Nonlinear Prediction Methods for Estimation of Clique Weighting Parameters in NonGaussian Image Models

Sean Borman†, Ken Sauer† and Charles Bouman*

†Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556

* School of Electrical Engineering, Purdue University, West Lafayette, IN 47907-0501

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 edge-preserving models which are 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

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

$\mathcal{C}$ is the collection of all cliques among neighboring pixels i, j, $\rho(\cdot, \cdot; \theta)$ 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.

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. Though a good deal of image reconstruction and estimation has included *ad hoc* choices for these parameters, much work has also been done in estimating them from either realizations of X or, in an incomplete data setting, from indirect observations for unsupervised reconstruction(see, e.g. Ref. 2 and bibliography). For Bayesian reconstruction, the scale parameter determines the degree of regularization, and is therefore of obvious importance. 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 or most useful.

This paper discusses estimation of $\mathbf{a}$ from observations of samples of X. In the absence of an apparent tractable form for maximum-likelihood (ML) estimation of these parameters due to the difficulty of evaluating the dependence of Z on $\mathbf{a}$, we investigate an approach which operationally resembles both parameter estimation via cross-validation,[3] and certain techniques for the design of weighted median filters.[4] Like these other approaches, our attempt is based on penalization of prediction error among image pixel values, with the form of prediction dictated by the particular MRF at hand. The prediction of the value of a pixel, given the remainder of an image, is a nonlinear function of neighboring pixels under edge-preserving, non-Gaussian MRFs. Penalization of the expected value, or observed average, of the error of this prediction forms a risk in terms of $\mathbf{a}$ whose minimization defines our estimate.

The following discussion and examples concentrate on the generalized Gaussian MRF (GGMRF),[5] which is described by (1) with

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

Author email addresses:borman.1@nd.edu, sauer@nd.edu, bouman@ecn.purdue.edu

but the same ideas apply to other MRFs as well. 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. We first discuss the estimation of the coefficients $\mathbf{a}$ with $q = 1$ in (2). In this case, the 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 begin by linking our estimation process to techniques for optimal weight selection for WMFs in seeking to estimate the coefficients of the GGMRF at $q = 1$. We then generalize the process for arbitrary values of q.

2. NONLINEAR PREDICTION AND COEFFICIENT ESTIMATION

Maximum likelihood (ML) parameter estimation is often applied to inference problems of the sort discussed here. Unfortunately, although the exponent in (1) is a simple function of the coefficients in $\mathbf{a}$, the normalizing constant Z has a dependence which is sufficiently complicated to make optimization of the log likelihood function for ML estimation apparently quite difficult. This is a classic problem in much MRF parameter estimation.[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 determined by $\mathcal{C}$. Additionally, symmetry of coefficients is implied, and we assume non-negativity to maintain convexity in the potential function. Since the end goal of the parameter estimation is Bayesian reconstruction, we begin by considering the optimization posed by $\log p(x)$ in sequential MAP updating of 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

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

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) when $q = 1$ 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 the lower plots of 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. (In the plots on the right of Fig. 1, we choose the midpoint of the specified interval as the predicted value.) Figure 1 shows also the form of the cost function which is minimized for local pixel updates.

Given this similarity, techniques developed for the selection of optimal weights in a WMF[4,6] may prove useful in our estimation of coefficients in the GGMRF. Similarly to optimal linear filters, WMF weights have been chosen with the goal of minimizing mean squared,[4] and mean absolute error in the output of the filter as a function of the weights. This optimization may be performed under constraints of positivity, or required to preserve specific image structure.[7] 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 nonlinear predictor of each pixel by our estimate of $\mathbf{a}$.

When $q > 1$, the log of the 1-D conditional PDF is differentiable, and the prediction is unique, as illustrated in Figure 2. This case no longer uses a weighted median as prediction value, but the form of the plots and the output value, when compared to the case $q = 1$ with the same neighboring pixel values in the left side of Fig. 1, show that the behavior of the predictor is continuously and relatively slowly varying in q.

Our estimation approach will be coefficient selection by minimization of nonlinear prediction error (MPE), with the choice of penalty for error the critical choice in definition of the estimator. In common applications, we rely on ergodicity and stationarity in using spatial averages in place of expectations. If we minimize the "natural" choice, mean squared error, $C(\hat{X}_i) = E[(X_i - \hat{X}_i)^2]$, where $\hat{X}_i$ is the nonlinear prediction of pixel i from its neighborhood, and $E[\,]$ is the expectation operator, we have the advantage of greater analytical tractability and possibly greater numerical stability. Should the MRF be Gaussian, minimization of mean squared error yields a pixel predictor which selects the conditional mean of each pixel as its prediction, coinciding with the MMSE criterion of optimality, as well as conditional median and MAP prediction of the pixel in the Gaussian case. Under nonGaussian models, the choice is not so obvious, and there is a good argument for matching the penalization of prediction error with the criterion

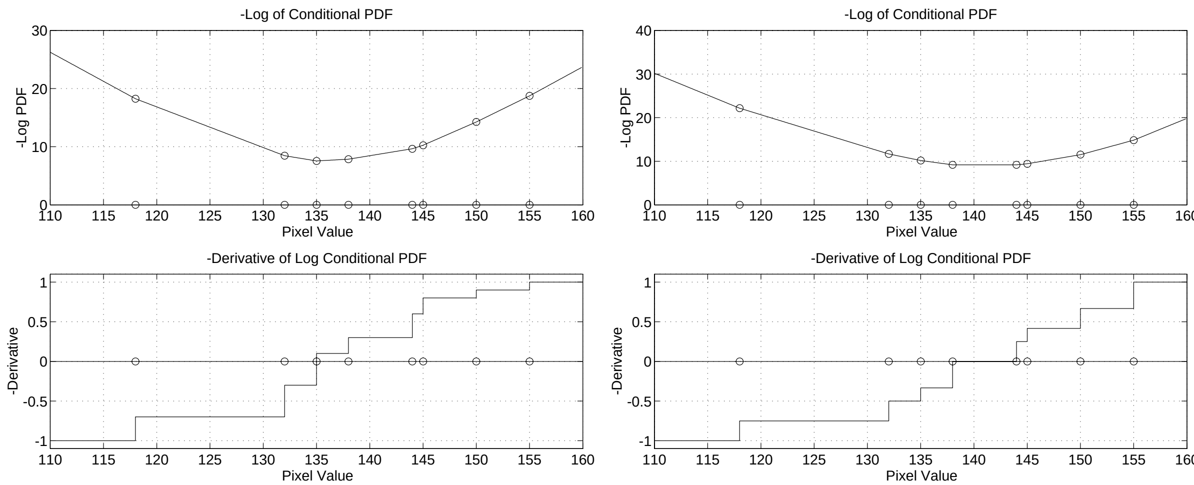

Figure 1. Optimization formed by choice of pixel conditioned on neighbors' values. The two upper plots show negative logarithm of conditional PDF of single pixel with neighbors taking on values marked by circles. The magnitudes of the step functions making up the derivative plots are twice the respective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

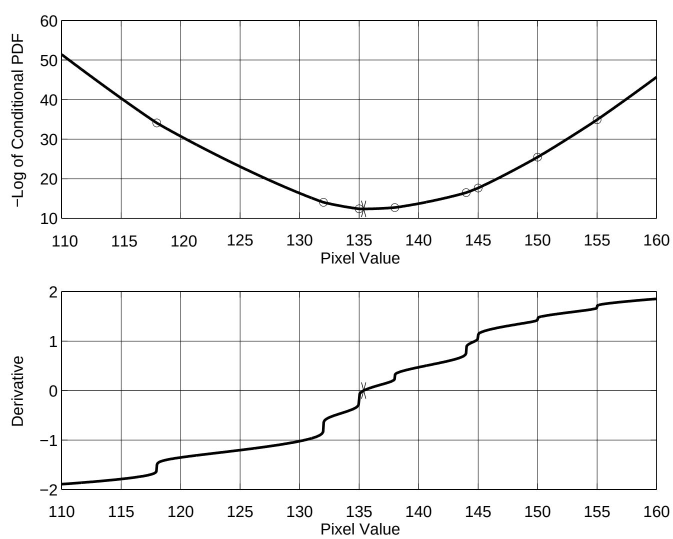

Figure 2. Negative logarithm of the conditional PDF of a pixel given its neighbors (above) and its derivative (below) for a GGMRF with q=1.3.

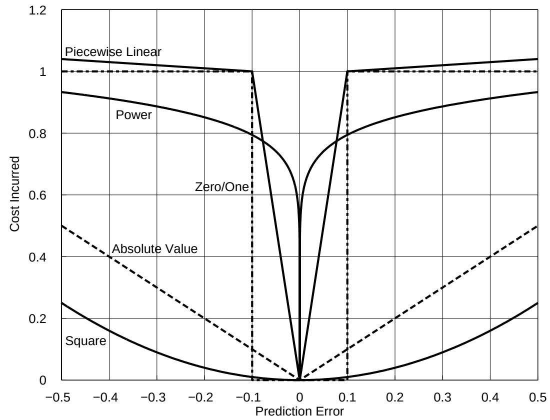

Figure 3. Cost functionals used for penalization of prediction error in MPE estimation. The cost functionals include the absolute value cost, squared cost, zero/one cost ($\alpha = 0.1$), power function cost ($\gamma = 0.1$), and piecewise linear cost ($\alpha = 0.1, \beta = 0.1$).

used to form the prediction. Since MAP estimators result from optimizing the cost function for the error Δ in the estimate $\hat{X}$ of the unknown parameter X

$$C(\Delta) = \begin{cases} 0 & \Delta = 0 \\ 1 & \Delta \neq 0 \end{cases} \tag{4}$$

our MAP pixel predictor under the chosen coefficients, if they should match the actual image model correctly, must minimize the risk, or average of the cost in (4). Thus, if the MAP pixel prediction is a consistent estimate, using this penalty will asymptotically yield the true parameters in **a**. To allow non-zero probability of having non-degenerate estimates, we must approximate (4); a few possibilities are shown in Figure 3. In the following results we will deal primarily with the quadratic cost and Δ^γ, which approaches (4) for small γ. As we will see, there are trade-offs to be made between this sense of optimality of coefficient estimate and difficulty of the associated optimization.

The behavior of the estimator depends also on the type of MRF modeled applied. As we depart from the Gaussian ($q = 2$), with its linear predictor, toward $q = 1$, the value of MAP pixel predictions show less simple dependence on the estimated entries in **a**, as is clear from the plots of Figures 1 and 2. The risk, approximated by the sample average of penalized prediction error, evolves as well, with generally greater ambiguity for smaller q. In Figure 4, for example, the case of a first-order MRF with mean squared error as penalty is illustrated. Due to symmetry, positivity, and normalization of coefficients to total unity, this problem has only one degree of freedom, and the horizontal axis represents the value of each member of one symmetric pair of coefficients. (Total weighting of pixel interactions is incorporated into the scale, or temperature, parameter, and may be estimated independently of the normalized **a**.)

3. OPTIMIZATION ISSUES

As illustrated in Figure 1, the choice of $\hat{X}$ requires minimization of a piecewise linear convex function when $q = 1$. 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 non-unique, 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 **a**, and possible non-uniqueness 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

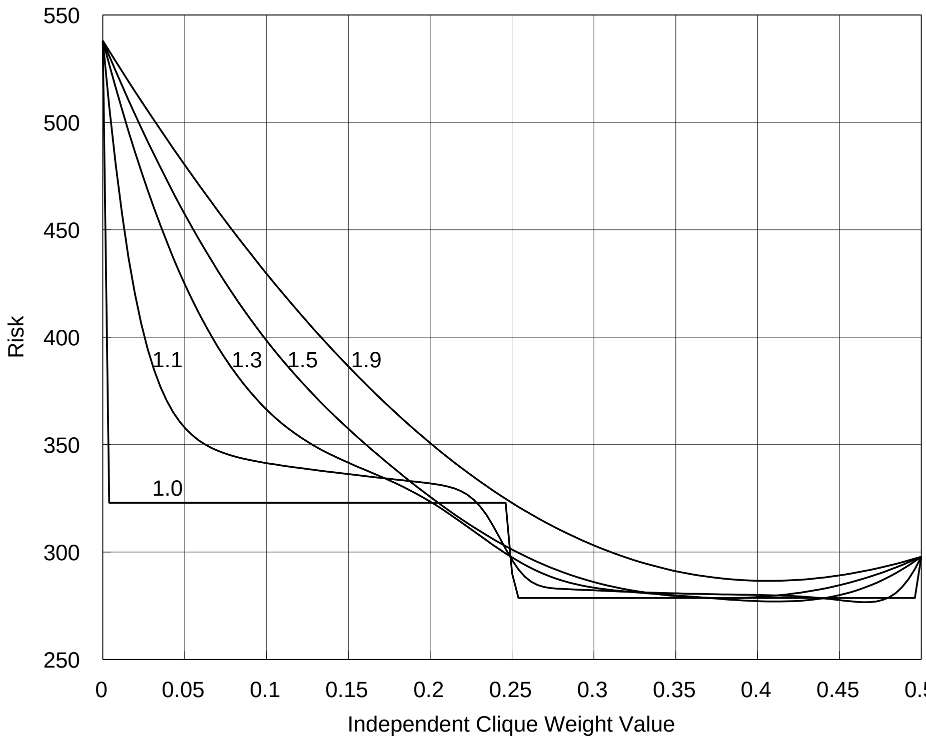

Figure 4. Risk as a function of independent clique weight for a 1^{st} order GGMRF predictor for various values of q. The input realization is the "lenna" test image. The square of the prediction error penalty was used.

under variation of the members of **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**.

For $q > 1$, the function plotted in Figure 2 is differentiable; consequently the risk is also a differentiable function of coefficients **a** *provided the prediction error penalty function is differentiable.*[8] Unfortunately, the penalty corresponding to MAP prediction, or its approximations from Figure 3 are non-differentiable at $\hat{X}_i = X_i$. This is reflected in an optimizational disadvantage for the penalty promoted earlier as best matching the simple MAP pixel predictor, also shown graphically in Figure 5. In Figure 5 and text below, we give coefficients for second order (eight neighbor) MRFs in sets of four entries representing the first four values beginning at the upper left. The remainder are symmetric with these.

Optimization techniques applied in our work thus far fall into three categories: (1) for $q = 1$ and constant-valued segments of risk functions, exhaustive search with "quantization" as discussed above; (2) for $q > 1$ and differentiable prediction error penalties, gradient descent; (3) for nondifferentiable penalties, and for avoiding local minima in case (2), simulated annealing. Though we have met good success with annealing in general,[8] the second case in Figure 6 shows an estimate which is significantly different from the global minimum.

4. EXPERIMENTS

We have performed several experiments to test the effectiveness of the predictive methods in estimating GGMRF coefficients. The more controlled trials involve images such as those of Figure 7, 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 set of images in Figure 8 is more realistic, providing the type of problem motivating this work.

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

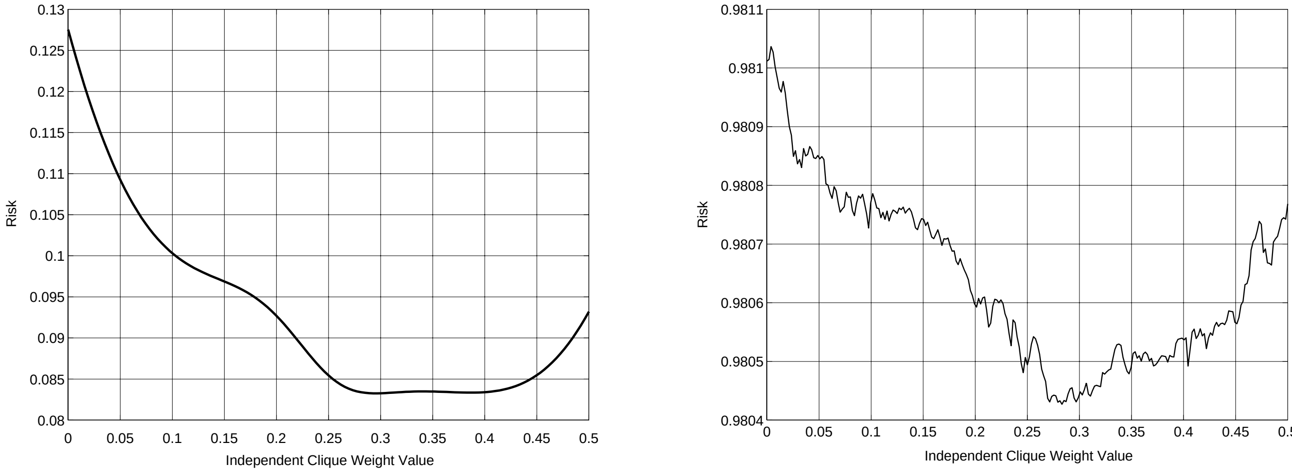

Figure 5. Left: Risk as a function of clique weight parameter for first-order GGMRF with $q = 1.3$ and prediction error penalty Δ^2. The input realization is a GGMRF with clique weight vector $\frac{1}{6}[0\ 2\ 0\ 1]^T$. Right: Risk for similar case with prediction error penalty $\Delta^{0.05}$.

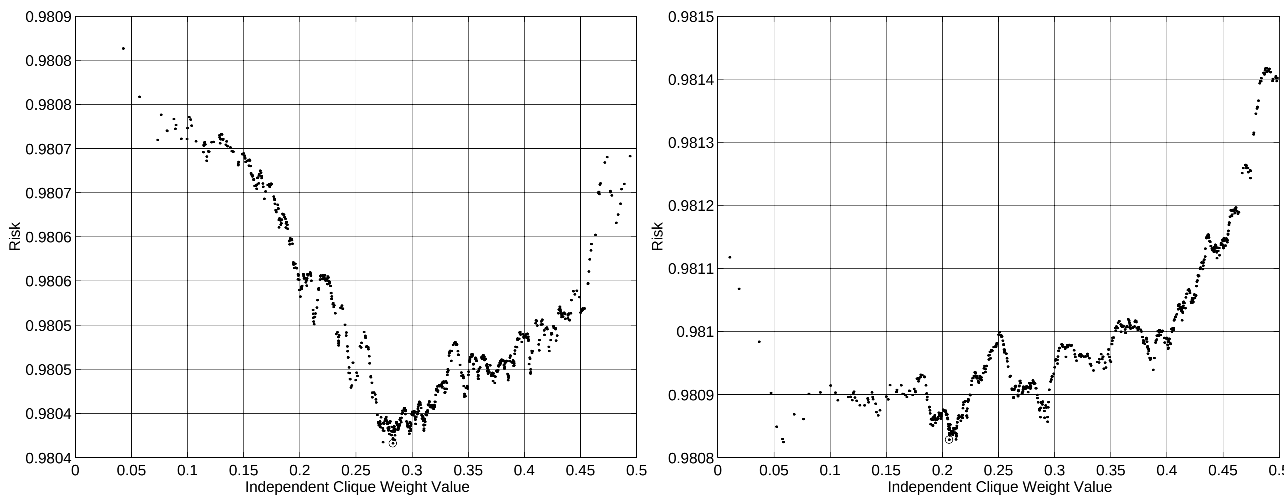

Figure 6. Sequence of estimates (one point per state) selected during iterations of simulated annealing with Metropolis algorithm for updates. In each case, final estimate is illustrated by circle. Minimum of risk in right plot is near 0.06.

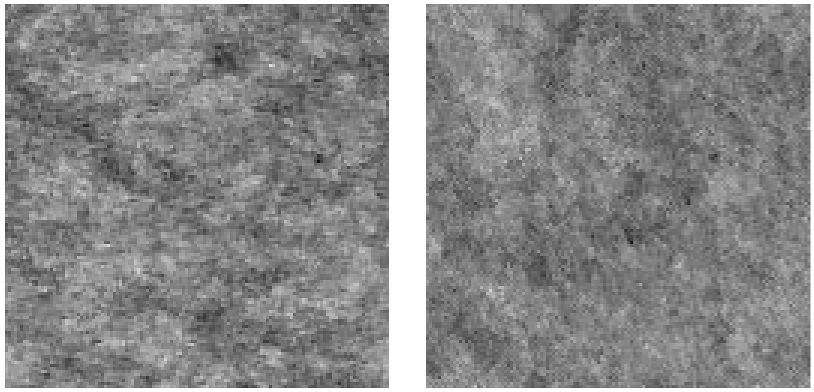

Figure 7. Images generated via the Metropolis algorithm under the GGMRF with $q = 1$. Left sample was generated with coefficients (clockwise from upper left in surrounding neighborhood) [0.0, 0.17, 0.0, 0.33], while the sample on right was generated with coefficients [0.2, 0.15, 0.1, 0.05].

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 left image of Figure 7 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 right image in Figure 7 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 images of Figure 8 yield coefficient estimates with $q = 1$ shown in Table 1 which are quite different in some cases from those which are typically assigned *ad hoc*.

In Table 2 we present the results of coefficient estimates from Monte Carlo simulations of images with GGMRF distributions and $q = 1$. The optimization is performed via gradient descent of the risk function, following initialization by the "quantized" coefficient estimate described above for the estimator under $q = 1$. For the sake of uniqueness of solutions, we use $q = 1.01$ in the estimator formulation. In spite of the fact that gradient descent is better suited to the smoother risk functions of the quadratic error penalty used in the center column, we find in some cases that the approximation to the consistent estimator under $C(\Delta) = |\Delta|^{0.05}$ yields substantially more accurate coefficient choices. Further study on these estimates will assess the potential improvements of these more accurate choices for clique weighting MRF coefficients.

5. CONCLUSION

Nonlinear prediction produces accurate estimates of MRF coefficients with potential function $|x|^q$, and can easily be applied to other MRF models. The optimization of the best-performing estimator, with penalty approximating the zero/one cost function, is computationally costly, but may be aided by good initial conditions from a simplified estimator. Though we have here estimated strictly from direct observations of images, it is also possible to extend these techniques to estimation from indirect observation through noisy, transformed data.

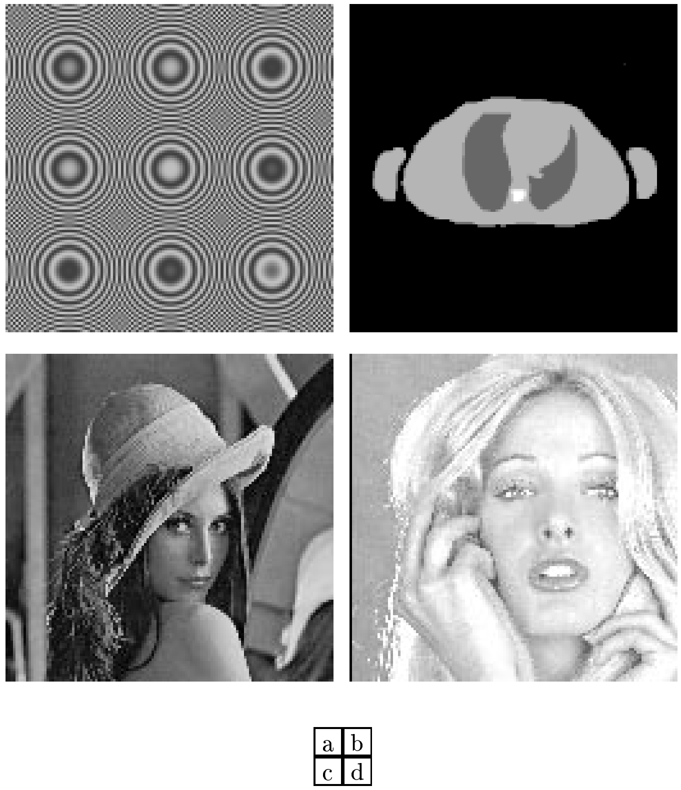

Figure 8. Above: Test images (128 pixels square) from which MRF clique weight parameters were estimated. (a)Aliased radial chirped sinusoid; (b)Hand segmented chest CT phantom (courtesy J. Fessler, G. Hutchins, Univ. of Mich.); (c)Lenna; (d)Tiffany.

Test Image	1^{st} Order Estimate			2^{nd} Order Estimate		
chirp		0.25		0.1875	0.0625	0.1875
	0.25		0.25	0.0625		0.0625
		0.25		0.1875	0.0625	0.1875
phantom		0.125		0.0625	0.1250	0.0625
	0.375		0.375	0.2500		0.2500
		0.125		0.0625	0.1250	0.0625
lenna		0.375		0.0000	0.3125	0.1250
	0.125		0.125	0.0625		0.0625
		0.375		0.1250	0.3125	0.0000
tiffany		0.375		0.0625	0.3125	0.0625
	0.125		0.125	0.0625		0.0625
		0.375		0.0625	0.3125	0.0625

Table 1. Performance of 1^{st} and 2^{nd} order, midpoint prediction, $q = 1$, estimates for various input test images.

ACKNOWLEDGMENTS

This research supported by the National Science Foundation under grants MIP93-00560 and MIP97-07763.

True Clique Weights			Estimated Clique Weights $C(\Delta) = \lvert\Delta\rvert^2$			Estimated Clique Weights $C(\Delta) = \lvert\Delta\rvert^{0.05}$		
0.0	0.0	0.0		0.001414			0	
0.5		0.5	0.498586		0.498586	0.5		0.5
0.0	0.0	0.0		0.001414			0	
0.0	0.5	0.0		0.499882			0.5	
0.0		0.0	0.000118		0.000118	0		0
0.0	0.5	0.0		0.499882			0.5	
0.0	0.0	0.5		0.260479			0.249852	
0.0		0.0	0.239521		0.239521	0.250148		0.250148
0.5	0.0	0.0		0.260479			0.249852	
0.5	0.0	0.0		0.245851			0.135741	
0.0		0.0	0.254149		0.254149	0.364259		0.364259
0.0	0.0	0.5		0.245851			0.135741	
0.00	0.25	0.00		0.252171			0.233276	
0.25		0.25	0.247829		0.247829	0.266724		0.266724
0.00	0.25	0.00		0.252171			0.233276	
0.000	0.333	0.000		0.382457			0.321800	
0.167		0.167	0.117543		0.117543	0.178200		0.178200
0.000	0.333	0.000		0.382457			0.321800	
0.000	0.167	0.000		0.096219			0.164676	
0.333		0.333	0.403781		0.403781	0.335324		0.335324
0.000	0.167	0.000		0.096219			0.164676	

Table 2. Gradient descent clique weight vector estimates using a 1^{st} order, GGMRF MPE estimator for various input MRF realizations.

REFERENCES

1. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **PAMI-6**, pp. 721–741, Nov. 1984.
2. S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," *to appear inIEEE Trans. on Image Processing* .
3. S. J. Reeves, "A cross-validation framework for solving image restoration problems," *Journal of Visual Communication and Image Processing* **3**, pp. 433–445, December 1992.
4. R. Yang, L. Yin, M. Gabbouj, and T. Neuvo, "2-d optimal weighted median filters," in *Proc. of the IEEE Workshop on Vis. Sig. Proc. and Com.*, pp. 31–35, (Raleigh, NC), Sept. 2-3 1992.
5. C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. on Image Processing* **2**, pp. 296–310, July 1993.
6. L. Yin, J. Astola, and Y. Neuvo, "Adaptive weighted median filtering under the mean absolute error criterion," in *Proc. of the IEEE Workshop on Vis. Sig. Proc. and Com.*, pp. 184–187, (Hsingchu, Taiwan, Republic of China), June 5-7 1991.
7. M. Gabbouj and E. Coyle, "Minimum mean absolute error stack filtering with structural conststraints and goals," *IEEE Trans. on Signal Processing* **ASSP-38**, pp. 955–968, June 1990.
8. S. Borman, *On Parameter Estimation for a Class of Markov Random Field Image Models.* Master's thesis, University of Notre Dame, Notre Dame, IN, May 1996.