Weighted Median Predictive Techniques for Coefficient Estimation in NonGaussian Markov Random Fields

Sean Borman (borman.1@nd.edu)
Ken Sauer (sauer.1@nd.edu)
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
(219)631-6999(Tel.) 631-4393(Fax)

Charles Bouman (bouman@ecn.purdue.edu)
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907-0501
(317)494-0340(Tel.) 494-6440(Fax)

Abstract- NonGaussian Markov image models are effective in the preservation of edge detail in Bayesian formulations of restoration and reconstruction problems. Included in these models are coefficients quantifying the statistical links among pixels in local cliques, which are typically assumed to have an inverse dependence on distance among the corresponding neighboring pixels. Estimation of these coefficients is a nontrivial task for NonGaussian models. We present results for coefficient estimation for a model which is particularly effective for edge preservation and noise suppression, using a predictive technique analogous to estimation of the weights of optimal weighted median filters.

1 Introduction

A variety of probabilistic image restoration and reconstruction problems call for non-Gaussian image models for preservation of discontinuities. A popular choice is the Markov random field (MRF)[1], with probability density function

$$p(x) = Z^{-1} \exp\left\{ -\sum_i \sum_{j\in\mathcal{N}_i} a_{ij}\rho(x_i, x_j; \theta) \right\}. \quad (1)$$

$\mathcal{N}_i$ is the spatial neighborhood of pixel i, $\rho(\cdot;\theta)$ is the potential function, and Z a normalizing constant. The vector θ generally includes one parameter specifying the shape of the potential function, and another indicating the scale of the data. The latter parameter is expressed by the local covariance in a Gaussian MRF.

We concentrate on the generalized Gaussian MRF (GGMRF)[2] in this paper, which is described by (1) with

$$\rho(x_i, x_j; \theta) = \left| \frac{x_i - x_j}{\sigma} \right|^q, \quad 1 \le q \le 2. \quad (2)$$

The GGMRF possesses several desirable properties, including invariance to scaling, convexity of the log prior density in x, and preservation of edges. The value of $q = 1$ is particularly effective for the recovery of discontinuities.

The formulation of the MAP problem requires either assumption of knowledge of the parameters $\{a_{ij}\}$ and θ, or estimation of their values from data. Recently, the estimation of σ, or equivalently the temperature parameter, was shown to be trivial for the GGMRF[3]. In most previous work on MRF parameter estimation, the set of a_{ij}, which we will denote $\mathbf{a}$, are arbitrarily fixed, with the values in $\mathbf{a}$ inversely related to the spatial distance between pixels i and j. There is no reason, however, to assume that this choice is always most accurate.

This paper discusses the estimation of the coefficients $\mathbf{a}$ with $q = 1$ in (2). The cost function $\log p(x)$ calls for an optimization step at each pixel in Bayesian estimation which has the same form as a weighted median filter (WMF) operation[4], with weights corresponding directly to $\mathbf{a}$. Due to this similarity, we exploit techniques for optimal weight selection for WMFs in seeking to estimate the coefficients of the GGMRF.

2 Nonlinear Prediction and Coefficient Estimation

Maximum likelihood (ML) parameter estimation is often applied to problems of the sort discussed here. Unfortunately, although the exponent in (1) is a simple function of the coef-

This research supported by the National Science Foundation under grant No. MIP93-00560.

ficients, the normalizing constant Z has a dependence which is sufficiently complicated to make optimization of the log likelihood function for ML estimation apparently infeasible. This is a classic problem in most MRF parameter estimation[1].

We will focus on techniques to attack this problem for the GGMRF with $q = 1$. We assume the MRF to be stationary, which means a number of distinct a_{ij} equal only to the number of pixels in the each neighborhood $\mathcal{N}_i$. Additionally, symmetry of coefficients is implied by Markov properties, and we assume nonnegativity to maintain convexity in the potential function. Since the end goal of the parameter estimation is Bayesian reconstruction, we consider first the optimization posed by $\log p(x)$ in sequentially updating pixels' values. If we let k index across unique relative positions of neighbors, and let $x_{(i,k)}$ represent the neighboring pixel to position i dictated by k, the solution is

$$\hat{x}_i = \arg\min_{\beta} \sum_k a_k |\beta - x_{(i,k)}|. \tag{3}$$

Suppose the pixel values $x_{(i,k)}$ are ordered from smallest to largest, forming a set $\tilde{x}_m$ with the property that if $i < j$, $\tilde{x}_i \leq \tilde{x}_j$, and the associated coefficients are similarly $\tilde{a}_k$. The solution to (3) is then $\tilde{x}_{k_T}$ with k_T the smallest index such that $\sum_{m=1}^{k_T} \tilde{a}_k \geq 0.5$, shown as the zero crossing in Fig. 1. This is equivalent to the weighted median of the neighboring pixel values, with weights normalized to total unity. Thus, MAP estimation under this MRF resembles a weighted median prediction, with $\mathbf{a}$ equivalent to relative weights.

Given this similarity, techniques developed for the selection of optimal weights in a WMF[4, 5] may prove useful in our estimation of coefficients in the GGMRF. Similarly to optimal linear filters, WMF weights may be chosen with the goal of minimizing mean squared[4], or mean absolute error[6] in the output of the filter. The form of the optimization in (3) is similar to a prediction filter, since the center pixel's value is not weighted. This makes our problem one of finding an optimal predictor of each pixel. The prediction is based on the weighted median operation, and therefore nonlinear.

Our approach will be coefficient selection by minimization of nonlinear prediction error, with the choice of penalty for error influenced by the form of the conditional marginal pixel distribution illustrated in Figure 1. While the plotted log PDFs are not of a common form, they approximate a quantized linear derivative over the region of primary interest, and the center pixel may be approximated as conditionally Gaussian, with mean at the weighted median of the neighbors. We minimize $E[(X_i - \hat{X}_i)^2]$, where $\hat{X}_i$ is the nonlinear prediction of pixel i from its neighborhood, and $E[]$ is the expectation operator. Note that using this form of inference on the GGMRF, no "structural constraints"[6] are necessary to form the predictor. In common applications, we rely on ergodicity and stationarity in using spatial averages in place of expectations.

As illustrated in Figure 1, the choice of $\hat{X}$ requires minimization of a piecewise linear convex function. The derivative is a sum of step functions. Thus two types of solutions to the prediction of a pixel are possible: (1) the log of the conditional density of the center pixel has zero derivative over an interval, making the maximum *a posteriori* probability prediction nonunique, or (2) the maximum of the log density is at a nondifferentiable point of the curve. In the former case, we use the midpoint of the interval as $\hat{X}$. Viewing these two cases in light of the MAP prediction of the center pixel, we arrive at both an abbreviation of the optimization problem necessary to estimate $\mathbf{a}$, and possible nonuniqueness of the minimum mean-squared error predictive estimate of the MRF coefficients.

To see this feature, consider any fixed ordering of a set of local pixel values, as in Figure 1, with either the unique or non-unique zero crossing of the derivative. In either case, the identical value of $\hat{X}$ can be preserved under quantization of the coefficients' values to multiples of $(2N)^{-1}$, where N is the total number of neighbors to each pixel. Thus, whatever the cost with which we penalize the prediction error, all possible values of the total penalty under variation of the members of $\mathbf{a}$ will be evaluated if we search on a grid of resolution $(2N)^{-1}$. While this makes the search for the optimal coefficient values relatively simple for modestly sized neighborhoods, it also results in non-uniqueness of the solution. By increasing the resolution beyond the required resolution in the vicinity of the minimum, we can make our coefficient estimate in general a *set* estimate, rather than point estimate, of **a**.

3 Experiments

We have performed several experiments to test the effectiveness of the WMF predictive method in estimating GGMRF coefficients. The more controlled trials involve images such as those of Figure 2, which represent samples from a chain of images with a known GGMRF. This allows us to ascertain exactly the accuracy of our estimates. The second is the type of problem motivating much of this work. This phantom represents a cross section of a human torso, taken from a hand-segmentation of a CT reconstruction, and used for attenuation estimation purposes in emission (PET) tomographic studies[7].

Minimization of weighted median prediction error appears to accurately estimate the form of the GGMRF model in sample fields, with precision limited by possible nonuniqueness as discussed above. For common, simple choices of the model, such as first-order processes with equal weights, estimates are exactly and reliably correct. With more arbitrary patterns, estimates are as precise as the "quantization" effect in values for coefficients discussed above permits. We may choose to increase the resolution of the search to better define the interval of the estimate.

The upper image of Figure 2 was generated from a second-order GGMRF with $q = 1$ and symmetric coefficients of $[0.0, 0.17, 0.0, 0.33]$. Estimates of these coefficients, with search on a grid of resolution 0.0625, are on the interval $[0.0, 0.125 \pm 0.0625, 0.0, 0.375 \pm 0.0625]$. The lower image in Figure 2 is from a second-order GGMRF with $q = 1$ and symmetric coefficients of $[0.2, 0.15, 0.1, 0.05]$. The estimate computed at resolution 0.0625 was $[0.1875, 0.125, 0.125, 0.0625]$. The cross section of Figure 3 yields coefficient estimates of [0.0, 0.125, 0.0625, 0.3215]. It is this type of function which especially benefits from edge-preserving models such as the GGMRF at $q = 1$[2].

4 Conclusion

Weighted median prediction produces accurate estimates of MRF coefficients with potential function $|x|$. While the GGMRF with $q = 1$ is emphasized here, reconstructions under this model are continuous in q, and for tractable optimization values near, but greater than one are very useful. For any $q > 1$, the log *a posteriori* density of the predicted pixel is everywhere differentiable and the MAP prediction is unique. In this case, greater precision of coefficient estimates is anticipated. The results shown here may serve well even in this case as initial estimates, from which gradient descent may be employed to seek an optimal estimate.

Acknowledgement

The authors thank Jeff Fessler and Gary Hutchins of the University of Michigan for generously providing tomography phantom data.

References

[1] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. PAMI*, vol. PAMI-6, no.6, pp. 721-741, Nov. 1984.

[2] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation," *IEEE Trans. Image Proc.,* vol. 2, no. 3, pp. 296-310, July 1993.

[3] C. Bouman and K. Sauer, "Maximum likelihood Scale Estimation for a Class of Markov Random Fields," *Proc. ICASSP '94*, Adelaide, Australia, pp. V537-V540.

[4] R. Yang, L. Yin, M. Gabbouj, and T. Neuvo, "2-D Optimal Weighted Median Filters," *Proc. 1992 IEEE Wkshp. Vis. Sig. Proc. & Comm.,* Raleigh, NC, Sept. 2-3, 1992, pp. 31-35.

[5] L. Yin, J. Astola, and Y. Neuvo, "Adaptive Weighted Median Filtering under the Mean Absolute Error Criterion," *Proc. 1991 IEEE Wkshp. Vis. Sig. Proc. and Comm.,* June 5-7, 1991, Hsingchu, Taiwan, pp. 184-187.

[6] M. Gabbouj and E. Coyle, "Minimum Mean Absolute Error Stack Filtering with Structural Constraints and Goals," *IEEE Trans. Acoust. Speech & Sig. Proc.,* vol. ASSP-38, no. 6, June 1990, pp. 955-968.

[7] J. Fessler, "Mean and Variance of Implicitly Defined Biased Estimators (such as Penalized Maximum Likelihood): Applications to Tomography," to appear in *IEEE Trans. on Image Proc.*

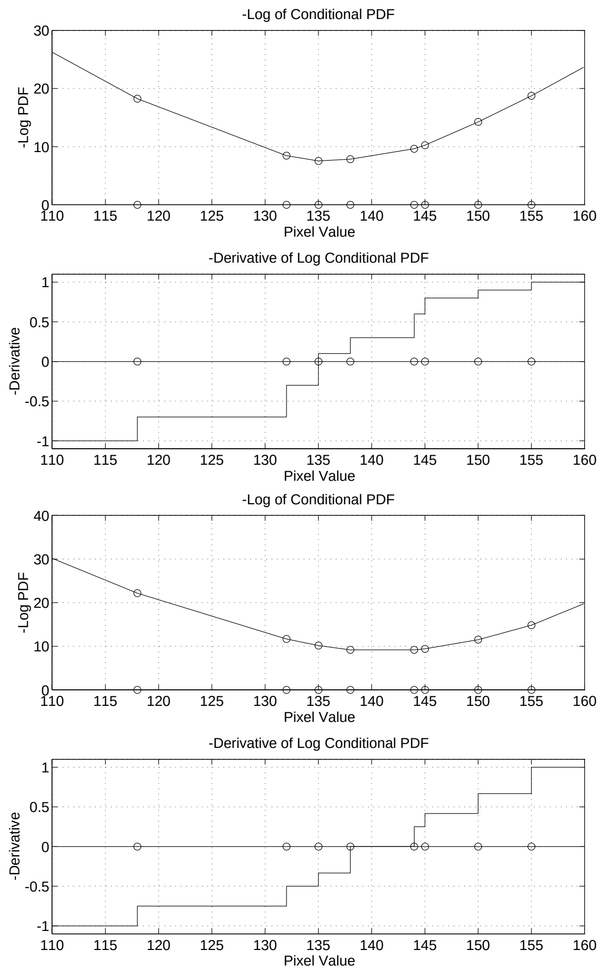

Figure 1: Optimization formed by choice of pixel conditioned on neighbors' values. First and third plots show negative logarithm of conditional PDF of single pixel with neighbors taking on marked values. The magnitudes of the step functions making up the derivative plots are twice the respective coefficients. (Upper) Unique choice for "weighted median" output. (Lower) Interval of non-unique choices for output; midpoint of interval used for prediction.

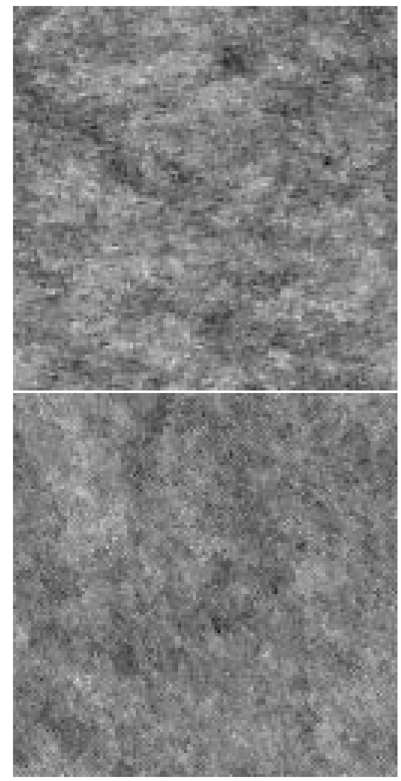

Figure 2: Images generated via the Metropolis algorithm under the GGMRF with $q = 1$. Upper sample was generated with coefficients (clockwise from upper left in surrounding neighborhood) [0.0, 0.17, 0.0, 0.33], while the lower sample was generated with coefficients [0.2, 0.15, 0.1, 0.05].

Figure 3: Phantom of cross section of human torso, manually segmented from CT scan.