BAYESIAN IMAGE RESTORATION AND SEGMENTATION BY CONSTRAINED OPTIMIZATION

S.Z. Li *K.L. Chan* *H. Wang*

School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798

Abstract

*A constrained optimization method, called the Lagrange-Hopfield (LH) method, is presented for solving Markov random field (MRF) based Bayesian image estimation problems for restoration and segmentation. The method combines the augmented Lagrangian multiplier technique with the Hopfield network to solve a constrained optimization problem into which the original Bayesian estimation problem is reformulated. The LH method effectively overcomes instabilities that are inherent in the penalty method (*e.g. *Hopfield network) or the Lagrange multiplier method in constrained optimization. An additional advantage of the LH method is its suitability for neural-like analog implementation. Experimental results are presented which show that LH yields good quality solutions at reasonable computational costs.*

1. INTRODUCTION

Image restoration is to recover a degraded image and segmentation is to partition an image into regions of similar image properties. Both can be posed generally as image estimation when the underlying image or segmentation map is to be recovered from a degraded image. Due to various uncertainties, it is often posed as an optimization problem. A criterion often used in optimal image estimation is maximum *a posteriori* (MAP) probability. In MAP estimation, it is assumed that both the prior probability distribution of the true image class and the conditional probability distribution of the data are known. In many MAP formulations, a Markov random field (MRF) or equivalently a Gibbs distribution is used as the prior distribution which encodes contextual constraints of the image [12, 21, 4, 9]. The posterior probability is generally also a Gibbs distribution. Minimizing the posterior probability is equivalent to finding the configuration which minimizes the energy function in the Gibbs distribution. This is the MAP-MRF framework [12, 5, 18].

When pixel values are from a discrete set, such as in multi-level image restoration and segmentation, the minimization is combinatorial. Many methods have been proposed to-date for finding the energy minimum in a discrete space. The most well-known ones include the iterative conditional modes (ICM) [4] and simulated annealing (SA) [17, 12]. Other algorithms of similar nature include the highest confidence first (HCF) [6, 7].

In this paper, we present a novel method, called the Lagrange-Hopfield (LH) method, for MAP-MRF multi-level image restoration and segmentation in which an image can be textured. The original MAP problem is converted into a constrained optimization problem. The LH method then solves the constrained optimization. It combines the augmented Lagrangian multiplier technique [24, 14, 23] with the Hopfield network [15, 16]. This effectively overcomes instabilities inherent in the penalty method (*e.g.* Hopfield network) or the Lagrange multiplier method in constrained optimization. The resulting algorithm solves a system of unconstrained differential equations, and is suitable for neural-like analog implementation. Experiments in both image restoration and segmentation are shown to compare the LH method with the iterative conditional modes (ICM), highest confidence first (HCF) and simulated annealing (SA) algorithms. The results show that LH performs better than ICM, HCF and sometimes SA in terms of the minimized energy values and error rates. Yet it is much more efficient than SA; it quickly yields a good solution after dozens of iterations.

Although in this paper the LH is used for image restoration and segmentation only, the method is general enough also for solving other problems that can be formulated as combinatorial optimization. Successful applications have been obtained for graph matching [19] and traveling salesman problem [20].

The rest of the paper is organized as follows: Section 2 introduces the basic MAP formulation with discrete valued MRF. Section 3 reformulates the MAP estimation, which is an un-constrained combinatorial optimization problem, as a constrained real optimization, and then describes the LH method for solving the latter optimization. Section 4 presents experimental comparisons. Conclusions are given in Section 5.

2. MAP-MRF FORMULATION

The underlying image signal is denoted as $f = \{f_i \mid i \in \mathcal{S}\}$ where $\mathcal{S} = \{1, \ldots, m\}$ is the set of sites where each $i \in \mathcal{S}$ indexes a pixel in the image lattice. Each pixel takes on a discrete value f_i in the label set $\mathcal{L} = \{1, \ldots, M\}$ and the f is the restored multi-level image or the segmented map. The spatial relationship of the sites, each of which is indexed by a single number in $\mathcal{S}$, is determined by a neighborhood system $\mathcal{N} = \{\mathcal{N}_i \mid i \in \mathcal{S}\}$ where $\mathcal{N}_i$ is the set of sites neighboring i. A single site or a set of neighboring sites form a clique denoted by c. In this paper, only up to pair-site cliques are considered.

The type of the underlying image f can be blob-like regions or a texture pattern. Different types are due to different ways that pixels influence each other, *i.e.* different contextual interactions. Such contextual interaction can be modeled as MRFs. According to the Markov-Gibbs equivalence [3, 12], an MRF is a Gibbs distribution

$$P(f) = Z^{-1} \times \mathrm{e}^{-\frac{1}{T}\sum_{c \in \mathcal{C}} V_c(f)} \tag{1}$$

where $V_c(f)$ is the clique potential function, $\mathcal{C}$ is the set of all cliques, T is a temperature constant, and Z is a normalizing constant.

Among various MRFs, the multi-level logistic (MLL) model [10, 8, 9] is a simple and yet powerful mechanism for encoding a large class of spatial patterns. In MLL, the pair-site clique potentials take the form

$$V_2(f_i, f_{i'}) = \begin{cases} \beta_c & \text{if } f_i = f_{i'} \\ -\beta_c & \text{otherwise} \end{cases} \tag{2}$$

where β_c is the parameter for cliques of type $c = \{i, i'\}$, while the single site potentials is defined by

$$V_1(f_i) = \alpha_I \quad \text{if } f_i = I \in \mathcal{L}_d \tag{3}$$

where α_I is the potential for label value I. For the 8-neighborhood system, we use $\beta_1, \beta_2, \beta_3$ and β_4 for pairwise cliques in the $0°$, $90°$, $45°$ and $135°$ directions, respectively.

When the true pixel values are contaminated by identical independently distributed (i.i.d.) Gaussian noise, the observed data is

$$d_i = f_i + e_i \tag{4}$$

where $e_i \sim N(0, \sigma^2)$ is the zero mean Gaussian distribution with standard deviation σ. The conditional distribution of d is

$$p(d \mid f) = \frac{1}{\prod_i^m \sqrt{2\pi\sigma^2}} \mathrm{e}^{-U(d \mid f)} \tag{5}$$

where

$$U(d \mid f) = \sum_{i \in \mathcal{S}} \frac{(f_i - d_i)^2}{2\sigma^2} \tag{6}$$

The energy in the posterior distribution, $P(f \mid d) \propto \mathrm{e}^{-E(f)}$, is

$$E(f) = \sum_{i \in \mathcal{S}} \frac{(f_i - d_i)^2}{2\sigma^2} + \sum_{\{i\} \in \mathcal{C}} V_1(f_i) + \sum_{\{i,i'\} \in \mathcal{C}} V_2(f_i, f_{i'}) \tag{7}$$

in which the MRF and noise parameters α, β and σ are assumed known for the MAP-MRF estimation. The MAP estimate for restoration and segmentation is defined as

$$f^* = \arg\min_{f \in \mathcal{L}^m} E(f) \tag{8}$$

The interested reader is referred to [18] for a thorough representation on the MAP-MRF framework.

3. THE LH METHOD

The minimization of the energy function $E(f)$ is in the discrete space $\mathcal{L}^m$ and hence combinatorial. Existing methods for the above minimization problem include SA, ICM and HCF. Here we present the LH method.

3.1. MAP-MRF Estimation as Constrained Optimization

We convert the original combinatorial problem formulation, in which the optimization is performed in a discrete space $\mathcal{L}^m$, into a constrained optimization in a real space. A convenient way is to consider it as a continuous label assignment problem [25]. We use an M element vector $p_i = [p_i(I) \mid I \in \mathcal{L}]$ to represent the state of the assignment for $i \in \mathcal{S}$. The real value $p_i(I) \in [0, 1]$ reflects the strength with which i is assigned label I. The matrix $p = [p_i(I) \mid i \in \mathcal{S}, I \in \mathcal{L}]$ is the state of the assignment.

The energy with the p variables is defined by

$$\begin{aligned} E(p) \quad = \quad & \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} r_i(I)\ p_i(I) + \\ & \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} \sum_{i' \in \mathcal{S}, i' \neq i} \sum_{I' \in \mathcal{L}} r_{i,i'}(I, I')\ p_i(I)\ p_{i'}(I') \end{aligned} \tag{9}$$

where $I = f_i$ and $I' = f_{i'}$,

$$r_i(I) = V_1(I \mid d) = (I - d_i)^2/2\sigma + V_1(I) \tag{10}$$

is the single-site clique potential function in the posterior distribution $P(f \mid d)$, and

$$r_{i,i'}(I, I') = V_2(I, I' \mid d) = V_2(I, I') \tag{11}$$

is the pair-site clique potential function in $P(f \mid d)$.

With such a representation, a combinatorial problem is reformulated as the following constrained minimization

$$\min_p \qquad E(p) \tag{12}$$

$$\text{subject to} \qquad C_i(p) = 0 \quad i \in \mathcal{S} \tag{13}$$

$$p_i(I) \geq 0 \quad \forall i \in \mathcal{S}, \forall I \in \mathcal{L} \tag{14}$$

where C_i's are some real functions of p variables. For image restoration and segmentation, and more generally labeling problems, we have

$$C_i(p) = \sum_I p_i(I) - 1 \tag{15}$$

The final solution p^* is subject to additional constraints

$$p_i^*(I) = 0 \quad \text{or} \quad 1 \qquad \forall i, I \tag{16}$$

3.2. Use of Augmented Lagrange and Hopfield Techniques

The LH method for solving the above constrained minimization problem combines the augmented Lagrange technique with the Hopfield network. First, we introduce internal variables $u_i(I) \in (-\infty, +\infty)$ $(\forall i, I)$ and relate them to $p_i(I)$ via

$$p_i(I) = \psi_T(u_i(I)) \tag{17}$$

where $\psi_T(x)$ is a sigmoid function

$$\psi_T(x) = 1/[1 + e^{-x/T}] \tag{18}$$

controlled by a temperature parameter $T > 0$. With this introduction of the internal u variables, the energy function can be considered as $E(u) = E(p(u))$. This treatment confines $p_i(I)$ to the range $(0, 1)$ so as to impose the inequality constraints of (14). When $T \to 0^+$, $p_i(I)$ is forced to be 0 or 1 depending on whether $u_i(I)$ is positive or negative. Thereby, the unambiguity constraints of (16) are imposed.

Next, we use the Lagrange method to impose the equality constraints of (13). Define the following Lagrange function

$$L(p, \gamma) = E(p) + \sum_i \gamma_i C_i(p) \tag{19}$$

where γ_i are called the Lagrange multipliers. It is a function of the $M \times m$ variables of p and the K variables of γ. For p^* to be a local minimum subject to the constraints, it is necessary that (p^*, γ^*) be a stationary point of the Lagrange function:

$$\begin{array}{l} \nabla_p L(p^*, \gamma^*) = 0 \\ \nabla_\gamma L(p^*, \gamma^*) = 0 \end{array} \tag{20}$$

If (p^*, γ^*) is a saddle point for which

$$L(p^*, \gamma) \leq L(p^*, \gamma^*) \leq L(p, \gamma^*) \tag{21}$$

then p^* is a local minimum of $E(p)$ satisfying $C_i(p^*) = 0$, *i.e.* a local solution to the the constrained optimization problem [13].

The following dynamic equations, called the basic differential multiplier method [23], can be used to find such a saddle point

$$\begin{array}{rcl} \dfrac{\mathrm{d}p_i(I)}{\mathrm{d}t} & = & -\dfrac{\partial L(p, \gamma)}{\partial p_i(I)} \\ \dfrac{\mathrm{d}\gamma_i}{\mathrm{d}t} & = & +\dfrac{\partial L(p, \gamma)}{\partial \gamma_i} \end{array} \tag{22}$$

It performs energy descent on p but ascent on γ. The convergence of this system is illustrated in [1].

The Lagrangian (19) can be augmented by adding penalty terms, such as quadratic terms $[C_i(p)]^2$, giving an augmented Lagrange function [24, 14]

$$L_\beta(p, \gamma) = E(p) + \sum_i \gamma_i C_i(p) + \frac{\beta}{2} \sum_i [C_i(p)]^2 \tag{23}$$

where $\beta > 0$ is a finite weighting factor. The introduction of the quadratic terms $\beta \sum_i [C_i(p)]^2$ with $C_i(p) = 0$ does not alter the location of the saddle point. The quadratic terms effectively overcomes the problems associated with the penalty method or the Lagrange method when used alone and stabilize the system [2, 11].

The dynamic equations for minimizing L_β are

$$\frac{\mathrm{d}p_i(I)}{\mathrm{d}t} = -\frac{\partial L_\beta(p, \gamma)}{\partial p_i(I)} \tag{24}$$

$$\frac{\mathrm{d}\gamma_i}{\mathrm{d}t} \quad = \quad +\frac{\partial L(p, \gamma)}{\partial \gamma_i} = +C_i(p) \tag{25}$$

where

$$\frac{\partial L_\beta(p, \gamma)}{\partial p_i(I)} = \frac{\partial E(p)}{\partial p_i(I)} + \sum_i \gamma_i \frac{\partial C_i(p)}{\partial p_i(I)} + \beta \sum_i C_i(p) \frac{\partial C_i(p)}{\partial p_i(I)} \tag{26}$$

In the above, $\frac{\partial C_i(p)}{\partial p_i(I)}$ equals 1 for $C_i(p)$ given in (15) and hence can be dropped from the equations. This corresponds to the modified differential multiplier method [23]. Our experiments show that the penalty terms are necessary for damping oscillations and hence helping the convergence of numerical computation.

In the LH method, the updating of label assignment is performed on u, rather than on p. Equation (24) is replaced by

$$\begin{aligned} \frac{\mathrm{d}u_i(I)}{\mathrm{d}t} &= -\frac{\partial L_\beta(p(u),\gamma)}{\partial p_i(I)} \qquad (27) \\ &= = -\left[q_i(I) + \sum_i \gamma_i + \beta \sum_i C_i(p)\right] \end{aligned}$$

where $q_i(I) = \frac{\partial E}{\partial p_i(I)}$. Because $\frac{\partial p_i(I)}{\partial u_i(I)} = \frac{\mathrm{e}^{-u_i(I)/T}}{T(1+\mathrm{e}^{-u_i(I)/T})^2}$ is always positive, $\frac{\mathrm{d}p_i(I)}{\mathrm{d}t}$ has the same sign as $\frac{\mathrm{d}u_i(I)}{\mathrm{d}t}$.

In numerical computation, the LH algorithm consists of the following three equations corresponding respectively to (27), (17) and (25)

$$u_i^{(t+1)}(I) \leftarrow u_i^{(t)}(I) - \mu\Delta_i(I) \qquad (28)$$

$$p_i^{(t+1)}(I) \leftarrow \psi_T(u_i^{(t+1)}(I)) \qquad (29)$$

and

$$\gamma_i^{(t+1)} \leftarrow \gamma_i^{(t)} + \beta C_i(p^{(t)}) \qquad (30)$$

In the above, $\Delta_i(I) = \left\{q_i^{(t)}(I) + \sum_i \gamma_i^{(t)} + \beta \sum_i C_i(p^{(t)})\right\}$ corresponds to $\frac{\partial L_\beta(p(u),\gamma)}{\partial p_i(I)}$, and μ is a step size factor. During the update, T may be decreased and β increased to speed up convergence. The updating is performed synchronously in parallel for all i and I. Comparing the LH algorithm with the mean field theoretical algorithms of Peterson and Soderberg [22], we see that the LH does not need the normalization operation required by the latter algorithm and thus is more convenient for analog implementation.

Using the augmented Lagrange multipliers, the LH method overcomes the instability with the penalty method and the standard Lagrangian, and hence improves convergence. Basically, one is converting a combinatorial optimization to a constrained real optimization, and solving the constrained problem using an unconstrained real optimization method. In the penalty method, the objective for the unconstrained optimization is a weighted sum of the original objective function and other terms which penalize the violation of the constraints. To produce valid solutions, the relative weight for the original objective function must be kept relatively small in order to verify the constraints; therefore the solutions using the penalty method can be significantly biased from the true minimum of the original objective function, owing to large influence from the penalty terms of constraints.

With the augmented Lagrangian, the weighting values for the penalty terms can be much smaller than those required by the penalty method; thus the relative weight for the original objective can be increased, which helps yielding a lower objective value while keeping the convergence. At the same time, the augmented Lagrange method also overcomes the zigzagging problem with the standard Lagrange method and improve the convergence. These advantages of the augmented Lagrangian plus the use of Hopfield method for imposing the inequality and unambiguity constraints have made the success of the LH method.

4. EXPERIMENTAL RESULTS

In the following, we present two experiments, one on MAP-MRF image restoration and the other on segmentation, to compare the performance of the following algorithms (1) LH, (2) ICM (scan-line visit order), (3) pICM (parallel ICM using codings), (4) HCF (the parallel version), and (5) SA (with the Metropolis sampler). The solution quality is measured by (1) the minimized energy value $E(f^*)$ according to (7), and (2) the error rate defined as the number of different pixel values between the restored image f^* and the true image divided by the total number of pixels.

For LH, the parameters for the iterative equations (28) – (30) are set as follows: $\beta = 1000$, $\mu = 100$ and $T = 100000$, λ are increased from 1 to 100 according to $\lambda \leftarrow 1.01\lambda$. The convergence criterion is when $\|u^{(t)} - u^{(t-1)}\|_\infty$ is smaller than 0.0001 where $\|\cdot\|$ is the infinity norm. The HCF algorithm is the parallel version described in [7] rather than the sequential one in [6] because the former almost always performs better than the latter as reported in [7]. SA is included as a benchmark. The schedule for SA is set as $T^{(t+1)} \leftarrow 0.999T^{(t)}$ with $T^{(t)} = 10^4$.

The first experiment are on MAP-MRF restoration performed on a synthetic image shown in Fig.1. In the figure, the original image has $M = 4$ gray levels, the label set is $\mathcal{L} = \{1,2,3,4\}$ with the corresponding gray pixel values also in $\{1,2,3,4\}$. The image was generated according to a Gibbs distribution determined by the clique potential parameters α_I ($I \in \mathcal{L}$), β_1, $\cdots$, β_4 where the parameter values used for the image are $\sigma = 1$, $\alpha_I = 0$, $\beta_1 = \beta_2 = \beta_3 = -1$, $\beta_4 = 1$. The Metropolis algorithm combined with an annealing procedure was used to generate the Gibbs distribution. In the observed noisy image, every pixel takes a real value which is the true pixel value plus zero-mean i.i.d.

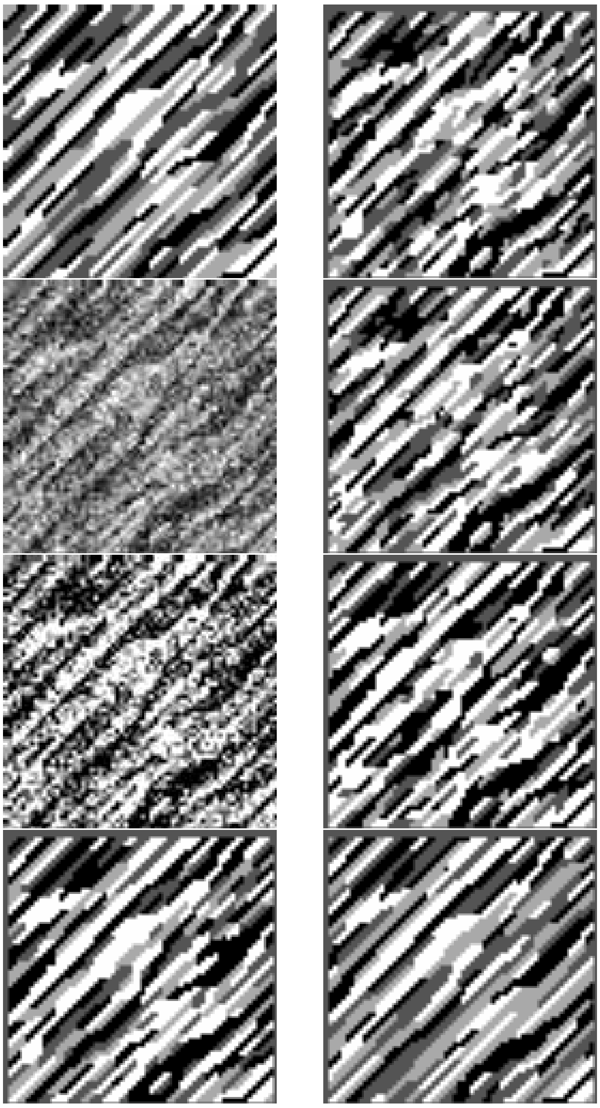

Figure 1: Restoration of synthetic image. Left column top down: • The original image. • The observed noisy image. • The maximum likelihood estimate. • The LH solution. Right column top down: • The ICM solution. • The pICM solution. • The HCF solution. • The SA solution.

Gaussian noise with a standard deviation $\sigma = 1$. The maximum likelihood estimate was used as the initial labeling for the five algorithms being compared.

Table 1 shows the minimized energy values, the error rates, and the required iteration numbers. The smaller the minimized energy value and the error rate, the better the solution quality. The smaller the iteration number, the faster the convergence. It can be seen from the table that objectively, LH performs the best out of the three deterministic algorithms in terms of both the minimized energy, *i.e.* maximized posterior probability, values and the error rates. Overall, the solution quality is ordered as "ICM < HCF < LH < SA" and mostly "pICM < ICM". A subjective evaluation based on the images in the figures would also agree to this ordering. According to the iteration numbers, LH takes much less iterations than SA, about 2.1% – 4.4% of that, to converge. When implemented using neural like hardware, the computation of thousands of LH iterations can be done in real-time.

	LH	ICM	pICM	HCF	SA
Min. Energy	-11049	-10003	-10049	-10269	-11988
Error Rate	0.268	0.349	0.354	0.367	0.164
Iterations	3654	7	6	31	83034

Table 1: The minimized energy values, error rates, and iteration numbers.

The second experiment compares the algorithms in performing MAP-MRF segmentation on the Lenna image of size 256×240 into a tri-level segmentation map. It is performed by using exactly the same implementation of all the five algorithms as used for the first experiment. The results are illustrated in Figure 2. The noisy input image is the original Lenna image corrupted by i.i.d. zero-mean Gaussian noise of standard deviation 10. The conditional density functions are assumed to be i.i.d. Gaussian distributions with the same standard deviation $\sigma = 10$ and three mean values 40, 125 and 200 for the three-level segmentation. An isometric MRF prior is used with the four β parameters being $(-1, -1, -1, -1)$. The maximum likelihood estimate was used as initial segmentation for LH, ICM, HCF and SA. Table 2 illustrates the minimized energy, *i.e.* maximized posterior probability, values and the iteration numbers numerically. The solution quality is ordered as "HCF < ICM < pICM < SA < LH". A subjective evaluation based on the images in Figure 2 would also agree to this rating.

The LH method may not be as slow as it might seem. LH converged in just over 1000 iterations, and this is with the stringent convergence criterion $\|u^{(t)} - u^{(t-1)}\|_\infty < 0.0001$. LH finds a solution better than ICM after only a dozen of iterations and the LH solution is better than the HCF solution in every iteration stage.

Finally, we have the following remarks regarding the use of the augmented Lagrange multiplier method: Without the γ_i terms, LH degenerates to a penalty method combined with the Hopfield method; without the C_i terms, it degenerates to a standard Lagrange method combined with the Hopfield method. The algorithm diverges for both degenerated cases. It con-

Figure 2: Segmentation of the Lenna image. Left column top down: • The noisy input image. • The maximum likelihood segmentation. • The LH solution. Right column top down: • The ICM solution (pICM yields a similar solution). • The HCF solution. • The SA solution.

verges only when both types of the terms are present to be a full LH algorithm.

	LH	ICM	pICM	HCF	SA
Min. Energy	-180333	-171806	-171946	-167167	-173301
Iterations	1255	7	7	38	593916

Table 2: Numerical comparison of the algorithms on the segmentation of the Lenna image.

5. CONCLUSIONS

The Lagrange-Hopfield (LH) method has been presented for MAP-MRF image restoration and segmentation. The LH method converts the original combinatorial optimization problem into a constrained real optimization, and solve the latter using the augmented Lagrange method combined with the Hopfield method. It is more stable and convergent than the standard Lagrangian and Hopfield methods. We have shown that the LH method gives better quality solutions than ICM and HCF, and comparable to SA in terms of both minimized energy values and error rates; yet it takes much less computational effort than SA to converge.

6. REFERENCES

[1] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in Linear and Nonlinear Programming.* Stanford University Press, 1958.

[2] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York, 1982.

[3] J. Besag. "Spatial interaction and the statistical analysis of lattice systems" (with discussions). *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.

[4] J. Besag. "On the statistical analysis of dirty pictures" (with discussions). *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.

[5] R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Applications.* Academic Press, 1993.

[6] P. B. Chou and C. M. Brown. "The theory and practice of Bayesian image labeling". *International Journal of Computer Vision*, 4:185–210, 1990.

[7] P. B. Chou, P. R. Cooper, M. J. Swain, C. M. Brown, and L. E. Wixson. "Probabilistic network inference for cooperative high and low level vision". In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*, pages 211–243, Boston, 1993. Academic Press.

[8] H. Derin and W. S. Cole. "Segmentation of textured images using using Gibbs random fields". *Computer Vision, Graphics and Image Processing*, 35:72–98, 1986.

[9] H. Derin and H. Elliott. "Modeling and segmentation of noisy and textured images using Gibbs random fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):39–55, January 1987.

[10] H. Elliott, H. Derin, R. Cristi, and D. Geman. "Application of the Giibs distribution to image segmentation". In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, pages 32.5.1–32.5.4, San Diego, March 1984.

[11] R. Fletcher. *Practical Methods of Optimization.* Wiley, 1987.

[12] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[13] B. S. Gottfried. *Introduction to Optimization Theory.* Prentice-Hall, 1973.

[14] M. R. Hestenes. "Multipler and gradient methods". *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[15] J. J. Hopfield. "Neurons with graded response have collective computational properties like those of two state neurons". *Proceedings of National Academic Science, USA*, 81:3088–3092, 1984.

[16] J. J. Hopfield and D. W. Tank. " 'Neural' computation of decisions optimization problems". *Biological Cybernetics*, 52:141–152, 1985.

[17] S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi. "Optimization by simulated annealing". *Science*, 220:671–680, 1983.

[18] S. Z. Li. *Markov Random Field Modeling in Computer Vision.* Springer-Verlag, 1995.

[19] S. Z. Li. "Relaxation labeling using Lagrange multipliers and Hopfield network". In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 266–269, Washington, D.C., 23-26 October 1995.

[20] S. Z. Li. "Improving convergence and solution quality of hopfield type network with augmented lagrange technique. *Submitted*, 1996.

[21] J. L. Marroquin. "Probabilistic solution of inverse problems". *A. I. Lab. Tech. Report* No. 860, MIT, Cambridge, MA, 1985.

[22] C. Peterson and B. Soderberg. "A new method for mapping optimization problems onto neural networks". *International Journal of Neural Systems*, 1(1):3–22, 1989.

[23] J. C. Platt and A. H. Barr. "Constrained differential optimization". In *Proceedings of the IEEE 1987 NIPS conference*, 1988.

[24] M. J. D. Powell. "A method of nonlinear constraints in minimization problems". In R. Fletcher, editor, *Optimization*, London, 1969. Academic Press.

[25] A. Rosenfeld, R. Hummel, and S. Zucker. "Scene labeling by relaxation operations". *IEEE Transactions on Systems, Man and Cybernetics*, 6:420–433, June 1976.