INFERENCE FOR HIDDEN MARKOV MODELS

by

Lars Smedegaard Andersen
Department of Statistics GN-22
Padelford Hall
University of Washington
Seattle, WA 98195
U.S.A.

present address:
H. Lundbeck A/S
Ottiliang 7–9
DK-2500 Valby
Denmark

ABSTRACT

Bayesian inference techniques can be applied to hidden Markov models for drawing inference concerning many features of interest in the unobserved signal. By resampling from the posterior of the signal conditionally on the observation, optimal restoration can be performed and local measures of variability can be found. The restoration estimate is qualitatively different from those of Iterative conditional modes and simulated annealing. The restoration algorithm described here is optimal with respect to a local loss function and consequently has little known global characteristics. Specific problems of interest in image analysis are studied in simple examples.

Research supported in part by the Office of Naval Research under contract number N00014-85-K-0422, on spatio temporal stochastic processes (Antonio Possolo principal investigator)

1. Introduction

The recent interest in statistical imaging has to some extent focused on procedures to re-estimate a signal from observation of a noisy reception thereof, Besag (1986) and Geman and Geman (1984). In this paper we investigate a technique based on sampling directly from the posterior distribution. These techniques are useful for both restoration and inference for spatial features in the signal. Grenander (1981) suggested the inference machine but this work has since received surprising little attention, we believe. Sampling from the posterior distribution was mentioned in Geman and Geman (1984) and studied in Marroquin, Mitter and Poggio (1987) in the context of restoration.

We believe that the method studied here is of particular interest when restoration is of secondary interest or if analysis of residuals is desired. The local measure of uncertainty of the restoration can be used to compare residuals, Haslett and Horgan (1986, 1987).

The methods described here are applicable to most discrete models and many models with continuous grey levels. The Gibbs sampler can be applied also for many continuous Markov random fields, Andersen (1989) and Revuz (1975).

Although the special cases studied in this paper involve only spatially independent noise more complicated degradation mechanisms can be accommodated by straight forward generalizations. Geman and Geman (1984) give a much more general model for which the results here are still valid. Line processes can be incorporated in several ways, by letting the signal consist of two layers of Markov random fields or just treat the signal as multidimensional, we shall not, however, venture into this.

1.1 The basis image models

We shall throughout this paper work in the following set up. Let $T \subseteq \mathbf{Z}^d$ be a finite rectangular lattice, and let x_T be an unobserved image on T we assume that x_T is a realization of a stationary discrete Markov random field $\mathbf{X}_T$, with respect to a translational invariant neighborhood system; we avoid edge effects by thinking of T as wrapped around a torus in $\mathbf{R}^{d+1}$. The observation available is y_T, and we assume that

$$p(\mathbf{Y}_T|\mathbf{X}_T) = \prod_{t \in T} p(\mathbf{Y}_t|\mathbf{X}_t) \tag{1.1}$$

and furthermore we assume that the set of possible values of $\mathbf{X}_t$ and $\mathbf{Y}_t$ are $\mathbf{A}$ and $\mathbf{B}$ respectively, where $\mathbf{A}$ and $\mathbf{B}$ are finite sets, the matrix $U_t = p(\mathbf{Y}_t|\mathbf{X}_t)$ is assumed to be square and invertible.

For simplicity we also assume that the matrix U and the local conditional distributions of $\mathbf{X}_T$ are known; in practice this will of course usually not be the case, and we will then need to estimate parameters on the basis of observing $\mathbf{Y}_T$, see Andersen (1988a, 1989), for further details. Adopting an empirical Bayes approach the prior distribution for $\mathbf{X}_T$ is that of the Markov random field, inference for x_T should thus be carried out using the posterior distribution $\mathbf{X}_T|\mathbf{Y}_T$, and estimation of parameters for the prior should be based on the marginal distribution $p(\mathbf{Y}_T)$ of the observed image. Knowing the matrix U means that the device through which we receive the signal is known.

The image models described above are usually chosen for their simplicity rather than through realistic considerations of mechanisms generating the images. For more structured models an alternative was proposed by Grenander (1981) that allows for locally highly specialized models.

2. Resampling methods for Markov random fields

Given a Markov random field the local conditional distributions are on the form

$$p(x_t|x_{H_t}) = \frac{1}{z(x_{H_t})} \exp(\sum_{c \in C_t} \nu_c(x_t; x_{H_t})), \tag{2.1}$$

where C_t is the set of cliques with respect to the neighborhood system $\{H_t : t \in T\}$, the functions $\nu_c : \mathbf{A} \times A^{H_t} \to R$ are often called the clique potentials, a term taken from statistical mechanics, where Markov random fields are known as Gibbs fields. $z(x_{H_t})$ is the normalizing constant making the left hand side of (2.1) a genuine probability distribution. One of the many niceties of Markov random fields is that replicates from the distribution

$$p(x_T) = \frac{1}{z} \exp(\sum_{c \in \mathbf{C}_t} \nu_c(x_t; x_{H_t})) \tag{2.2}$$

usually can be simulated using Metropolis algorithm:

1. Choose a starting value $x_T^{(0)}$, from the set of configurations $\mathbf{A}^T$.
2. Iteratively with n, pick a pixel t at random (such that all pixels are visited infinitely often) and let $x_{T-\{t\}}^{(n)} = x_{T-\{t\}}^{(n-1)}$ and generate $x_t^{(n)}$ using (2.1).

It can be proved, by a Markov chain argument, that $\mathbf{X}_t^{(n)}$ is a recurrent Markov chain with state space $\mathbf{A}^T$. The stable distribution of $\mathbf{X}_T^{(n)}$, (as $n \to \infty$), exists and is the Markov random field with local conditional distributions (2.1).

2.1 Local conditional distributions

As we argued in section 1 the relevant distribution for restoration of the observed image y_T is that of $\mathbf{X}_T|\mathbf{Y}_T$, denoted here for convenience by $p_{y_T}(x_T)$; formally this may be written as

$$p_{y_T}(x_T) = \frac{p(y_T|x_T)p(x_T)}{p(y_T)}, \tag{2.3}$$

where by the local conditional distributions of $\mathbf{X}_T$ conditional on $\mathbf{Y}_T$ are given by

$$p_{y_T}(x_t|x_{H_t}) = \frac{p(y_t|x_t)p(x_t|x_{H_t})}{\sum_{x'_t \in \mathbf{A}} p(y_t|x'_t)p(x'_t|x_{H_t})} \tag{2.4}$$

That is conditionally on $\mathbf{Y}_T$ the true signal is a Markov random field with non stationary local conditional distributions given by (2.4). For models with spatially dependent noise, $p(y_t|x_T, y_{T-\{t\}} = p(y_t|x_{H_t}, y_{H_t}$, say, the posterior distribution is still a Markov random field, only (2.4) involves more complicated components.

2.2 Local optimality of resampling restoration schemes

Contrary to the restoration algorithms of simulated annealing and iterative conditional modes the resampling restoration schemes typically will not produce a global or local maximum posterior likelihood estimate. The optimality results we are to state and prove here will be of definite subjective nature, such as minimum mean squared error, minimum misclassification restoration etc. This, however, is also the key to the pixel to pixel variation measure since we acquire more information about the distribution of $\mathbf{X}_T|\mathbf{Y}_T$ than just a point estimate of the mode. By introduction of objective functions and coupling of pixels we are in a position where we can make minimum cost decisions, and for reasonably small blocks of pixels, B, say the full distribution of $\mathbf{X}_B|\mathbf{Y}_T$ can be well approximated by simple Monte Carlo.

Definition 2.1. A structure identifier, F, say is a function from $\mathbf{A}^T(=\underset{t\in T}{\times}\mathbf{A}_t) \to \mathbf{M}$, where $\mathbf{M}$ is a metric space. □

The structure identifiers will play the role of statistics of interest in classical statistics, such as sample averages, mean squared sample error, parameter estimators, shape and size characteristics etc. Given a structure identifier, F, the resampling of images form (2.3) facilitates approximation of the conditional distribution of $F(\mathbf{X}_T)|\mathbf{Y}_T$, that is the a posterior distribution usually used for estimation and forecasting. Below we define a special class of structure identifiers: The local objective functions.

Definition 2.2. A local objective function is a real valued additive structure identifier, F, depending on x_T, such that

$$F_{x_T}(x'_T) = \sum_{t\in T} f_{t,x_t}(x'_t)$$

and

$$f_{t,x}(x) = 0, f_{t,x}(y) \geq 0, \text{ for all } x, y \in \mathbf{A}, \text{ and } t \in T. \qquad \square$$

With this definition in hand we are able to state and prove theorems of optimality for the resampling estimation schemes.

Theorem 2.1. *Let $x_1^*, \ldots, x_n^*$ be resampling replicates from (2.3) using Metropolis algorithm as described above given a local objective function, defined by $f : \mathbf{A}^2 \to R$ there exist functions $\phi^{(n)} : (\mathbf{A}^T)^n \to \mathbf{A}^T$ such that*

$$E\{F_{\mathbf{X}_T}(\phi^{(n)}(x_1^*, \ldots, x_n^*))\} \to \min_{\psi} \mathbf{E}\{F_{\mathbf{X}_T}(\psi(\mathbf{Y}_T))\} \text{ as } n \to \infty$$

where the minimum is taken over all functions $\psi : \mathbf{A}^T \to \mathbf{A}^T$, and the expectation is over $(\mathbf{X}_T, \mathbf{Y}_T)$.

Proof. Note that

$$\begin{aligned}\mathbf{E}\{F_{\mathbf{X}_T}(\phi^{(n)}(x_1^*, \ldots, x_n^*))\} &= \mathbf{E}\{\mathbf{E}\{F_{\mathbf{X}_T}(\phi^{(n)}(x_1^*, \ldots, x_n^*))|\mathbf{Y}_T\}\} \\ &= \mathbf{E}\{\sum_{t\in T} \mathbf{E}\{f_{t,\mathbf{X}_t}(\phi^{(n)}(x_1^*, \ldots, x_n^*))|\mathbf{Y}_T\}\}\end{aligned}$$

where by it suffices to prove that

$$\mathbf{E}\{f_{t,\mathbf{X}_t}(\phi_t^{(n)}(x_1^*,\dots,x_n^*))|\mathbf{Y}_T\} \to \min_{\psi_t} \mathbf{E}\{f_{t,\mathbf{X}_t}(\psi_t(y_T))|\mathbf{Y}_T\}, \text{ as } n \to \infty.$$

The right hand side is calculated by

$$\sum_{x\in\mathbf{A}} f_{tx}(\widehat{y})p(x;\mathbf{X}_t|\mathbf{Y}_T). \tag{2.5}$$

where $\widehat{y}$ minimizes (2.5). Since $\widehat{p}_t$, the bootstrap empirical distribution function for $\mathbf{X}_t$ converges to $p(.,\mathbf{X}_t|\mathbf{Y}_T)$ with probability one $\widehat{\widehat{y}}$ minimizing

$$\sum_{x\in\mathbf{A}} f_{tx}(\widehat{\widehat{y}})\widehat{p}_t(x)$$

converges with probability one to $\widehat{y}$, where by the theorem follows, since $\mathbf{A}$ and $\mathbf{B}$ are finite.

To appreciate the above theorem fully we turn to a few examples; let $f_{t,x}(y) = 1_{\{x\neq y\}}$, then the majority vote estimate (among the resampling replicates) converges with probability one to the minimum mean misclassification estimate, as the number of iterations per resampling replicate, and the number of bootstrap samples tend to infinity. In the case where $\mathbf{A}$ and $\mathbf{B}$ are not finite, but complete metric spaces the theorem still holds, added further assumptions for f and the local conditional distributions.

Corollary 2.1. *Assuming that the Markov chain defined by Metropolis algorithm is ergodic, i.e. that it has a unique stable distribution independent of* $\mathbf{X}_T^{(0)}$. *Let* $f_t : \mathbf{A}^2 \to R$ *be an objective function, such that* f *is convex and everywhere differentiable, then there exist functions* $\phi^{(n)} : (\mathbf{A}^T)^n \to \mathbf{A}^T$ *such that*

$$\mathbf{E}\{F_{\mathbf{X}_T}(\phi^{(n)}(x_1^*,\dots,x_n^*))\} \to \min_{\psi} \mathbf{E}\{F_{\mathbf{X}_T}(\psi(\mathbf{Y}_T))\}, \text{ as } n \to \infty,$$

where the minimum is taken over functions $\psi : \mathbf{A}^T \to \mathbf{A}^T$, *and the mean is over* $(\mathbf{X}_T, \mathbf{Y}_T)$.

Proof. As above, only minimization of (2.5) is performed through differentiation. As to conditions sufficient for the existence and uniqueness of the stable distribution for the simulation Markov chain see Andersen (1989) and Revuz (1975). □

As a special case of corollary 2.1 we get a resampling approximation to the classical two-filter and Kalman filter for times series models that are special cases of Markov random fields on one dimensional lattices (Solo (1982) and Meinhold and Singpurvalla (1983)), along with a natural generalization of these methods to no Gaussian cases and spatial Gaussian and non Gaussian Markov random fields, see Andersen (1989), Ripley (1981), Besag (1974, 1975) and Green, Jennison and Seheult (1985). For Gaussian Markov random field models the mean squared error optimal restoration is identical to the maximum posterior mode restoration.

Assume for instance that after obtaining the restored image special interest is focused on a region, B, say by resampling again an optimal estimate can be found for x_B or any structure identifier thereof, with respect to local objective functions treating B as one single pixel; this can be carried out either marginally or conditionally on $\mathbf{X}_{T-B} = \widehat{x}_{T-B}$, sampling from either $\mathbf{X}_B|\mathbf{Y}_T$ or $\mathbf{X}_T|\mathbf{Y}_T$ fixing $\mathbf{X}_{T-B} = \widehat{x}_{T-B}$, respectively. As B, the local region of interest grows larger, however, this becomes increasingly computer time consuming for some choices of structure identifier.

Proposition 2.1. *The local asymptotic measure of performance of the restoration:*

$$\lim_{n\to\infty} \mathbf{E}\{f_{t,x_t}(\widehat{x}_t)|\mathbf{Y}_T\}$$

is consistently estimated, but in general not unbiasedly, by

$$\sum_{x\in\mathbf{A}} f_{t,x}(\widehat{x}_t)\widehat{p}_t(x),$$

where $\widehat{p}_t(x)$ is the observed frequency of the event $\mathbf{X}_t = x$, among the resampling replicates. □

In practise we need to choose the number of samples from $\mathbf{X}_T|\mathbf{Y}_T = y_T$, and in doing this we also set the internal precision of the stochastic approximation by Monte Carlo. In general it is our belief that fewer samples of higher quality are preferable to more samples of lesser quality, which in practice amounts to letting the local updating procedure from Metropolis algorithm run longer and taking fewer samples. Using these procedures, however, for simple structure identifiers does not require excessive computer time.

Example 2.1. For restoration purposes local smoothing algorithms are known to work very well for $\mathbf{A}$ categorical although local smoothing becomes less useful. We have chosen the local conditional distributions such that

$$\log\left\{\frac{p(x_t = i|x_{H_t})}{p(x_t = j|x_{H_t})}\right\} = \alpha_1(n_1(i) - n_1(j)) + \alpha_2(n_2(i) - n_2(j)), \tag{2.6}$$

where $n_1(i), i = 1,2,3$ is the number of pixels among the first four nearest neighbors with pixel value "i", $n_2(i)$ correspondingly is the number of pixels among the four second nearest neighbors with pixel values "i". The error process is described by

$$p(y_t|x_t) = \begin{cases} 1-p & \text{if } y_t = x_t \\ p & \text{if } y_t \neq x_t \end{cases}.$$

Figure 2.1 (a), the original image was generated from (2.6) using 5×10^6 local updates, and $\alpha_1 = 1.05, \alpha_2 = 0.90$. Since T is wrapped around a torus, we have in fact only three major connected patches. (b) is the corrupted image, $p = 0.35$, (c), the minimum misclassification restoration using 200 resampling replicates and 5×10^4 local updates per replicate shows 150 misclassifications, (d). Figure (e) shows this a local variability measure, again picking out the interior boundaries very successfully

- (f) shows this "estimated" interior boundary process overlaid the restored image. □

3. Stochastic Inference

The practical inference problems of spatial statistics are many-sided and in image analysis often highly context dependent–there are, however, basic problems in formalized hypothesis testing since the hypothesis are not easily described in terms of the Markov random fields models.

3.1 Examples

Example 3.1

Consider figure 3.1 below, (a) showing the original image, generated by the same Markov random field as in example 2.2, super imposed on (a), in the smaller frame is a black object drawn by hand. Figure 3.1 (b) shows the corrupted image: The original image received through a symmetric noisy channel with 35 percent misclassified pixels. In (c) is the restored image using the minimum misclassification estimation with 200 resampling replicates and 50,000 local updates per replicate; Lets assume that we were interested in the posterior distribution of the area of B, the black object in the smaller frame. In figures (d), (e) and (f) we have drawn up the approximated conditional distributions of $F_1|\mathbf{Y}_T, F_2|\mathbf{Y}_T$ and $F_3|\mathbf{Y}_T$, respectively, where F_1 is the area of the largest connected set of black pixels intersecting B, and F_2 is the largest area of the smallest subsets of T intersecting B and surrounded by a connected subset of non black pixels; that is F_2 is the area of the largest connected set of black pixels intersecting B, and its interior; F_3 is the naive estimate, counting the number of black pixels in the smaller frame–from the plots in (d), (e) and (f) it seems that F_1 and F_2 and F_3 are all biased toward larger volumes then the true of 163, by repetition of this experiment from the same original image it seems that F_1 and F_2 are biased toward smaller volumes while F_3 seems biased toward larger volumes; the biases of course depend on the actual realization of the error process. The true volume of B is 163, and the volume of the estimated object $\widehat{B}$ is 174, different from $EF_1|\mathbf{Y}_T, EF_2|\mathbf{Y}_T$ and $EF_3|\mathbf{Y}_T$. □

3.2 Optimality of decision rules

Having chosen the simple Markov random field image models as the general frame work we must accept that all configurations over T are possible, the model parameters can not be used to confine the true image to have a specific structure. For pattern recognition and classification purposes this characteristic makes the introduction of structure identifiers almost impossible to avoid; in order to make decisions on the basis of images, using the Markov random field models, we must be able to classify any given configuration on $\mathbf{A}^T$. As usual we also require the knowledge of a cost function.

With the resampling methods in hand we can, however, given a structure identifier and a cost function prove optimality in terms of minimum loss.

Theorem 3.1. *Let $F : \mathbf{A}^T \to \{1, \ldots, k\}$ be a structure identifier, and let W be a cost function such that W is invertible, $W_{i,i} = 0$ and $W_{i,j} \geq 0$, then the classification rule determined by the minimum cost identifier of $F(x_T^{(i)}), i = 1, \ldots, n$, with respect to $\widehat{p}(i|\mathbf{Y}_T)$, $i = 1, \ldots, k$ is the minimum cost classifier of x_T into*

$$\{F^{-1}(1), \ldots, F^{-1}(k)\}$$

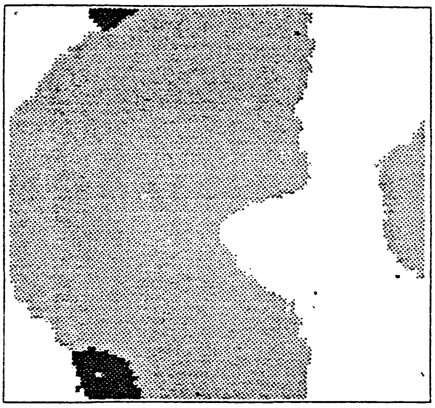

Figure 2.2 a), Orginal image 100 × 100 pixels, realization of Markov random field after 5×10^6 local updates. parameters are $\alpha_1 = 1.05$, $\alpha_2 = 0.9$

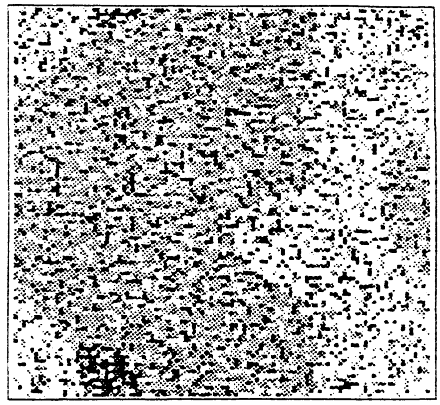

Figure 2.2 b), Observed image 35 percent miclassification rate, symmetric noisy channel. Actually observed misclassified pixels 36.5 percent.

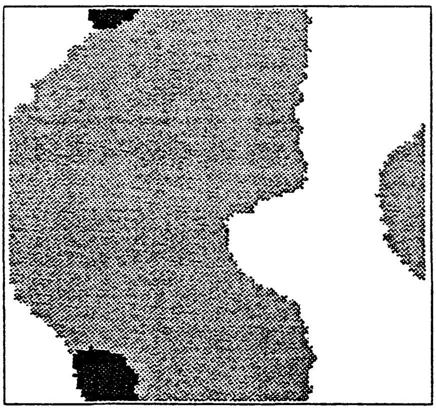

Figure 2.2 c), Minimum misclassification restoration, on the basis of 200 replicate samples and 50000 local updates per sample. Misclassification rate 1.50 percent, estimated misclassification rate 2.92 percent.

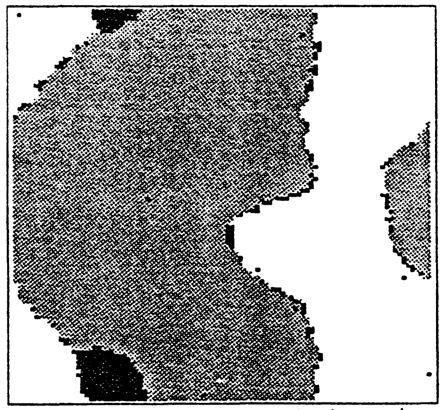

Figure 2.2 d), The 150 errors overlayed on the restored image, errors are black.

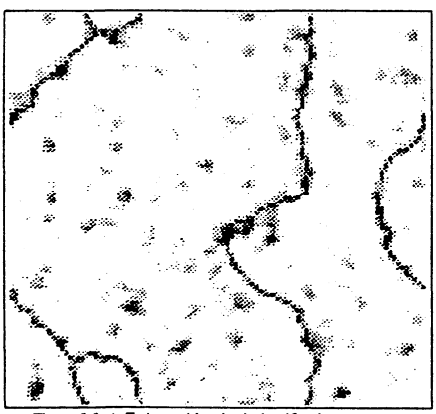

Figure 2.2 e), Estimated local misclassification measure from 200 independent replicates. $\hat{p}_t = p(x_t = \hat{x}_t \mid Y_T)$, image shows $\hat{p}_t\,(1-\hat{p}_t)$.

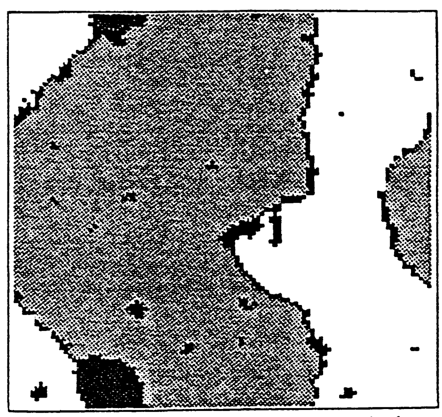

Figure 2.2 f), The highest 5 percent estimated local misclassification rates overlayed on the restored image.

Figure 2.2

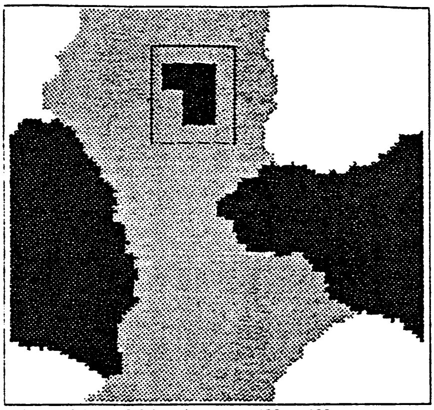

Figure 3.1 a), Original image on 100 × 100 square lattice, the image consists of a realization from the Markov random field in example 2.2, and a black object super imposed in the smaller frame.

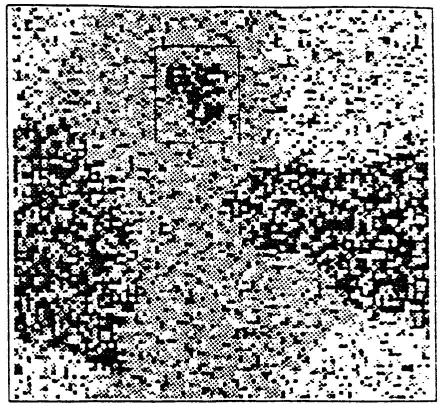

Figure 3.1 b), Corrupted image 35 per cent misclassified pixels.

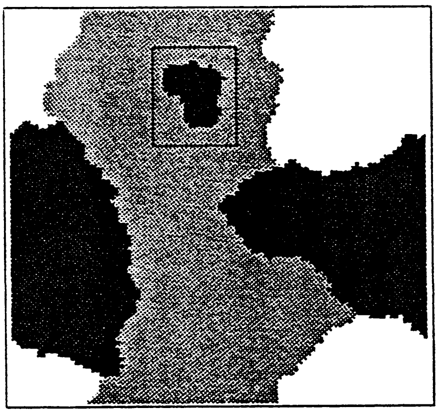

Figure 3.1 c), the minimum misclassification resoration, using the algorithm, and the parameters from example 2.2. 200 bootstrap samples with 50000 local updates per sample. The super imposed object shows up clear in the smaller frame. The number of misclassified pixels is 187.

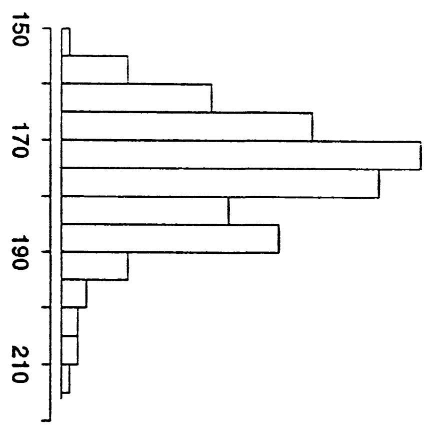

Figure 3.1 d), emperical distribution function of F_1, the mean value is 175.4.

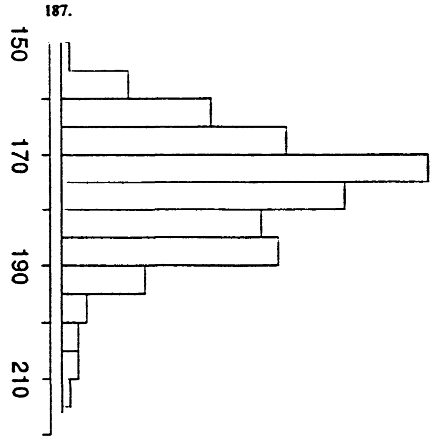

Figure 3.1 e), emperical distribution function of F_2, the mean value is 175.8.

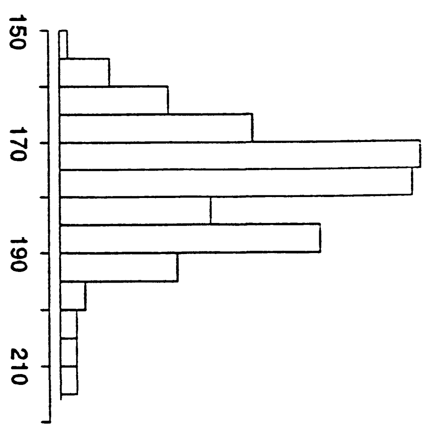

Figure 3.1 f), emperical distribution function of F_3, the mean value is 177.3.

Figure 3.1

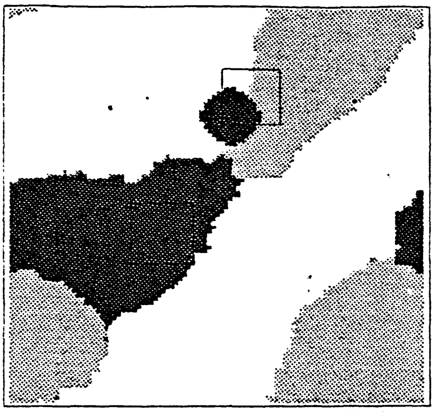

Figure 3.2 a), Realization of Markov random field, as in example 2.2; the black ball in the upper half of the image was super imposed artificially.

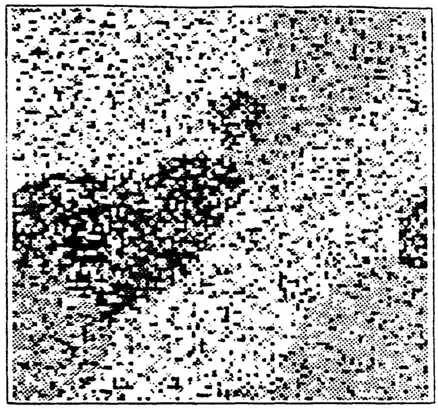

Figure 3.2 b), 35 percent corruption of a), symmetric noisy channel.

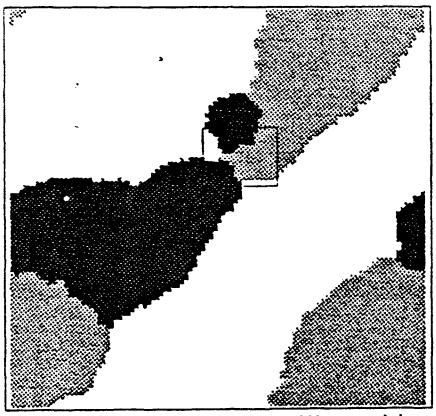

Figure 3.2 c), Restoration using 200 resampled replicates from Markov random fields model; example 2.2.

Figure 3.2

Proof. The proof is trivial since we can approximate the distribution function to $F(\mathbf{X}_T)|\mathbf{Y}_T$ arbitrarily well by theorem 2.1. □

Example 3.3. Consider figure 3.2, in (a) is a realization of the Markov random field from example 2.2–the ball intersecting the top frame was hand drawn and super imposed such that the two black patches are separated by only two pixels; (b) shows the corrupted image and (b) the minimum misclassification restoration of (a), misclassification rate: 1.96 percent; we are interested in finding a minimum misclassification rule to answer the question: Is the black patch intersecting the frame super imposed on figure 3.2 (c) one or two objects? Let the misclassification costs be symmetric, we then need a function $F\{1,2,3\}^T \to \{0,1\}$, by which we can classify any given image. We choose the following simple suggestion:

$F = 1$ if the largest set of connected black pixels intersecting the top frame and the largest set of connected black pixels intersecting the lower frame of figure 3.2 (a) is the same.

$F = 0$ otherwise.

In 135 of the 200 replicates $F = 0$ was observed. The procedure, thus, correctly classified the observed image in figure 3.2 (b).

Figure 3.3 shows the minimum misclassification restoration given that the black "blobs" are figure 3.3 (a) is really two objects, misclassification rate: 1.95 percent. Note again that since the Markov random field models can not force global features for the restoration there is no guarantee that $F(\widehat{x}_T) = 0$. In figure 3.3 (b) we have displayed the pixels of disagreement between figures 3.2 (c) and 3.3 (a) as black pixels overlaid on figure 3.3 (a), there is a total of 13 black pixels in figure 3.3 (b). □

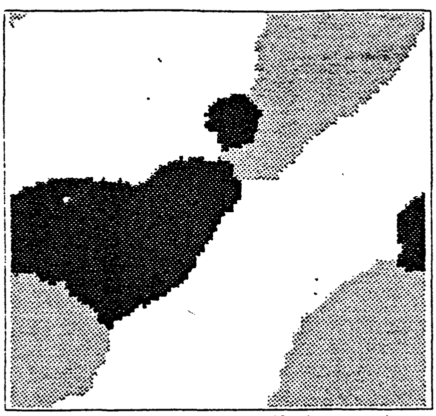

Figure 3.3 a). minimum misclassification restoration conditional on $F(X_T)=0$, restoration on the basis of 135 replicates.

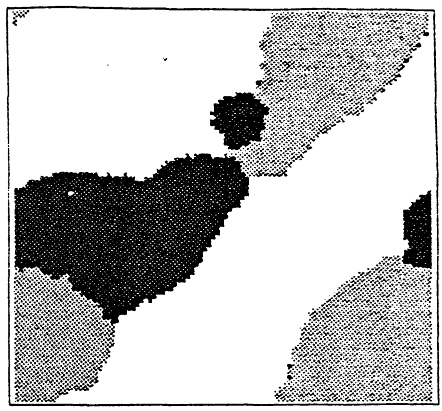

Figure 3.3 b). disagreements between figures 3.2 c) and 3.3 a) overlayed as black pixels on figure 3.3 a).

Figure 3.3

Acknowledgements

The author is grateful to The Danish Natural Science Research Council for financial support during his stay at Department of Statistics, University of Washington. Inspiration and helpful discussions with students, faculty and guests at University of Washington are also gratefully appreciated.

References

Andersen, L. S. (1988a). *Consistent Parameter Estimation for Corrupted Markov Random Fields: With Applications to Image Analysis.* Technical Report no.117, Dept. of Statistics, University of Washington.

Andersen, L. S. (1988b). *Stochastic Inference for Spatial Statistics.* Technical Report no.120, Dept. of Statistics, University of Washington.

Andersen, L. S. (1989). *Stochastic Inference in Spatial Statistics.* Ph.D. Thesis, University of Washington.

Besag, J. (1974). Spatial Interaction and The Statistical Analysis of Lattice Systems. *Journal of Royal Statistical Society* B **34**, 75–83.

Besag, J. (1975). Spatial Analysis of Non Lattice Data. *The Statistician* **24**, 179–195.

Besag, J. (1986). On The Statistical Analysis of Dirty Pictures. *Journal of Royal Statistical Society* B **47**, (with discussion).

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and The Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geman, S., & Graffigne, C. (1986). *Markov Random Field Image Models and Their Applications to Computer Vision.* Proceedings of the International Congress of Mathematicians, Berkeley, California.

Green, P., Jennison, C., & Seheult, A. (1985). Analysis of Field Experiments by Least Squares Smoothing. *Journal of Royal Statistical Society* B **47**, 299–315.

Grenander, U. (1981). *Regular Structures.* Lectures in pattern theory, Volume 3. Springer-Verlag, New York.

Haslett, J., & Horgan, G. (1986). *Linear Models in Spatial Discriminant Analysis.* (to appear).

Haslett, J., & Horgan, G. (1987). Spatial Discriminant Analysis–A Linear Discriminant Function for The Black/While Case, Devijver, P. A. & Kittler, J. (1987, eds.), *Pattern Recognition Theory and Applications*, NATO-ASI series F, Springer-Verlag, 47–55.

Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic Solution of Ill–Posed Problems in Computer Vision. *Journal of American Statistical Association* **82**, 76–89.

Meihhold, R. J., & Singpurwalla, N. D. (1983). Understanding The Kalman Filter. *American Statistician* **43**, 123–127.

Revuz, D. (1975). *Markov Chains.* Volume II. North-Holland, Amsterdam.

Ripley, B. D. (1981). *Spatial Statistics.* Wiley and Sons, New York.

Ripley, B. D. (1987). *Stochastic Simulations.* Wiley and Sons, New York.

Solo, V. (1982). Smoothing Estimation of Stochastic Processes: Two Filter Formulas. *IEEE Transactions on Automation Control* AC **27**, 473–476.