2}$$

$$\frac{\partial E}{\partial z_{i,j}} = 0 \Rightarrow z_{i,j} \stackrel{\text{def}}{=} \frac{1}{1 + \exp\{-\beta(\lambda\,[f_{vv}^2(i,j) + 2f_{hv}^2(i,j) + f_{hh}^2(i,j)] - \alpha)\}} \tag{8}$$

where $X_1 \stackrel{\text{def}}{=} \sum_{t\theta} c_{t\theta,ij}$, and X_2 and X_3 are functions of the $\mathbf{f}$ and $\mathbf{z}$ that are in the neighborhood of $f_{i,j}$. More details on this derivation (and for the WM case as well) can be found in [16,6]. Below we summarize the DA algorithm:

Pseudo-code for the MAP-DA algorithm
Initialize β to β_0, $\mathbf{f}$ and $\mathbf{z}$ to random values.
Begin A: Do A until $z_{i,j} \leq \tau$ or $z_{i,j} \geq (1 - \tau)$.
  Begin B: Do B until $\frac{E_n - E_{n-1}}{E_n - E_0} < \epsilon$.
  Update complete data $\mathbf{c}$ using (7) (E-step).
  Update $\mathbf{f}$ and $\mathbf{z}$ using (8) (M-step).
  End B
$\beta \leftarrow (\beta \times \gamma)$ where γ governs the annealing schedule.
End A

3. Experiments and Results

We performed 2-D simulation studies with projection data from a 128×128 "ground truth" autoradiograph phantom [7], with 128 projection angles over 360° and 192 detector bins at each projection. To generate projection data, we simply add Poisson noise to each of the attenuated projections of the phantom, using the noiseless projection value as the mean in a Poisson random number generator.

We tested the four reconstruction algorithms—MAP-DA with the WM prior, MAP-DA with the WP prior, and ML-EM with two different stopping rules on the phantom shown in Fig. 1. For convenience, we will refer to these simply as the WM, WP, EM-1, and EM-2 algorithms.

For the ML-EM reconstructions, we used two different stopping rules. Our EM-1 algorithm adhered to a stopping rule based on minimum RMS error in the reconstruction. With our simulations, the average number of iterations for the minimum RMS error averaged 16. For our noise trials, we fixed the number of iterations at these average values. Our EM-2 algorithm adhered to a stopping rule related to the χ^2 stopping rule reported in [17].

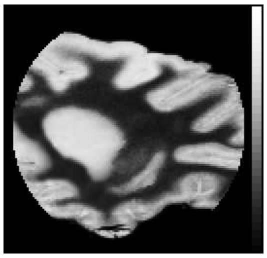

Figure 1. "Ground-truth" autoradiograph phantom used in experiments

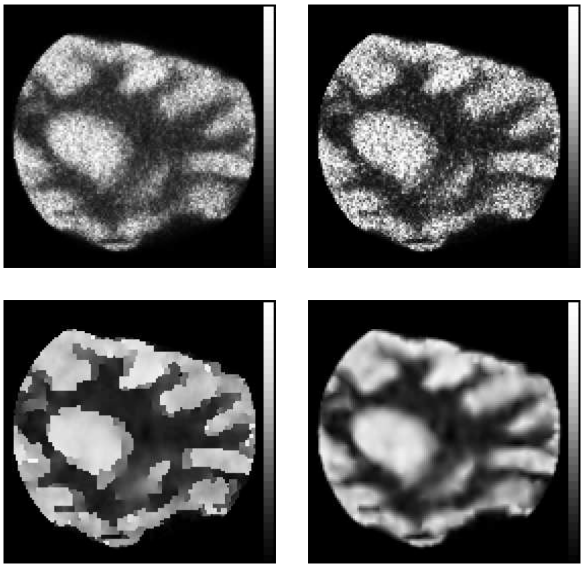

Figure 2. Anecdotal reconstructions. Top left: EM-1 reconstruction. Top right: EM-2 reconstruction. Bottom left: WM reconstruction. Bottom right: WP reconstruction.

Figure 2 shows anecdotal reconstructed images from the four reconstruction algorithms. Several qualitative observations may be noted in Fig. 2. The WM reconstruction, Figs. 2(c), looks artificially patchy, a result not unexpected since WM tends to favor piecewise constant reconstructions. The gradient limit effects ("patchiness") may be reduced by increasing λ, but then a tradeoff results and the reconstructed images become excessively smooth. The WM attempts to create step edges even in the ramp regions, resulting in spurious discontinuities. By extension to a higher order model, the WP reconstructions avoid the artifacts of WM reconstructions as seen in Figs. 2(d). Comparisons of WP to the EM reconstructions in Fig. 2 show that WP reduces noise without introducing artifacts. Close inspection reveals that the WP reconstruction captures subtle aspects of the phantom in Fig. 1 that are missed by the EM algorithms. More quantitative analyses using bias/variance metrics can be found in [16].

4. Conclusions

We have introduced the weak plate prior—a higher order mechanical model—to emission tomography. The new prior is capable of representing inhomogeneous source distributions frequently assumed to be uniform. A novel Generalized EM deterministic annealing algorithm is used to obtain MAP estimates. Our initial efforts [18] in the difficult problem of estimating the hyperparameters λ and α, utilize maximum pseudo-likelihood methods with training data, and recently other efforts have been reported for determining λ [19].

References

1. L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, **1**, pp. 113–122, 1982.
2. S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, **4**, pp. 1–58, 1992.
3. S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bulletin of the International Statistical Institute*, **LII-4**, pp. 5–21, 1987.
4. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, **8**, pp. 194–202, June 1989.
5. G. Gindi, M. Lee, A. Rangarajan, and G. Zubal, "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Trans. Med. Imag.*, **12**, pp. 670–680, Dec. 1993.
6. G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and G. Zubal, "Bayesian reconstruction for emission tomography via deterministic annealing," in *Info. Proc. Med. Imag.*, H. Barrett and A. Gmitro, eds., pp. 322–338, Springer–Verlag, 1993.
7. G. Gindi and A. Rangarajan, "What can SPECT learn from autoradiography?," in *Proc. Nucl. Sci. Symp./Med. Imag. Conf.*, vol. 4, pp. 1715–1719, IEEE Press, 1994.
8. D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Patt. Anal. Mach. Intell.*, **14**, pp. 367–383, March 1992.
9. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, **6**, pp. 721–741, Nov. 1984.
10. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
11. A. Rangarajan and R. Chellappa, "Markov random field models in image processing," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, ed., pp. 564–567, MIT Press, 1995.
12. I. Elfadel, "Convex potentials and their conjugates in analog mean-field optimization," *Neural Computation*, **7**, pp. 1079–1104, Sept. 1995.
13. A. Rangarajan and R. Chellappa, "A continuation method for image estimation using the adiabatic approximation," in *Markov Random Fields: Theory and Application*, R. Chellappa and A. K. Jain, eds., pp. 69–91, Academic Press, 1993.
14. D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRFs: Surface reconstruction," *IEEE Trans. Patt. Anal. Mach. Intell.*, **13**, pp. 401–412, May 1991.
15. R. Hathaway, "Another interpretation of the EM algorithm for mixture distributions," *Stat. and Prob. Lett.*, **4**, pp. 53–56, 1986.
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.*, **14**, pp. 669–680, Dec. 1995.
17. J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, **8**, pp. 186–193, 1989.
18. S.-J. Lee, G. Gindi, G. Zubal, and A. Rangarajan, "Using ground-truth data to design priors in Bayesian SPECT reconstruction," in *Info. Proc. Med. Imag.*, Y. Bizais, C. Barillot, and R. D. Paola, eds., pp. 27–38, Kluwer Acad. Pub., 1995.
19. Z. Zhou, R. M. Leahy, and E. U. Mumcuoglu, "Maximum likelihood hyperparameter estimation for Gibbs priors with applications to PET," in *Info. Proc. Med. Imag.*, Y. Bizais, C. Barillot, and R. D. Paola, eds., pp. 39–51, Kluwer Acad. Pub., 1995.