JOINT PARAMETER ESTIMATION AND RESTORATION USING MRF MODELS AND HOMOTOPY CONTINUATION METHOD

P. K. Nanda,[1] U. B. Desai and P. G. Poonacha

Department of Electrical Engineering
Indian Institute of Technology- Bombay
Bombay 400 076
India
e-mail:ubdesai@ee.iitb.ernet.in

ABSTRACT

This paper presents a joint strategy for parameter estimation of Markov Random Field (MRF) model and image restoration. The proposed scheme is an unsupervised one in the sense that no a priori knowledge of the actual image is assumed. The technique of homotopy continuation method is employed to estimate the model parameters. The model considered involves line fields and is tested on real images. Simulation results are presented for gray level images.

1. INTRODUCTION

During the past decade Markov Random Field (MRF) models have been used extensively in image restoration, edge detection, segmentation and other early vision problems. The literature on this is very vast nevertheless, we cite a few papers ([1] - [12]) to illustrate the breadth of applicability of MRF models. An MRF prior model is specified in terms of certain parameters called the clique parameters. In most of the applications, the performance of the algorithms using MRF models depends on the choice of the MRF model parameters. Often, these parameters are selected on an adhoc basis. Therefore, to enhance the performance, several parameter estimation strategies have been proposed that can be broadly categorized as: (i) the generalized cross validation scheme of Wahba [6], and (ii) minimization of the likelihood function or pseudolikelihood function of Besag [7]. However, most of the approaches assume the availability of a good initial image which is often not the case in practice. Lakshmanan and Derin [8] and Younes [9] have considered

the problem of simultaneous parameter estimation and segmentation where they have assumed to have a priori knowledge of the noise variance. Kang and Katsaggelos [13] have also suggested an iterative scheme for parameter estimation and restoration with a priori knowledge of the noise variance. Recently, some unsupervised schemes are suggested in ([10]-[12]) where noise variance is also estimated along with the clique parameters.

In this paper we focus on the solution to the joint problem of MRF model parameter estimation and image restoration. Moreover, we also let the noise variance to be unknown; and hence the proposed scheme is an unsupervised one. In this scheme we alternate between parameter estimation and image restoration. The parameter estimation steps employ some of the recent advances in continuation method ([14]-[17]) together with the conditional pseudolikelihood function to estimate the clique parameters as well as noise variance. Image restoration is achieved using simulated annealing algorithm.

2. IMAGE MODEL

In this paper we consider the following model. $Y_{ij} = X_{ij} + W_{ij}, \forall (i,j) \in (N \times N)$. The above model with a lexicographical ordering can be written as $Y = X + W$, where, $Y = [Y_{ij}]$ = observed image random field, $X = [X_{ij}]$ = unknown image random field, $W = [W_{i,j}]$ = noise random field and $(N \times N)$ is the rectangular lattice over which random fields are defined. In the above model X is modeled as a Markov random field with respect to a neighbourhood system η and is described in terms of the local characteristics.

$$P(X_{i,j} = x_{i,j} \mid X_{k,l} = x_{k,l}, (k,l) \in (N \times N), (k,l) \neq (i,j))$$

$$= P(X_{i,j} = x_{i,j} \mid X_{k,l} = x_{k,l}, (k,l) \in \eta)$$

Here we consider the first order neighbourhood structure, that is the neighbourhood consisting of the closest four neighbours of each pixel. Besides Markovian property we make the following assumptions. (a) $W_{i,j}$ is a white Gaussian sequence with zero mean and variance σ^2 (σ unknown). (b) $N_{i,j}$ is statistically independent of $X_{k,l}$, for all (i,j) and (k,l) belonging to $N \times N$. (c) $x_{i,j}$ takes any gray level value from the set $G = (1, \ldots, N_G)$, (typically $N_G = 256$).

It is known [1] that X is a MRF with respect to the neighbourhood system η if and only if $P(X = x)$ is Gibbs distributed with respect to η. This is expressed as

$$P(X = x \mid \phi) = \frac{1}{Z}e^{-U(x,\phi)}$$

where $Z = \sum_x e^{-U(x,\phi)}$ is the partition function, ϕ represents the clique parameter vector, the exponent term $U(x,\phi)$ is called the energy function and is of the form $U(x,\phi) = \sum_{C:(i,j)\in c} V_c(x,\phi)$, with $V_c(x,\phi)$ being referred to as the potential. In general, the unknown parameter $\theta = [\phi^T, \sigma^2]^T$.

3. JOINT PARAMETER ESTIMATION AND RESTORATION PROBLEM

As suggested by Lakshmanan and Derin [8] a general approach for joint parameter estimation and say the restoration problem would be to solve the following problem :

$$(x^*, \theta^*) = arg \quad \max_{x,\theta} P(X = x, \theta \mid Y = y) \qquad (1)$$

To find the solution to the above problem is a formidable task and to our knowledge no algorithm is available. Hence, we reformulate the problem which of course would give a suboptimal solution to (1). In the proposed scheme we alternate between parameter estimation and image restoration steps. Let at iteration k $\theta^k = [\phi^k, (\sigma^2)^k]^T$ be the estimate of the parameters, and x^k be the estimate of the image X. Now consider the following problems :

$$x^{k+1} = arg \quad \max_x P(X = x \mid Y = y, \theta^k) \qquad (2)$$

and

$$\theta^{k+1} = arg \quad \max_\theta P(X = x^{k+1} \mid Y = y, \theta) \qquad (3)$$

Problem (2) can be solved as a maximum a posteriori (MAP) estimation problem using a Bayesian approach [1]. It is easily shown that

$$P(X = x \mid Y = y, \theta^k)$$

$$= \frac{P(Y = y \mid X = x, \theta^k)P(X = x \mid \theta^k)}{P(Y = y \mid \theta^k)}$$

Since y is known, the denominator is a constant and with the assumption mentioned earlier, (2) is equivalent to

$$x^{k+1} = arg \quad \max_x \left[e^{\frac{-\|y-x\|^2}{2\sigma^2}} e^{-U(x,\phi^k)} \right] \qquad (4)$$

which can be solved using for example the simulated annealing algorithm.

Considering the parameter estimation problem given in (3) the conditional probability can be expressed as

$$P(X = x^{k+1} \mid Y = y, \theta) = \frac{\frac{1}{(2\pi\sigma^2)^{N^2/2}}e^{\frac{-\|y-x\|^2}{2\sigma^2}}\frac{1}{Z}e^{-U(x,\phi)}}{P(Y = y \mid \theta)}$$

But $P(Y = y \mid \theta)$ is no longer constant. However, It can be shown that

$$P(Y = y \mid \theta) = \sum_\xi \frac{1}{(2\pi\sigma^2)^{N^2/2}}e^{\frac{-\|y-\xi\|^2}{2\sigma^2}}\frac{1}{Z}e^{-U(\xi,\phi)}$$

which implies

$$P(X = x^{k+1} \mid Y = y, \theta) = \frac{e^{\frac{-\|y-x^{k+1}\|^2}{2\sigma^2}}e^{-U(x^{k+1},\phi)}}{\sum_\xi e^{\frac{-\|y-\xi\|^2}{2\sigma^2}}e^{-U(\xi,\phi)}} \qquad (5)$$

In (5) the summation is over all possible realizations of X. Thus, from a computational standpoint, handling (5) would be practically impossible. One can view (5) as a likelihood function to be maximized for estimating θ. To overcome the computational problem, we approximate (5) using the pseudolikelihood function.

$$\prod_{i,j} P(X_{i,j} = x_{i,j}^{k+1} \mid X_{m,n} = x_{m,n}^{k+1}, \ (m,n) \in \eta_{i,j}, Y = y, \theta)$$

$$\underline{\underline{\Delta}}\hat{P}(X = x^{k+1} \mid Y = y, \theta) \approx P(X = x^{k+1} \mid Y = y, \theta) \qquad (6)$$

This is essentially the pseudolikelihood function of Besag [7], except we are approximating the posterior probability distribution instead of the a priori probability distribution. It can be shown that

$$\hat{P}(X = x^{k+1} \mid Y = y, \theta) =$$

$$\prod_{i,j} \left[\frac{e^{\frac{-(\nu_{i,j}-x_{i,j}^{k+1})^2}{2\sigma^2}}e^{-\sum_{C:(i,j)\in c} V_c(x_{i,j}^{k+1},\phi)}}{\sum_{x_{i,j}^{k+1}\in G}\left\{ e^{\frac{-(\nu_{i,j}-x_{i,j}^{k+1})^2}{2\sigma^2}}e^{-\sum_{C:(i,j)\in c} V_c(x_{i,j}^{k+1},\phi)} \right\}} \right] \qquad (7)$$

Due to space limitation, we have omitted the derivation of (7), nevertheless it can be found in [18]. In (7) the summation is over the all possible G gray levels of the pixel $x_{i,j}^{k+1}$, which is much smaller, typically 256. Now the parameter estimation problem is recast as

$$\theta^{k+1} = arg \quad \max_{\theta} \hat{P}(X = x^{k+1} \mid Y = y, \theta) \quad (8)$$

4. HOMOTOPY CONTINUATION METHOD FOR PARAMETER ESTIMATION

It is clear from Section 2.2 that the parameter estimation problem has been reduced to maximization of (7) with respect to θ. Towards this end let

$$f(\theta) = \frac{\partial}{\partial \theta} \left\{ \log \left[\hat{P}(X = x^{k+1} \mid Y = y, \theta) \right] \right\} \quad (9)$$

Now the homotopy method is employed to solve $f(\theta) = 0$. To have an arbitrary starting point for the path, we have considered the fixed point homotopy map given by

$$h(\theta, \lambda, q) = \lambda f(\theta) + (1 - \lambda)(\theta - q) \quad (10)$$

where $0 \leq \lambda \leq 1$ and q is an arbitrary starting point. Here the predictor-corrector method is employed to track the path defined by the homotopy in (10). The procedure can be briefly outlined as follows:
Let $(\theta^k, \lambda^k, \theta^{k-1})$ be a point that satisfies (10) . Therefore, the point thus considered is on the path. Tracking the path involves computing the adjacent point on the path. This is determined in the following way. Increment λ^k by some small value $\Delta\lambda$ thus giving the next point $\lambda^{k+1} = \lambda^k + \Delta\lambda$ and evaluate equation (10) at $(\theta^k, \lambda^{k+1}, \theta^{k-1})$. If the value of the map $h(\theta^k, \lambda^{k+1}, \theta^{k-1})$ is not equal to zero , then the point $(\theta^k, \lambda^{k+1}, \theta^{k-1})$ is not on the path. Since $h(\theta^k, \lambda^{k+1}, \theta^{k-1}) \neq 0$, we try to obtain an estimate of θ^k, say $\hat{\theta}^k$ corresponding to λ^{k+1} such that $h(\theta^k, \lambda^{k+1}, \theta^{k-1}) \approx 0$ To achieve this one could use Newton's algorithm, namely,

$$\hat{\theta}_{i+1}^k = \hat{\theta}_i^k - J_{\hat{\theta}}^{-1}[h(\hat{\theta}_i^k, \lambda^{k+1}, \theta^{k-1})]h(\hat{\theta}_i^k, \lambda^{k+1}, \theta^{k-1}) \quad (11)$$

Where the superscript i denotes the ith Newton iteration and $J_{\hat{\theta}}^{-1}$ is the inverse of the Jacobian of h with respect to the coefficient of the parameter vector θ. But if θ_0^k is too far from the value $\hat{\theta}^k$ which makes $h(\hat{\theta}^k, \lambda^{k+1}, \theta^{k-1}) \approx 0$, then (11) may not converge. Thus to improve the convergence of (11) , we select

the initial point as $\hat{\theta}_0^k = \theta^k$. Suppose $|\hat{\theta}_{M+1}^k - \hat{\theta}_M^k| \leq \gamma$ then we set $\hat{\theta}_M^k = \hat{\theta}^k = \theta^{k+1}$. The initial point θ_0^k for the correction step is made closer to the desired solution by considering the following equation.

$$\hat{\theta}_0^k = \theta^k - \Delta\lambda J_{\theta}^{-1}[h(\theta^k, \lambda^{k+1}, \theta^{k-1})]\frac{\partial h}{\partial \lambda}[h(\theta^k, \lambda^{k+1}, \theta^{k-1})] \quad (12)$$

The derivation of (12) is analogous to the derivation of (5a) of Stonick and Alexander [16] for our homotopy map (10). Equation (12) corresponds to the prediction of the next point by taking a step in the direction of the path's slope. For the fixed point homotopy map considered, (12) becomes

$$\hat{\theta}_0^k = \theta^k - \Delta\lambda \left\{ \frac{I}{1 - (\lambda^k + \Delta\lambda)} - \frac{I}{(1 - (\lambda^k + \Delta\lambda))^2} \left[\frac{F_{\theta}^{-1}(\theta^k)}{(\lambda^k + \Delta\lambda)} + \frac{I}{1 - (\lambda^k + \Delta\lambda)} \right]^{-1} \right\} \{f(\theta^k) - (\theta^k - \theta^{k-1})\} \quad (13)$$

Where I is the identity mtrix. If $\hat{\theta}_0^k$ estimated by (12) is not on the path then it is taken as the initial point in the correction step (11). Otherwise $\hat{\theta}_0^k$ is considered as the next point on the path. If θ_0^k is not on the path, then the correction step is carried out by Newton's method .

5. PARAMETER ESTIMATION AND RESTORATION

Here we consider the MRF model suggested in [1] with line fields for a 256 gray level image. The a priori energy function for the model is $U(x, h, v, \phi) = \sum_{i,j} \alpha[(x_{i,j} - x_{i,j-1})^2(1 - h_{i,j}) + (x_{i,j} - x_{i-1,j})^2(1 - v_{i,j})] + \beta(h_{i,j} + v_{i,j})$
The corresponding posterior energy function is

$$U_p(x, h, v, \theta) = \sum_{i,j} \frac{(y_{i,j} - x_{i,j})^2}{2\sigma^2} + U(x, h, v, \phi) \quad (14)$$

Substituting (14) in (7) and (9) we obtain

$$f(\theta) = \sum_{i,j} \frac{\partial}{\partial \theta} [\{-U_p(x, h, v, \theta)\} - \log \left\{ \sum_{x_{i,j} \in G} e^{-U_p(x, h, v, \theta)} \right\}] \quad (15)$$

707

The basic steps in the algorithm for simultaneously updating x^k and θ^k are

Algorithm 1

1. Input noisy image Y.

2. Initialize the parameter vector θ to θ^0.

3. Given θ^k estimate x^{k+1} by minimizing

$$\frac{\|y - x^k\|^2}{2\sigma^2} + U(x^k, h^k, v^k, \phi^k)$$

Here, we used simulated annealing algorithm for minimization.

4. Having determined x^{k+1}, estimate θ^{k+1} using the homotopy map (10) and the corresponding update equations (11) and (13) and with $f(\theta)$ given in (15). In the parameter estimation part the estimated image x^{k+1} and the noisy image Y are used as input.

5. If a stopping criterion is met, stop or else go to Step 3.

In our simulations the following stopping criterion was used. If $\frac{1}{N^2}\sum_{i,j}(x_{i,j}^{k+1} - x_{i,j}^k)^2 \leq$ threshold, then stop. θ_0 is selected based on some apriori knowledge; for example we may have an idea about the parameters for a given class of image from prior experience.

6. SIMULATION AND RESULTS

Here in our simulation we have considered a typical *autopart* image as shown in Figure 1 to validate our approach. The corresponding noisy image with $SNR = 1dB$ is shown in Figure 2. The SNR is defined by SNR= $10.0\log_{10}\left\{\frac{1}{N^2}\sum_{i,j}(x_{i,j} - m)^2/\sigma^2\right\}$ where σ=standard deviation of the zero mean white Gaussian noise and m=mean of the image. The noisy image is the only input to Algorithm 1. With an initial guess of the parameter vector the SA block is executed first with the noisy image. This is run for 2000 iterations. The the estimated image of SA algorithm as shown in Figure 3 and the observed image of Figure 2 are input to the parameter estimator as the original and noisy image respectively. The Parameter thus estimated is used subsequently in the SA algorithm to obtain the image estimate. The initial parameter vector and the SA algorithm constitute the first iteration. However, the subsequent iterations consist of one parameter estimator followed by the SA algorithm. The SA algorithm parameters like initial temperature, number of iteration, cooling rate are specified at the outset and are unaltered during the subsequent iterations.

The initial parameter vector was choosen as $\theta_0 = [0.015, 10.0, 25.0]$. Figures 3, 4, and 5 are the image estimates obtained at 1st, 2nd and 3rd iteration respectively. The corresponding parameters are tabulated in Table 1. The SNR of the estimated image at 3rd iteration is 7.68, an improvement of $6.68dB$. Here we would like to remark that the estimated noise variance in the process of the joint solution is close to the actual one.

Table 1

itera-tion no.	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma^2}$
1	0.015	10.0	25.0
2	0.00766	3.16	37.02
3	0.001869	4.79	29.435

7. CONCLUSION

The proposed unsupervised scheme has been successfully tested on different class of images, due to space constraint we have presented the "autopart" image to reinforce the validity of our strategy.

Suppose we implement our scheme in a supervised mode, that is assuming the knowledge of the initial image, and estimate only the parameters. Let this parameter estimate be θ^*. Then from our simulation we found that $\hat{\theta}$, obtained from the unsupervised scheme is very close to θ^* [18]. Currently we are investigating the sensitivity of the algorithm to arbitrary choice of the initial guess θ^0.

8. REFERENCES

[1] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs distribution and Bayesian restoration of images", *IEEE Trans. on Pattern Anal. and Machine Intell.*, Vol-6, No. 6, pp 721-741, 1984.

[2] H. Derin, H. Elliot, R. Gristi and D. Geman, "Bayes smoothing algorithm for segmentation of binary images modelled by MRF fields.", *IEEE Trans. Patt. Annal. Machine Intell.*, Vol-6, pp 707-720, Nov 1984.

[3] F. R. Hansen and H. Elliot, "Image segmentation using simple Markov random field models", *Compu. Vision Graph. and Image Proce.*, Vol-20, pp 101-132, 1982.

[4] S. Kapoor, P. Y. Mundkur and U. B. Desai, "Depth Recovery using MRF.", *Sadhana*, Vol-18, Part-1, pp 17-29, March 1993.

[5] M. R. Bhatt and U. B. Desai, "Robust image rstoration algorithm using Markov random field model", *CVGIP: Graphical Models and Image Processing*, Vol-56, No-1, pp. 61-74, Jan 1994.

[6] G. Wahba, "Practical approximate solutions to linear operator equations when the data are noisy", *SIAM J. Numer. Anal.*, Vol.14, 1977.

[7] J. Besag, "On the statistical analysis of dirty pictures", *J. Roy. Stat. Soc. Series-B*, Vol-48, pp 259-302, 1986.

[8] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing", *IEEE Trans. Patt. Annal. and Machine Intell.*, Vol 11, pp 799-813, 1989.

[9] L. Younes, "Estimation and annealing for Gibbsian fields", *Ann. Inst. Henri Poincar'e.*, 24(2), pp 269-294, 1988.

[10] C. S. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields", *CVGIP: Graphical Models and Image Processing*, Vol. 54, No. 4, pp. 308-328, July 1992.

[11] B. S. Manjunath and R. Chellapa, "Unsupervised texture segmentation using Markov random fields", *IEEE Trans. Pattern Anal. Mach. Intelligence*, PAMI-13, pp 478-482, 1991.

[12] H. H. Hguyen and Paul Cohen, "Gibbs random fields, fuzzy clustering and the unsupervised segmentation of textured images", *CVGIP: Graphical Models and Image Processing*, vol-55, No-1, pp. 1-19, Jan 1993.

[13] M. G. Kang and A. K. Katsaggelos, "Simultaneous iterative image restoration and evaluation of the regularization parameter", *IEEE Trans. on Signal Processing*, Vol-40, No-9, pp. 2329-2334, Sept 1992.

[14] W. I. Zangwill and C. B. Garcia, "Pathways to solutions fixed points and equlibria", *Prentice-Hall Inc, Englewood Cliffs*, 1981.

[15] L. T . Watson, "Numerical linear algebra aspects of globally convergent homotopy methods", *SIAM Review*, Vol. 28, No-4, pp 529-545, Dec 1986.

[16] V. L. Stonick and S. T. Alexander, "A Relationship between recursive least square update and homotopy continuation methods", *IEEE Tran. Signal Processing*, Vol. 39, No. 2, pp 530-532, Feb. 1991.

[17] V. L. Stonick and S. T. Alexander, "Globally optimal rational approximation using homotopy continuation methods", *IEEE Tran. Signal Processing*, Vol. 40, No. 9, pp 2358-2361, Sept 1992.

[18] P. K. Nanda, U. B. Desai and P. G. Poonacha, "A homotopy continuation method for parameter estimation in MRF models and image restoration", *Tech. Report SPANN-94.1*, Deptt.of Electrical Engg. IIT-Bombay, India 1994.
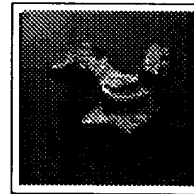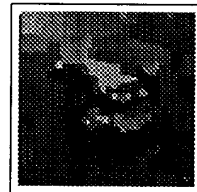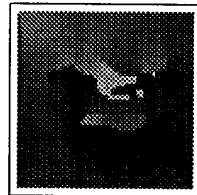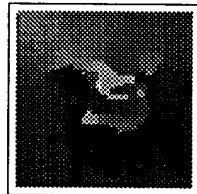
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5