Edge Detection, Spatial Smoothing, and Image Reconstruction With Partially Observed Multivariate Data

Sarat C. Dass and Vijayan N. Nair

Situations with incomplete multivariate spatial data on a lattice are considered. The goal is to impute the missing data in the presence of edges or boundaries and recover the image. Two methods based on Bayesian hierarchical models that iterate between edge detection and spatial smoothing to impute the missing data within identified homogeneous regions are examined. Their performance is compared with another method that imputes the missing values using edge-preserving spatial smoothers with locally varying weights. The performances of the three methods are compared on artificial and real datasets. It is seen that information from the multivariate data is critical in recovering the images. An application with color images where only one of three primary colors (red, green, or blue) is observed at each pixel is used to illustrate the results.

KEY WORDS: Edge processes; EM algorithm; Imputation; Kriging; Markov random field; Missing data; Spatial mixture model.

1. INTRODUCTION

Consider the following application that arises in the recovery of color images. A digital camera of resolution $m \times n$ is mounted with a single charged couple device (CCD) at each pixel location. A single CCD records the intensity of only one primary color: red, green, or blue (R, G, B). The goal is to impute the missing values of the (R, G, B) color intensities so that we can reconstruct the digital image. Similar problems also arise in other applications, such as flat panel displays.

Various designs have been used to determine the choice of color to be recorded at each pixel in a rectangular lattice (see Holst 1998, p. 180). One design that is widely used for sampling colors (sometimes called the Bayer design) is shown in Figure 1. In this scheme, G is sampled twice as often as R and B, because variation in the intensity of G is more sensitive to the eye compared that of to R and B. Thus it is necessary to get more accurate readings for G. Partially sampled images that are obtained from digital cameras using this design are referred to as *Bayer images*.

One interpolation scheme for the missing colors takes a simple average of the observed colors at k nearest neighbors of the pixel of interest. Often k is chosen to be 4, made up of the east–west–north–south neighbors. We refer to this as *Bayer interpolation*. Figure 2 shows the result of this smoothing applied to two simple images: panels (a) and (c) are the true images, and panels (b) and (d) are the corresponding recovered images using Bayer interpolation. It is clear from this figure that problems arise from smoothing across the boundaries and can result in blurred reconstructions. Thus there is a need to incorporate edge or boundary information while imputing the missing data. We consider two different approaches based on the incomplete data: iterative methods for edge detection and spatial smoothing/interpolation, and edge-preserving smoothers with locally varying weights.

Techniques for edge detection and image reconstruction have been discussed extensively in the literature. What makes the present problem of special interest is the presence of incomplete multivariate data. As we discuss, borrowing information from the multivariate data from neighboring pixels is critical in determining the edges and imputing the missing values.

We begin with a general formulation of the problem. Let S denote the set of all pixels on a $m \times n$ rectangular lattice. For every $s \in S$, let x_s denote the p-variate vector of attributes of interest at that site. In the color image application, this will be the true color intensity (R, G, B). Denote the collection of all attributes by $X = \{x_s, s \in S\}$, and denote the observable data by $Y = \{y_s, s \in S\}$. This represents the partially observed version of X with possible degradation or noise in the form of iid normal errors with mean 0 and known variance σ^2. For example, in the Bayer design, there is a measurement at every pixel, but only one component of the trivariate vector (R, G, B) is available. In other applications, all of the p-dimensional attributes $(x_1, \ldots, x_p)$ may be unobservable at some pixels. It is possible that the attributes of interest, X, are measured without error whenever the measurements are available. This case can be handled by taking $\sigma^2 \equiv 0$ in the foregoing formulation.

We use the following notation throughout. For each site, s, let obs_s denote the component of colors observed at site s, and let mis_s denote the components that were unobserved at that site. In the color image application, $x_s = (x_{1s}, x_{2s}, x_{3s}) = (R_s, G_s, B_s)$. So if R is observed at pixel s, then $obs_s = \{1\}$ and $mis_s = \{2, 3\}$. In other applications, if all of the component are missing at a site s, then obs_s will be empty. Let $obs = \bigcup_{s \in S} obs_s$ and $mis = \bigcup_{s \in S} mis_s$. Let $y_{obs} = \{y_i, i \in obs_s, s \in S\}$ denote the vector of all observed colors, and let $y_{mis} = \{y_i, i \in mis_s, s \in S\}$ denote the missing components.

Sarat C. Dass is ———, Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824 (E-mail: *sdass@msu.edu*). Vijayan N. Nair is ———, Department of Statistics, Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor, MI 48109-1285 (E-mail: *vnn@umich.edu*). The authors thank Dr. C. K. Lakshminarayan of Motorola for initial discussions that led to the formulation of this problem and for providing the CCD images. They are also grateful to reviewers for insightful comments that led to improvements in the manuscript. This research began while the first author was visiting the Department of Statistics, University of Michigan. The research was supported in part by National Science Foundation grant DMS-9803281.

Journal of the American Statistical Association
March 2002, Vol. 00, No. 0, Theory and Methods
DOI 10.1198/00

G	R	G	R	G	⋯
B	G	B	G	B	⋯
G	R	G	R	G	⋯
B	G	B	G	B	⋯
⋮	⋮	⋮	⋮	⋮	⋱

Figure 1. The Bayer Design.

Hierarchical models such as the one described earlier have been used in many image restoration and segmentation problems (see, e.g., Geman and Geman 1984; Derin and Elliot 1987; Geman and McClure 1987; Jeng and Woods 1991). However, the performance of these models has been investigated only in situations with no missing observations. Our goal here is to reconstruct the multivariate data X from the (incomplete) observed data Y. We can handle situations in which we just want to recover some function of X, say $Z = f(X)$ as a special case, by using $\widehat{Z} = f(\widehat{X})$. Alternative approaches that use information about the structure of $f(\cdot)$ explicitly are not considered here.

If the image is homogeneous, then spatial interpolation or smoothing techniques (kriging, cokriging, or related techniques) can be used to impute the missing data. These are discussed at some length in Section 2, because they form the building blocks for later sections. However, when natural edges are present in an image, interpolating across these boundaries will cause blurring in the reconstructed image, as seen in Figure 2. Thus it is critical that information about the boundaries be incorporated into the imputation process.

The rest of the article is organized as follows. Sections 3 and 4 impute the missing values iteratively by first doing edge detection and then spatial smoothing/interpolation within the identified homogeneous regions. Both are based on Bayesian hierarchical methods; Section 3 uses a Markov random field (MRF) to model the edge processes, whereas Section 4 is based on underlying class processes. Instead of explicitly doing edge detection, Section 5 considers edge-preserving imputation via spatial smoothing with locally varying weights. The idea can also be motivated as hierarchical Bayesian modeling through the use of locally weighted pairwise difference priors. Section 6 compares the results of various interpolation schemes on simulated data and on a real CCD image.

2. SPATIAL SMOOTHING AND IMPUTATION IN HOMOGENEOUS REGIONS

We first consider the case with a homogeneous region. We make the following model assumptions throughout the article: the underlying multivariate spatial process X is a Gaussian Markov random field (GMRF) and, given $X = x$, the observed data Y_{obs} are iid $N(x_{obs}, \sigma^2)$.

We begin with a brief introduction to MRFs. There is an extensive literature on the subject (see, e.g., Mardia 1988; Chellapa and Jain 1991; Guyon 1995; Winkler 1995). Given a set of sites S, let $\mathcal{N} = \{N_s : s \in N_s\}$ be a collection of neighborhoods for all of the sites $s \in S$. For any pair of sites s and t, the neighborhood system is assumed to have the properties $s \notin N_s$ and $s \in N_t$ if and only if $t \in N_s$. For a subset C of sites in S, we call C a *clique* if all pairs of sites in C are neighbors. Thus the neighborhood system as defined earlier will generate a collection of cliques, $\mathcal{C}$. Specifying a collection of cliques or a neighborhood system on S for a MRF is equivalent. When $S = \{\text{all edge sites}\}$, we specify a MRF for the edge process via a collection of cliques, whereas when $S = \{\text{all pixels in the image}\}$, we model the color intensity distribution on S via a neighborhood system.

As noted earlier, we consider Gaussian models for the data. In the digital camera application, the color intensity process (R, G, B) is modeled as a trivariate GMRF. In the sequel, we will continue to use the term "colors" to also refer to the general situation with arbitrary attributes. We include here a brief description of general p-variate GMRF models (see Mardia 1988 for a more extensive discussion).

Let $\mathcal{X} = \{x_s : s \in S\}$ be a collection of multivariate random variables on S. The total number of sites in S is N, say. Denote by $x_s = (x_{1s}, x_{2s}, \ldots, x_{ps})^T$ the vector of attributes of interest at site s. Also, let x_{-s} denote the attributes at all pixels in S excluding pixel s. For a given neighborhood system $\mathcal{N}$ and for $\mu \in R^p$, assume that the conditional distribution of x_s given x_{-s} is

$$\pi(x_s \mid x_{-s}) = N_p\left(\mu + \phi \sum_{t \in N_s} B_{st}(x_t - \mu), G_s\right), \qquad (1)$$

where $0 < \phi \leq 1$, B_{st} is a $p \times p$ matrix of weights, and G_s is a $p \times p$ positive definite matrix. The mean term μ in (1) is the underlying common mean for the homogeneous region. It

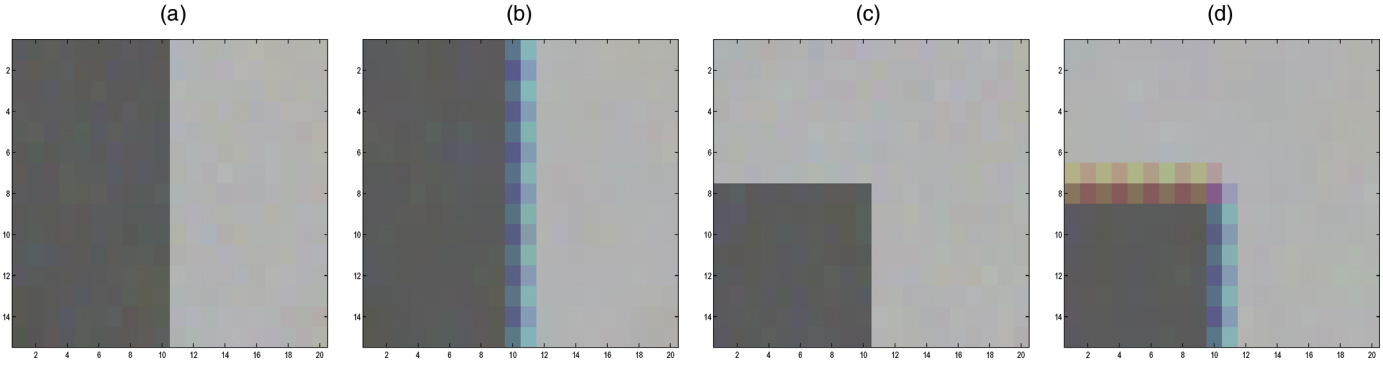

Figure 2. Blurring of Edges Using Bayer Interpolation Scheme. (a) Original image; (b) Bayer reconstruction; (c) original image; (d) Bayer reconstruction.

is well known that the foregoing conditionals may not define a consistent joint distribution for $\mathcal{X}$ on S. However, it can be shown (Guyon 1995) that the joint distribution is uniquely defined and is multivariate Gaussian under the consistency condition

$$G_s^{-1}B_{st} = (G_t^{-1}B_{ts})^T, \tag{2}$$

for $t \neq s$. For $t = s$, define $B_{ss} = -I_{p\times p}$. Also, define $M =$ block diagonal$((G_s^{-1}))$ and $B = ((B_{st}))$. Under (2), the joint distribution of $X = (x_1^T, x_2^T, \ldots, x_N^T)^T$ is given by

$$X \sim N(\nu, \Sigma), \tag{3}$$

where $\nu = \underline{1}_N \otimes \mu$ and $\Sigma^{-1} = -M \cdot B$.

In the hierarchical specification, we assume that the observed colors, y_{obs}, are recorded with error with a known noise variance σ^2. This parameter σ^2 can also be viewed as a tuning parameter that controls the degree of smoothing for reconstructing the underlying color process X. As $\sigma^2 \to 0$, smoothing tends to an interpolation scheme for the missing colors, x_{mis}. As shown later, this leads to precisely the cokriging equations for x_{mis} given y_{obs}.

To write the expressions for the spatial predictor of the missing values explicitly, reorder the components of X as $X^* = (x_{obs}, x_{mis})^T$. Then

$$X^* \sim N(\nu^*, \Sigma^*), \tag{4}$$

where ν^* and Σ^* are obtained from the ν and Σ corresponding to the reordering of the rows of X in (3). We decompose ν^* and Σ^* into the observed and missing parts, $\nu^* = (\nu^*_{obs}, \nu^*_{mis})^T$ and

$$\Sigma^* = \begin{pmatrix} \Sigma_{obs,\,obs} & \Sigma_{obs,\,mis} \\ \Sigma_{mis,\,obs} & \Sigma_{mis,\,mis} \end{pmatrix}, \tag{5}$$

with $\Sigma_{mis,\,obs} = \Sigma^T_{obs,\,mis}$. Writing $\underline{y} = (y_{obs}, y_{mis})^T$, the joint distribution of $(y_{obs},\ x_{mis})^T$ is given by

$$(y_{obs},\ x_{mis})^T \sim N(\nu^*, \Sigma^{**}) \tag{6}$$

with

$$\Sigma^{**} = \begin{pmatrix} \Sigma_{obs,\,obs} + \sigma^2 I & \Sigma_{obs,\,mis} \\ \Sigma_{mis,\,obs} & \Sigma_{mis,\,mis} \end{pmatrix}. \tag{7}$$

The minimum variance predictor of x_{mis} given y_{obs} is the expectation of the conditional distribution of x_{mis} given y_{obs} as before. From (6), we get

$$E(x_{mis} \mid y_{obs}) = \nu^*_{mis} + \Sigma_{mis,\,obs}(\Sigma_{obs,\,obs} + \sigma^2 I)^{-1} \times (y_{obs} - \nu^*_{obs}). \tag{8}$$

Also, for the observed sites, x_{obs} is estimated using the minimum variance predictor

$$E(x_{obs} \mid y_{obs}) = \nu^*_{obs} + \Sigma_{obs,\,obs}(\Sigma_{obs,\,obs} + \sigma^2 I)^{-1} \times (y_{obs} - \nu^*_{obs}). \tag{9}$$

In (8) and (9), σ^2 controls the degree of smoothing for estimating x_{mis} and x_{obs}. The case $\sigma^2 = 0$ corresponds to spatial interpolation. Note that $E(x_{obs} \mid y_{obs}) = y_{obs}$ in (9) when $\sigma^2 = 0$. Similarly, for x_{mis}, (8) reduces to the well-known cokriging equations with known μ and Σ when $\sigma^2 = 0$. When μ is unknown, it can be estimated from the data, which results in the usual cokriging equations for unknown means. (See Ver Hoef and Cressie 1993 and Ver Hoef and Barry 1998 for details on cokriging and general multivariable spatial prediction for homogeneous regions, and Le and Zidek 1992 and Le, Sun, and Zidek 1997 for a fully Bayesian approach to kriging and cokriging.) We discuss some choices of Σ next.

Spatial smoothing/interpolation schemes such as cokriging that use information from the multivariate dependence structure are more efficient than marginal methods such as kriging. However, we need to estimate the multivariate correlation structure to use these schemes. For the color image application, only a single coordinate (color) is observed at each pixel, so there is no way to estimate the multivariate correlation structure without making further model assumptions. One reasonable class of models to consider is

$$\Sigma = U \otimes \Gamma, \tag{10}$$

where $U = ((u_{st}))$ depends only on the spatial locations (s, t) and $\Gamma = ((\gamma_{kl}))$ is the covariance between x_{ks} and x_{ls}, which is assumed to be independent of s. In other words, we model the covariance between attribute components at two different sites as a product of two terms, the first term measuring only the correlation between spatial locations and the second term measuring only the correlation between attributes (colors). This model allows estimation of the covariance under the Bayer design or other similar missing-data schemes. This kind of multispectral decomposition for GMRFs has been used in other image reconstruction problems (see, e.g., Guyon 1995 and references therein). Such decompositions of Σ also enable an enormous reduction in computational complexity.

There can be many choices for U, for example, $u_{st} = H(s - t)$ for some known positive definite function H (see Cressie 1993 for examples of H). For the color image application, the choice of U that we consider can be motivated through the conditional GMRF specification in (1). With $p = 3$, we take B_{st} and G_s in (1) as

$$B_{st} = \begin{pmatrix} \frac{1}{n_s} & 0 & 0 \\ 0 & \frac{1}{n_s} & 0 \\ 0 & 0 & \frac{1}{n_s} \end{pmatrix} \tag{11}$$

and

$$G_s = \Gamma / n_s, \tag{12}$$

where n_s is the number of neighboring sites of s. This simply means that we impute the missing values at site s as a weighted average of neighboring attributes and the mean attribute μ. Equation (12) implies that there is a common unknown variance covariance matrix for x_s for every $s \in S$. In

view of (11) and (12), we have $\Sigma = U \otimes \Gamma$ with $U^{-1} = ((v_{st}))$, where

$$v_{st} = \begin{cases} n_s & \text{if } t = s \\ -\phi & \text{if } t \in N_s \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

(provided that U^{-1} is invertible) with $\phi = 1$.

When $\phi = 1$, (1) says that each color is interpolated as the (unweighted) average color of its neighbors. When $\phi = 1$, the GMRF model is noninformative in the direction of the unknown mean. The posterior becomes informative if each component of (R, G, B) is observed at least once (possibly at different sites) in S (see Dass 2000). In our simulations discussed in Section 5, we report on image reconstructions for values of ϕ close to 1 ($\phi = .99, .95$), for which the prior GMRF model is proper. There were no visible differences in the reconstructions. For the simulations, we chose σ^2 to be a small positive value ($\sigma^2 = .01$).

The spatial smoothing/interpolation schemes discussed thus far do not account for edges or boundaries present in images. In the next three sections, we discuss methods that incorporate edge information and prevent interpolation across boundaries when imputing the missing data.

3. MODELING WITH AN UNDERLYING MRF EDGE PROCESS

In this section we consider methods based on a hierarchical MRF model for edge processes (see Geman and Geman 1984). Specifically, we assume that edge sites are located between each pair of pixels in the lattice, and they take value 0 if no edge is present and 1 if an edge is present. The presence or absence of an edge site is determined by whether there is a sudden change in intensity of the colors as we move between adjacent pixels. The cliques that we consider are the same as those of Geman and Geman (1984), which consists of the four edge sites corresponding to a group of four image pixels as shown in Figure 3.

The notation $e(s, t)$ stands for the edge between sites s and t. For each edge clique C, consisting of edges $e(s_1, s_2)$, $e(s_2, s_3)$, $e(s_3, s_4)$, and $e(s_4, s_1)$, we define a potential function V_C based on the values of $e(s_1, s_2)$, $e(s_2, s_3)$, $e(s_3, s_4)$, and $e(s_4, s_1)$. It follows that the MRF for the edge process is uniquely defined if we specify the values of the potentials V_C for all possible values of $e(s_1, s_2)$, $e(s_2, s_3)$, $e(s_3, s_4)$, and $e(s_4, s_1)$ for all cliques C. Following Geman and Geman (1984), we set $V_C = 0$ for "no edges," $V_C = .8$ for "lines" and "corners," and $V_C = 2.7$ for "T junctions" and "four-way junctions." It is reasonable to believe that "no edges" is the most likely occurrence in an image (having the lowest value of V_C), whereas "T junctions" and "four-way junctions" are least likely to occur (having the highest values of V_C). With these specifications of potential values, the MRF model for the edge process is given by

$$\pi(E) = \frac{1}{K} \cdot \exp\left\{ -\sum_{C \in \mathcal{C}} V_C(E) \right\}, \tag{14}$$

where K is the normalizing constant.

Given the edges, the multivariate GMRF model in Section 2 is used to model the data within homogeneous regions. Specifically, given E, a realization of the edges on S, we use the conditional distributions

$$\pi(x_s \mid x_{-s}, E) = N_3\left(\sum_{t \in N_s} B_{st}(E) x_t, G_s(E) \right), \tag{15}$$

where the sum now extends over all neighboring pixels t with $e(s, t) = 0$; that is, no edge is present between pixels s and t. In other words, the conditional distribution of x_s given a realization of the edge process depends only on neighbors of x_s that are not separated by an edge (or within the same homogeneous region of the image). For every realization of the edge process, we must ensure that the modified conditionals defined as in (4) consistently define a unique joint distribution on S. The consistency conditions in the model with edges entail that

$$G_s^{-1}(E) B_{st}(E) = \left(G_t^{-1}(E) B_{ts}(E) \right)^T \tag{16}$$

should hold for all pairs of pixels s and t. Choosing B_{st} and G_s to satisfy (2) for every edge realization E, we see that the consistency condition (16) holds. So we have the following hierarchical specification:

- The distribution of E is an MRF with the specified values of the potential function.
- Given E, the underlying color process X is multivariate GMRF with local characteristics defined by (1).
- Given X, $\{Y_i\}$ are independent with

$$Y_i \sim N(x_i, \sigma^2), \qquad i \in obs_s, \tag{17}$$

where σ^2 is a known constant.

The foregoing hierarchical model specifies the full joint distribution of the observable process $Y = (y_{obs}, y_{mis})$. Using the notation $(a, b, \ldots \mid c, d, \ldots)$ to denote the posterior of $(a, b, \ldots)$ given $(c, d, \ldots)$, we can write the full posterior of X and E given y_{obs} as

$$\pi(X, E \mid y_{obs}) \propto \left(\prod_{s \in S} \prod_{i \in obs_s} \exp\left\{ -\frac{1}{2\sigma^2} (y_i - x_i)^2 \right\} \right) \cdot \pi(X \mid E) \cdot \pi(E). \tag{18}$$

We are interested in finding the maximum a posteriori (MAP) estimate of (X, E) that maximizes the posterior in (18). Finding this MAP estimate can be difficult for two reasons. First, the observed likelihood can be intractable for maximization due to the presence of missing observations; second, the joint maximization of (X, E) can itself be hard. The difficulty

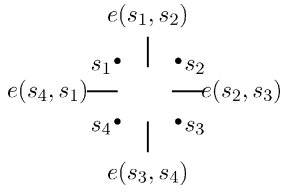

Figure 3. Example of an Edge Clique.

of handling missing values can be overcome by using the EM algorithm of Dempster, Laird, and Rubin (1977) for the observed posterior $(X|y_{obs})$. The E step of the EM algorithm substitutes each missing value at step $t+1$ with its expected value at step t under the posterior distribution $(y_{mis}|X, E, y_{obs})$. It is clear that this expected value is just the underlying missing color intensity at step t. The M step of the EM algorithm involves joint maximization of (X, E), which is difficult. Thus the M step is broken down into two conditional maximization (CM) steps, maximizing the posteriors of $(X \mid E, y)$ and $(E \mid X, y)$ with respect to X and E. Maximizing the conditional posterior distributions $(X \mid E, y)$ and $(E \mid X, y)$ is relatively easy. In the case of $(E \mid X, y)$, because we used an MRF model for the prior edge process, it follows that $(E \mid X, y)$ is also a MRF with an updated collection of cliques. Thus we can use simulated annealing to obtain the conditional maximum of $(E \mid X, y)$ (see Geman and Geman 1984). The conditional posterior distribution $(X \mid E, y)$ is a mixture of independent GMRF models, with conditional modes given by the posterior means. The EM algorithm and the CM updates will ensure only that a local maximum is obtained, not that it will always converge to a global maximum.

We illustrate and compare the performance of this method in Section 6. It turns out that in general, using the edge process does not work as well in the presence of missing data. The algorithm is too flexible and ends up with too many edges. It appears that the missing data cause the algorithm to get stuck at suboptimal local maxima.

4. MODELING WITH AN UNDERLYING CLASS PROCESS

In this section we consider an alternative approach based on a spatial mixture model with underlying class processes. This approach fixes the number of homogeneous regions and thus limits the tendency inherent in the edge process to detect too many spatially homogeneous regions.

Let G be the total number of classes (spatially homogeneous regions). We use a MRF to model membership in the classes. Specifically, define a collection of random variables $C = \{c_s, s \in S\}$, with each c_s taking values in $\{1, 2, \ldots, G\}$. The distribution of C can be taken to be a variation of Strauss's color model (Strauss 1975, 1977) with joint density

$$\pi(C) = K \cdot \exp\left\{\sum_{u=1}^{G} \alpha_u n_u + \sum_{u \neq v} \beta_{\{uv\}} n_{\{uv\}}\right\}, \tag{19}$$

where n_u is the number of pixels with $c_s = u$, $n_{\{uv\}}$ is the number of neighboring pairs with $c_s = u$ and $c_t = v$, and K is the normalizing constant. The joint distribution can also be written as

$$\pi(C) = K \cdot \exp\left\{\sum_{u=1}^{G} \alpha_u \sum_{s \in S} I\{c_s = u\} + \sum_{u \neq v} \beta_{\{uv\}} \sum_{s \sim t} I\{c_s = u\} \times I\{c_t = v\}\right\}, \tag{20}$$

where $s \sim t$ denotes the sum over all pairs of neighbors $\{s, t\}$. The parameters α_u measure the overall intensity of each class value in the image, whereas $\beta_{\{uv\}}$ measures the degree of association between values u and v of the class assignment. The local characteristics of c_s, given c_{-s}, are

$$\pi(c_s = u \mid c_{-s}) \propto \alpha_u + \sum_{v \neq u} \beta_{\{uv\}} \sum_{t \in N_s} I\{c_s = u, c_t = v\}, \tag{21}$$

where $\mathcal{N} = \{N_s, s \in S\}$ denotes the neighborhood system of choice. We consider two choices for the neighborhood system—namely, the first- and second-order neighborhood systems—for the reconstruction of images. We take $\beta_{\{uv\}} = \beta$ for all pairs (u, v) and take $\alpha_u = \alpha$ for all $u = 1, 2, \ldots, G$. In other words, we assume that all component pairs have the same degree of association and that each component color has the same overall intensity value. It turns out that the value of α plays no role in the reconstruction, so we take it to be 0 throughout. Thus our model for the class process turns out to be the Potts model (Potts and Ward 1955; Bowsher et al. 1996). The parameter β controls the degree of spatial correlation between neighboring class assignments; higher values of β correspond to similar neighborhood class assignments, whereas small values of β allow neighboring classes to differ from one another. With these choices, (21) becomes

$$\begin{aligned}\pi(c_s = u \mid c_{-s}) &\propto \beta \sum_{t \in N_s} \sum_{\substack{v=1 \\ v \neq u}}^{G} I\{c_s = u, c_t = v\} \\ &= \beta \cdot (n_s - n_{uu}),\end{aligned} \tag{22}$$

where $n_s = |N_s|$ is the total number of neighbors of s and n_{uu} is the number of neighbors, t, of s such that $c_t = u$.

Given a realization of $C = \{c_s, s \in S\}$, the underlying spatial process X is modeled as a GMRF, similar to the one in Section 2, with local characteristics

$$x_s \mid x_{-s} \sim N\left(\mu + \sum_{t \in N_s} B_{st}(C)(x_t - \mu), \Gamma_s(C)\right). \tag{23}$$

$B_{st}(C)$ is chosen as

$$B_{st}(C) = \frac{w_{st}(C)}{\sum_{t \in N_s} w_{st}(C)} \cdot I_{p \times p} = \frac{w_{st}(C)}{n_s(C)} \cdot I_{p \times p}, \tag{24}$$

say, for some nonnegative weights $w_{st}(C)$ with $n_s(C) = \sum_{t \in N_s} w_{st}(C)$. Also, $\Gamma_s(C)$ is chosen as

$$\Gamma_s(C) = \Gamma / n_s(C). \tag{25}$$

It is easily verified that the consistency conditions hold if $w_{st}(C) = w_{ts}(C)$. In Section 3, w_{st} was taken to be 1 if no edge was present between pixels s and t, and 0 otherwise. The choice of $w_{st}(C)$ is taken to be

$$w_{st}(C) = \begin{cases} 1 & \text{if } c_s = c_t \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

For this choice of weights, the joint distribution of (c_s, x_s) given (c_{-s}, x_{-s}) is

$$\pi(c_s = u, x_s \mid c_{-s}, x_{-s}) \propto \pi(x_s \mid c_s = u, x_{-s}, c_{-s}) \cdot \pi(c_s = u \mid c_{-s}), \tag{27}$$

where

$$\pi(x_s \mid c_s = u, x_{-s}, c_{-s}) = \exp\left\{-\frac{1}{2}\cdot\sum_{t\in N_s} w_{st}(C)(x_s-\bar{x}_s)^T \times \Gamma^{-1}(x_s-\bar{x}_s)\right\}, \quad (28)$$

with $\bar{x}_s = 1/n_s \sum_{t\in N_s} x_t$ and

$$\pi(c_s = u \mid c_{-s}) \propto \exp\{\alpha+\beta\cdot(n_s - n_{uu})\}, \quad (29)$$

as given in (22).

The reconstruction algorithms for retrieving (X, C) from the observable data y_{obs} are based on the iterative conditional mode (ICM) algorithm. This algorithm iteratively maximizes the conditional modes of the posterior distributions

$$\pi(x_s \mid x_{-s}, C, y_{obs}) \propto \exp\left\{-\frac{1}{2\sigma^2}(y_{obs_s} - x_{obs_s})^2\right\} \cdot \pi(x_s \mid c_s = u, x_{-s}, c_{-s}) \quad (30)$$

and

$$\pi(c_s \mid c_{-s}, X, y_{obs}) \propto \pi(x_s \mid c_s = u, x_{-s}, c_{-s}) \cdot \pi(c_s = u \mid c_{-s}), \quad (31)$$

where the two expression on the right of (31) are as given in (28) and (29). Equation (30) updates the attributes, whereas (31) updates the class assignments. From (28), the estimate of x_{mis_s} depends only on the x_t's for which $c_s = c_t$. Thus, while updating attributes using (28), the imputation algorithm interpolates within boundaries of homogeneous regions. The class updates are as follows. From (28), we see that the first term on the right side of (31) implies that the class c_s should chosen to be close to the c_t's for which x_s is close to $\bar{x}_s$ in the Mahalanobis distance $\Delta = (x_s-\bar{x}_s)^T\Gamma^{-1}(x_s-\bar{x}_s)$. Thus in the presence of an edge, c_s is chosen to be the same class as its neighbors within the same homogeneous region so that x_s will be close to $\bar{x}_s$, and hence Δ will be small, satisfactory reconstruction in practice.

Implementation of the foregoing procedure assumes that the number of classes is known. Estimating the number of classes is similar to the problem of estimating the number of components in a mixture model and is known to be difficult. There is an extensive literature on this topic, and many different approaches have been discussed, including the use of likelihood-based methods and information-theoretic techniques such as the Akaike information criterion and Bayes information criterion (see McLachlan and Peel 2000 for a recent review). The statistical theory and inference associated with this is complicated because it is a nonregular estimation problem. In the kinds of application that we have in mind, computational speed is also an important requirement. Thus here we have restricted attention to heuristic techniques for estimating the number of classes based on multidimensional clustering.

Specifically, we considered a measure of dissimilarity between pixels based on both spatial distance and attribute information and used a standard clustering algorithm to group the pixels into homogeneous groups. We have examined different heuristic rules for deciding on the number of clusters. When using with a hierarchical clustering algorithm, one can interact visually with the cluster tree to decide where to cut the tree and determine the number of clusters. This seems to work well when only few edges or boundaries are present. With more complex images, we have found that partitioning them into smaller subregions and applying the clustering algorithm to the subproblems works better.

The dissimilarity measure between pixels that we considered uses both attributes and spatial distances as follows:

$$D(s,t) = \phi(D_C(x_s, x_t), D_S(s,t)), \quad (32)$$

where $D_C(x_s, x_t)$ is a (normalized) distance or dissimilarity measure of the attributes x_s and x_t and $D_S(s,t)$ is a (normalized) measure of spatial distance of the pixels s and t. This measure can be motivated from the expression of the conditional posterior of (x_s, c_s) given (x_{-s}, c_{-s}). (See also Chu et al. 1998 for a discussion of similar measures in the context of robust edge-preserving smoothers.) In our simulations, we have specifically used $D(s,t) = aD_C(x_s, x_t) + (1-a)D_S(s,t)$, where $0 < a < 1$ is a tuning parameter.

Another advantage of this approach is that it also provides an initial assignment of the pixels to classes. In our algorithm, we use this class assignment as the starting values in the iterative procedure (with ICM and spatial smoothing) to update the classification. The simulation results are given in Section 6.

5. IMPUTATION USING EDGE-PRESERVING SMOOTHERS WITH LOCALLY VARYING WEIGHTS

One drawback of the previous approaches is that they rely on explicitly detecting the edges and then doing imputation within the regions identified to be homogeneous. The method in Section 4 requires determining the number of classes, which can be time-consuming. Here we consider an alternative based on edge-preserving spatial smoothing. To keep the notation and discussion simple, we assume that that the observed attributes (colors) are recorded without any errors ($\sigma^2 \equiv 0$). The ideas can be easily extended to the case with additional degradation.

We consider local estimation of missing colors at each site s. We need the following notation, which is specific to site s. By rearranging the components of x_s as necessary, let $x_s = (x_{mis_s}^T, x_{obs_s}^T)^T = (x_{1s}, x_{2s})^T$. For $t \in N_s$, let x_{1t} be the attributes of x_t that correspond to the missing components of x_s, and let x_{2t} be those corresponding the observed components of x_s. Correspondingly, rearrange the variance covariance matrix Γ as

$$\Gamma = \begin{pmatrix} \Gamma_{1s,1s} & \Gamma_{1s,2s} \\ \Gamma_{2s,1s} & \Gamma_{2s,2s} \end{pmatrix}.$$

We use the color image example to motivate the basic idea. Suppose that R is the only observed color at pixel s. Then, according to our local notation, $x_s = (B_s, G_s, R_s)$ and $x_t = (B_t, G_t, R_t)$ for all $t \in N_s$. We can estimate the missing value, B_s, by computing an average of the B_t's in the local neighborhood, N_s. However, we want to average only over pixels that are in the same homogeneous region as pixel s.

Because we are not explicitly considering edge detection here, an alternative is to use a weighted average that gives greater weight to pixels that are more likely to be in the same homogeneous region as pixel s. The weights are data-based and depend on how similar the observed value of R_s is to the values of R_t for $t \in N_s$. If an edge is present between pixels s and t, then R_s and R_t are very different, so the weight should be small. Therefore, we choose the weights to decrease the influence of neighboring sites that are not similar. This is the basis of the iterative updating scheme discussed later.

Formally, for a fixed set of weights, the local cokriging equation for $x_{1s} = x_{mis_s}$ is the conditional expectation of x_{1s} given the observed data in the local neighborhood N_s and is given by

$$\sum_{t \in N_s} w_{st} x_{1t} + \Gamma_{1s,2s} \Gamma_{2s,2s}^{-1} \cdot \sum_{t \in N_s} w_{st} (x_{2s} - x_{2t}). \tag{33}$$

We use the estimator in (33) with data-dependent weights to do adaptive imputation and take Γ to be the identity matrix. Our weights depend on the absolute differences $|x_{2s} - x_{2t}|$. In our reconstruction algorithms, we use weights of the form $w_{st} = K_2 \cdot w(x_{2s} - x_{2t})$, where

$$w(x) = \frac{1}{\left(|x|^{\alpha^*} + \theta^*\right)} \tag{34}$$

for positive real numbers α^* and θ^* and K_2 is the normalizing constant such that the weights w_{st} sums to unity. We present a Bayesian motivation for such weights in the Appendix. The addition of θ^* also prevents singularities at $x = 0$.

The iterative algorithm works as follows. First, an initial estimate of the partially observed image, such as the Bayer estimate, is obtained. Then the missing attributes are estimated using the local cokriging equations given in (33) for all the sites. The specific order of sites to be sequentially updated is arbitrary; the reconstructed images in Section 6 were obtained by raster scan updating of the sites. The updating of attributes at all sites is cycled several times until convergence.

This approach using adaptive weights offers several advantages. First, the number of components in a spatial mixture need not be estimated. This entails significant reduction of computational time. By choosing adaptive weights, we are able to preserve edges and overcome the problem with ordinary cokriging. The local nature of the algorithm also enables us to perform the estimation procedure quickly. This step avoids the inversion of matrices of large dimensions. In the color image application, only the inverse of matrices of dimension 1, $\Gamma_{2s,2s}$, [see (33)] must be found.

The local cokriging equations in (33) can also be motivated through the use of a multivariate pairwise difference prior for the spatial process X, namely

$$\pi(x) = K \cdot \exp\left\{ -\frac{1}{2} \sum_{s \sim t} w_{st} (x_s - x_t)^T \Gamma^{-1} (x_s - x_t) \right\}, \tag{35}$$

where w_{st} are positive weights and K is the appropriate normalizing constant (see the Appendix for details). Large (small) values of w_{st} indicate that the color intensities x_s and x_t are similar (dissimilar). Thus the normalized weights, $w_{st} / \sum_{t \in N_s} w_{st}$, represent a continuous version of the 0–1 edge process in Section 3. We show in the Appendix that the criterion in (34) turns out to be the posterior mode of w_{st} for a suitable choice of prior. The weights are chosen adaptively at the updating step. The foregoing procedure can also be viewed as Bayesian model averaging, in which 0–1 weights (or edges) are averaged and this average weight is used to compute the estimates of missing colors in (33). Our final goal is interpolation of missing colors, and so it is sufficient to obtain a locally adaptive estimate of the missing colors at each updating step without having to explicitly recover boundaries, as was done in Section 3.

We now discuss some connections of the aforementioned procedure with local M estimation. The estimate of x_{mis_s}, using (33), can also be obtained as

$$\hat{x}_{mis_s} = \arg\min_{x_{mis_s}} \sum_{t \sim s} w_{st} \rho(x_s - x_t), \tag{36}$$

where $x_u = (x_{mis_u}^T, x_{obs_u}^T)^T$ for $u = s, t$ and $\rho(x) = x^T \Gamma^{-1} x$. In other words, the estimated value of x_{mis_s} can be seen as a local M estimate for the function $\rho(x) = x^T \Gamma^{-1} x$ (see Chu, Glad, Godtliebsen, and Marron 1998). Chu et al. obtained the local M estimator by choosing weights to be inversely proportional to the distance between covariate observations; imputation of the dependent variable at site s would involve observations with similar covariate values as those of site s. In the case of spatial interpolation, information on location of sites was also included as covariate information, resulting in local smoothing based on neighboring sites. In our interpolation scheme, positive weights are assigned only to nearest neighbors of a site. (This corresponds to hard thresholding for the location covariate information in the case of local M estimators.) Covariate information at neighboring sites s and t is taken to be x_{2s} and x_{2t}. Thus at each iteration of our imputation scheme, x_{mis_s} is obtained as an M smoothing estimator with covariate information (s, x_{2s}) and (t, x_{2t}). As we move from one site to another in raster scan fashion, we estimate different components of x_s (either R, G, or B) with corresponding observed components as covariate information according to the Bayer design.

We have only investigated the performance of the local M estimator based on the quadratic distance $\rho(x) = x^T \Gamma^{-1} x$. In our simulation results, this choice of $\rho(\cdot)$ satisfactorily recovers edges between homogeneous regions in the presence of missing observations. Several other choices of $\rho(\cdot)$ in the multivariate case can be considered, including multivariate versions of more robust analogs such as $\rho(\cdot) = |x|$, or any function of the form $\psi(x) = \rho'(x)$, where $\psi(x)$ is bounded. These may lead to better edge-preserving imputation with missing data, although problems may occur due to multiple local minima in (36). This will be considered in future research.

6. SIMULATION RESULTS

6.1 Artificial Images

Figures 4 and 5 provide a comparison of the three methods on two simulated images. Figures 4(a) and 5(a) correspond to the true images [same as the true images in panels Figure 2(a) and (c)]. They were generated using GMRFs with mean color intensities for (R, G, B) (90, 90, 90) (darker)

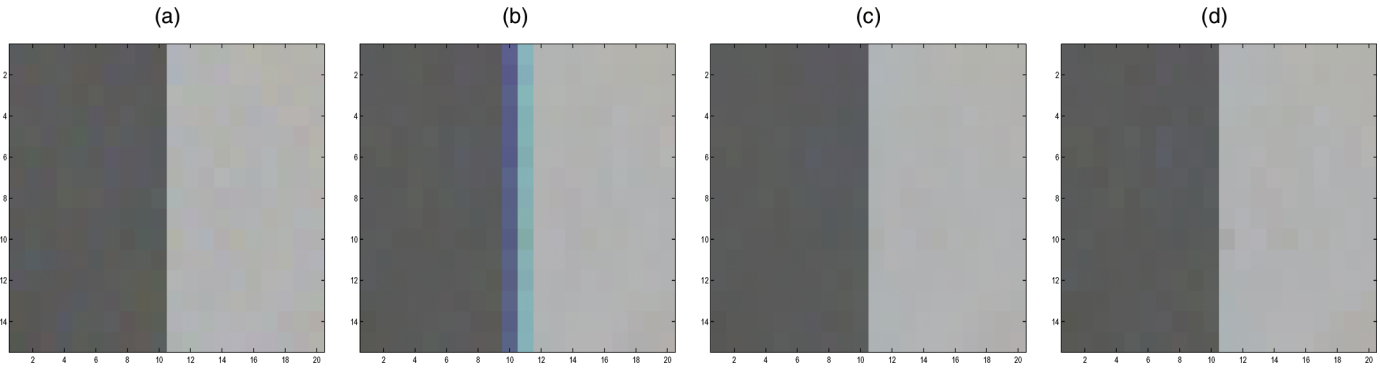

Figure 4. Reconstructed Images. (a) Original image, (b) edge process, (c) class process, (d) edge-preserving smoothers with adaptive weights.

and (180, 180, 180) (lighter), variance 20 and correlation of .4 between R, G, and B. The value of ϕ was chosen to be .99.

Figures 4(b) and 5(b) show the results of the imputation and recovery using underlying edge processes. Although the reconstruction is smoother than the Bayer interpolation in Figure 2, the algorithm is unable to recover the true boundary sharply. The algorithm appears to be too flexible and to detect too many edges.

The reconstructions using class processes are shown in Figures 4(c) and 5(c). A second-order neighborhood structure (3×3 neighborhood) was used for the MRF class process. The starting value for the initial class assignment was obtained from a multidimensional cluster analysis. We also estimated the number of spatial mixtures as two from this analysis. The particular resconstruction shown is based on the (relatively large) value of $\beta = 20$. We tried a number of different values of β ranging from 1 to 20, and the reconstructions were robust to changes in the β values in these two examples. This may be due in part to the fact that the initial class assignments based on cluster analysis correctly determined the image boundaries. We consider this issue in more detail later.

Figures 4(d) and 5(d) show the reconstructions based on edge-preserving imputations with adaptive weights. We used $\alpha^* = 1.0$ and $\theta^* = .01$ in selecting the weights [see (34)]. We see that the methods using class processes and edge-preserving smoothers do well in both of these examples and recover the edges exactly.

We now consider some additional simulations to study the impact of initial class assignment and the tuning parameters (neighborhood structure and choice of β) on the method in Section 4. We use a bivariate problem with two colors (R, G), with B held fixed. Figure 6 shows the two original images. They were simulated using a spatial mixture of GMRFs for (R, G) with mean intensities (90, 90) (darker side) and (150, 150) (lighter side), variance 30, and correlation .8. B was held fixed at 100 for all pixels. The missing-data pattern corresponded to observing R or G at only alternating pixels.

Figures 7 and 8 show the reconstructed images based on underlying class processes. The top and bottom rows correspond to first- and second-order neighborhoods for the class process. The columns correspond to three different values of the tuning parameter, $\beta = 2, 10, 20$. Higher values of β correspond to a greater penalty for different neighboring class assignments.

We used two different initial class assignments for the two panels in Figure 6. For Figure 6(a), pixels in the interior were all assigned correctly, those on the left boundary (column 5) were assigned to class 1, and those on the right boundary (column 6) were assigned to class 1 or class 2 alternately, starting with class 2 for the topmost pixel. The initial assignment of pixels to classes for Figure 6(b) was done as follows: pixels in the interior were all assigned correctly, but those on the boundary were assigned to classes 1 or 2 randomly with probability .5. The ICM reconstruction algorithm was run on both panels until it converged. In all of the experimental settings, it converged within 30 iterations.

Figures 7 and 8 allow us to draw several conclusions about the impact of the tuning parameters. In both cases, the

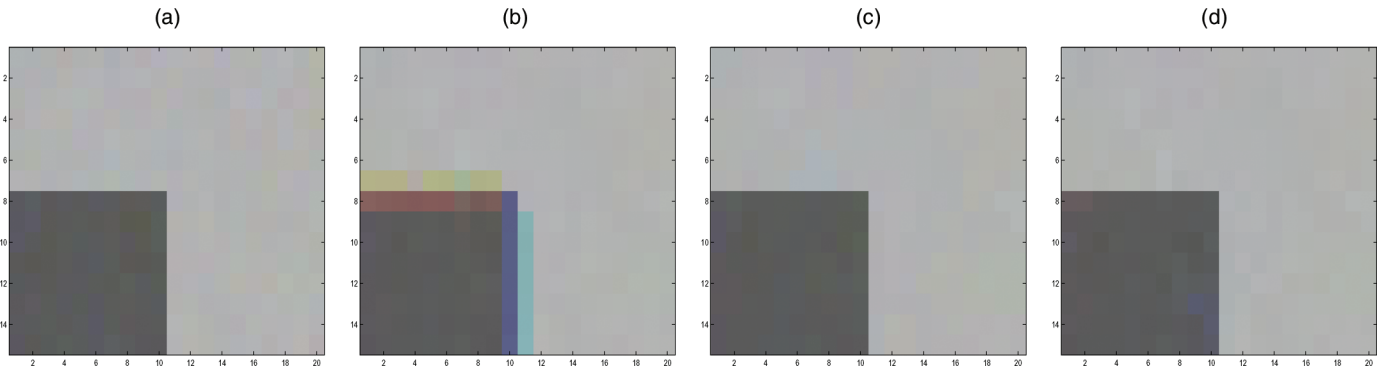

Figure 5. Reconstructed Images. (a) Original image, (b) edge process, (c) class process, (d) edge-preserving smoother with adaptive weights.

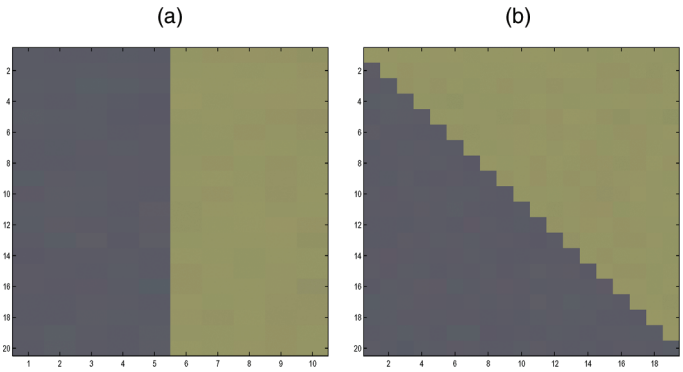

Figure 6. Original Images Used in the Two-Color Simulation Study. (a) Edge image; (b) diagonal image.

use of second-order neighborhoods worked well for all values of β. Reconstructions based on first-order neighborhoods did not work as well with small β. Increasing β improves performance and allows one to recover from the bad initial class assignments for the diagonal image in Figure 8, but not for the straight line edge in Figure 7. This is an artifact of the "confounding" between the neighborhood structure, edge pattern, and initial class assignments in this particular case. In general, higher values of β allow better recovery of edges. It should be noted, however, that the boundaries are sharply defined in these examples. In situations with less well-defined boundaries, large values of β can lead to underestimation of the number of classes and thus oversmoothing across class boundaries. The quality of edge reconstruction depends heavily on the choice of good starting values. In this respect, second-order neighborhood structures appear to be more robust with respect to initial starting values.

Figure 9 shows the reconstructed images based on (essentially) kriging (marginal information in just each of the components). More precisely, the missing values of R were imputed using only the observed Rs (similarly for the missing values of G). The reconstruction is based on second-order neighborhoods with $\beta = 20$ was for the class processes. These should be compared with Figures 7(f) and 8(f) which correspond to (essentially) cokriging. We see that borrowing multivariate information from neighboring sites enables one to better decide whether an edge is present at a given pixel location.

We also examined the influence of the tuning parameters on the edge-preserving imputation method in Section 5. We reconstructed the images in Figures 7 and 8 with several values of α^*: $\alpha^* = .2, .5, 1.0$, and 2.0. The value of θ^* was fixed at .01. Both first- and second-order neighborhoods were considered. All of these reconstructions picked up the edges well, giving essentially the same recovered images as in Figures 7(f) and 8(f). Computing the reconstructions were fast for both the first- and second-order neighborhood structures; convergence was obtained within 30 iterations of the algorithm.

6.2 Real Images

Figure 10 is the Bayer interpolation of a color CCD image from a real (professional) football game. We will use this to compare the two methods based on class processes and edge-preserving imputations. We do not consider the method based on edge processes here. The algorithm based on the

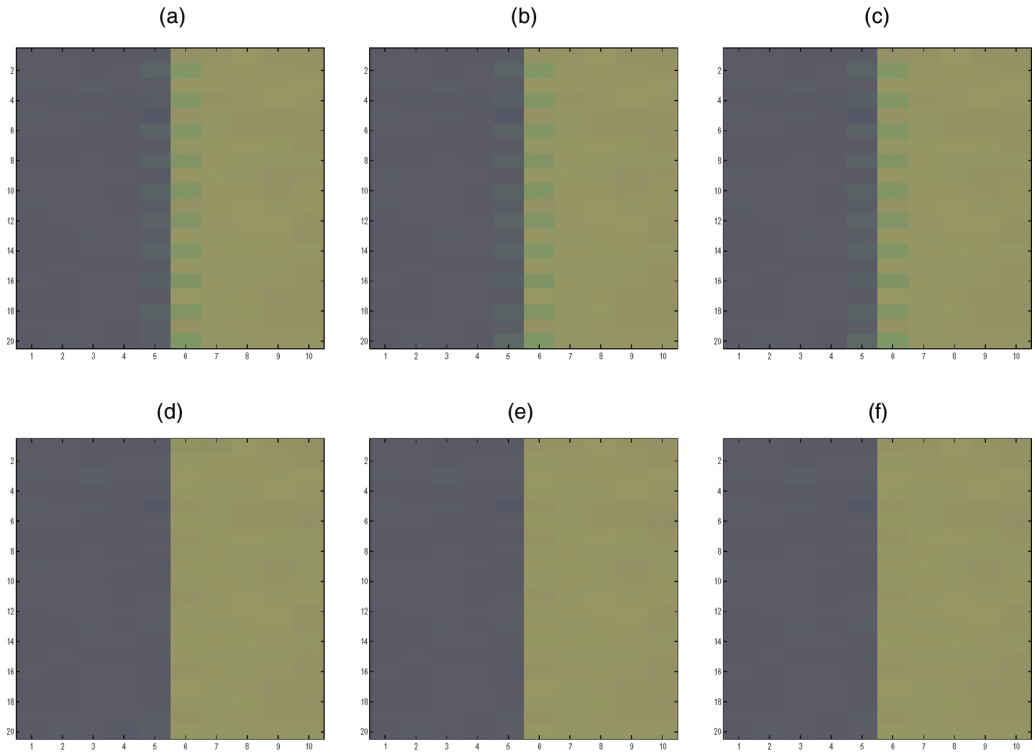

Figure 7. Reconstructions of Straight Line Edge Image. (a)–(c) First-order neighborhood with $\beta = 2, 10, 20$; (d)–(f) second-order neighborhood with $\beta = 2, 10, 20$.

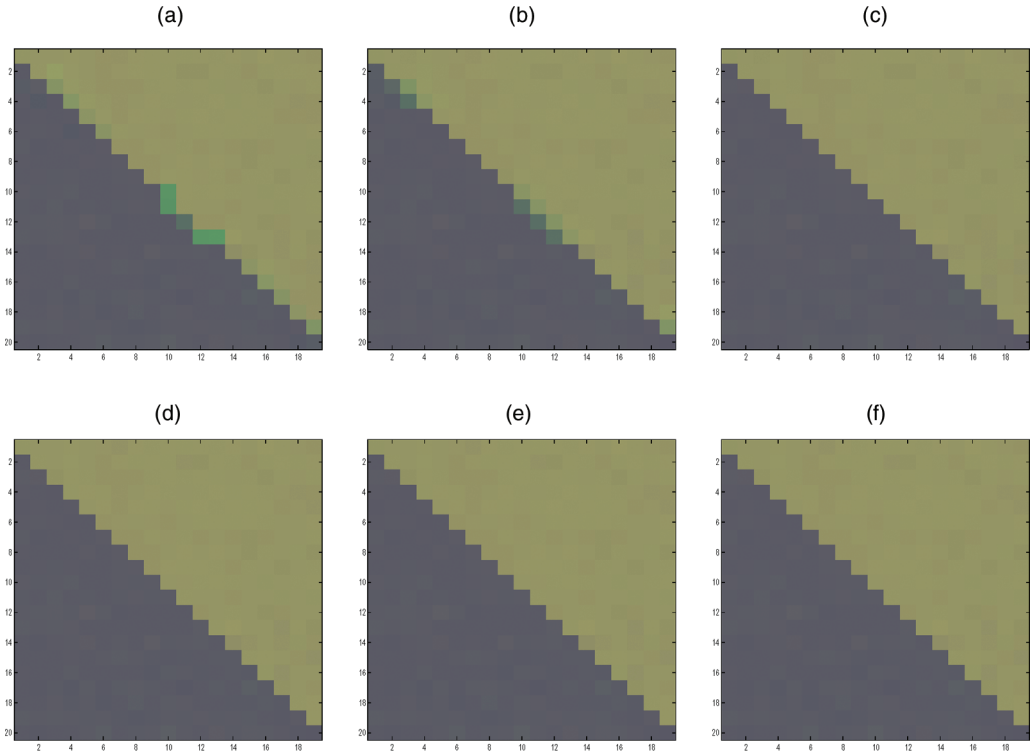

Figure 8. Reconstructions of Diagonal Edge Image. (a)–(c) First-order neighborhood with $\beta = 2, 10, 20$; (d)–(f): second-order neighborhood with $\beta = 2, 10, 20$.

class process was applied to the image in Figure 10 by sliding a 15×20 window across the entire image, each time carrying out Gaussian interpolation after class assignments were determined by the clustering procedure. The difference in the reconstructed images can be clearly seen when the images are examined at higher levels of detail. Figure 11(a) is the Bayer interpolation of one of the player's thighs [corresponding to pixels around the coordinate $(x, y) = (90, 200)$ in Fig. 10]. Figures 11(b) and (c) are reconstructions based on class processes and edge-preserving imputation. We see that both of them do a much better job of recovering the edge sharply. Figure 12(a) is another thigh image [corresponding to pixels around the coordinate $(x, y) = (350, 75)$ in Fig. 10]. Here the original Bayer scheme was adequate, although the boundary is still somewhat blurred. The two methods discussed here give better reconstructions. Due to space restrictions, we do not present the entire reconstructed images here.

7. SUMMARY AND CONCLUSIONS

We have considered several methods for imputing missing multivariate data and recovering images in the presence of edges. The information from the multivariate data is seen to be critical in determining the edges and doing spatial imputation.

The method based on underlying edge processes appears to detect too many edges and is not recommended. The method based on class processes works quite well in the examples that we have seen. However, in situations where the boundaries are not as sharply defined, the posterior distributions for the class assignments can be badly multimodal, leading to underestimation of the number of classes. This is particularly true with large values of β or larger neighborhood structures. In general, edge-preserving imputation with locally varying weights performs best. One advantage of this approach is that it does not require explicit reconstruction of the edges or classes. We show in the Appendix that this method can be viewed as modeling with an underlying edge process, analogous to Section 3, for a suitable prior. However, instead of actually reconstructing the edges at each stage of the algorithm and then smoothing within boundaries, this method uses the posterior information about the edges to compute locally weighted averages and thus impute the missing values. This is one explanation for the more stable reconstructions. One can also view this method, at least conceptually, as Bayesian model averaging of results from binary edge processes.

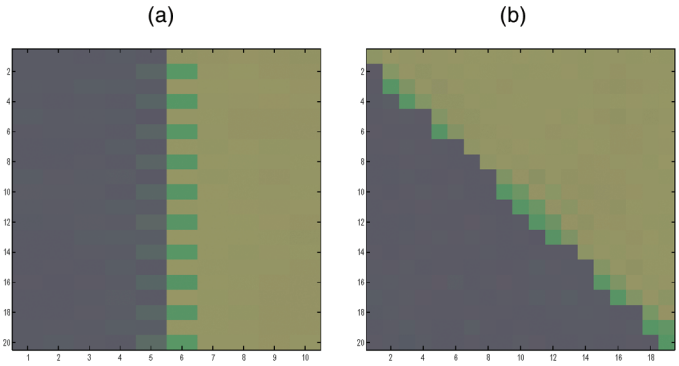

Figure 9. Reconstructed Images With Kriging; Second-Order Neighborhood With $\beta = 20$.

Figure 10. Original Bayer Reconstruction; Image Size 281 × 400.

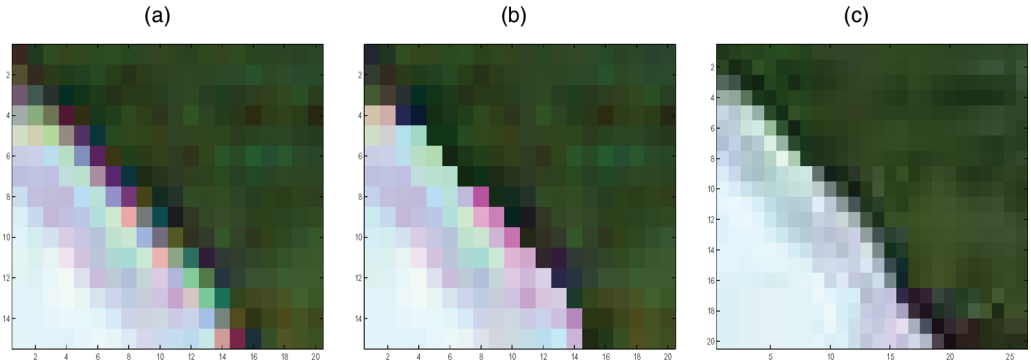

Figure 11. (a) Bayer Interpolation of Player's Thigh, (b) Reconstruction Using Class Process, and (c) Reconstruction Using Adaptive Weights.

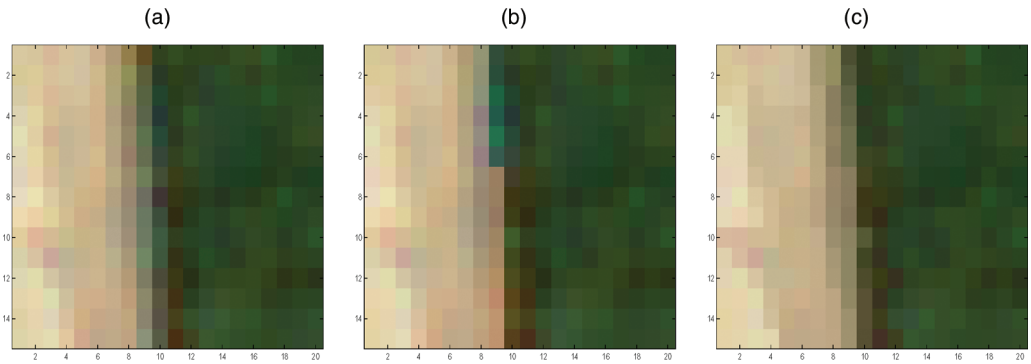

Figure 12. (a) Bayer Interpolation of Another Player's Thigh, (b) Reconstruction Using Class Process, and (c) Reconstruction Using Adaptive Weights.

APPENDIX:

We establish a connection between the results in Sections 5 and 3 by motivating the choice of locally adaptive weights for interpolation using a suitable prior distribution on w_{st}. Define a matrix $U = ((u_{st}))$ by $u_{ss} = \sum_{t \in N_s} w_{st}$, $u_{st} = -w_{st}$ with $w_{st} = w_{ts}$ for neighbors s and t, and $u_{st} = 0$ elsewhere. Denote the generalized Gaussian density for $x = (x_1, x_2, \ldots, x_N)$ by

$$\pi(x) = \frac{1}{(2\pi)^{d/2}} |U|^{1/2} \cdot \exp\{-x'Ux\}, \tag{A.1}$$

where d is the number of nonzero eigenvalues of U and $|U|$ is the product of the nonzero eigenvalues. Equation (A.1) can be rewritten as the first-order pairwise difference prior

$$\pi(x) = \frac{1}{(2\pi)^{d/2}} |U|^{1/2} \cdot \exp\left\{ - \sum_{\substack{(s,t) \\ \text{neighbors}}} w_{st}(x_s - x_t)^2 \right\}. \tag{A.2}$$

Consider the prior distribution on the vector of weights, $w = (w_{st})$,

$$\pi(w) \propto \frac{1}{|U|^{1/2}} \cdot \prod_{\substack{(s,t) \\ \text{neighbors}}} \exp\{-w_{st} a_{st}\} \prod_{\substack{(s,t) \\ \text{neighbors}}} w_{st}^{b_{st}}, \tag{A.3}$$

for some positive constants a_{st} and b_{st}. Using Bayes's rule, the posterior distribution of w_{st} (given x) can be seen to be independent with

$$\pi(w_{st}) \propto \exp\{-w_{st}((x_s - x_t)^2 + a_{st})\} \cdot w_{st}^{b_{st}}, \tag{A.4}$$

with mode at

$$\hat{w}_{st} = \frac{b_{st}}{\left((x_s - x_t)^2 + a_{st}\right)}. \tag{A.5}$$

This corresponds to (34) with $\alpha^* = 2$, $a_{st} = \theta^*$, and $b_{st} = 1$. Equation (A.5) shows how the weights can be chosen adaptively for interpolation using information based on x_s and x_t. Assume that site s belongs to the interior of a homogeneous region. In this case the differences $x_s - x_t$, $t \in N_s$ are of the same order of magnitude, making the weights $\hat{w}_{st}/(\sum_{t \in N_s} \hat{w}_{st})$ approximately constant and equal to $1/|N_s|$. On the other hand, if s is an edge site, then the differences $x_s - x_t$, $t \in N_s$ will greatly vary, with the smaller values of weights assigned to the w_{st} with large differences $x_s - x_t$, $t \in N_s$. A renormalized version of the weights consequently puts high (low) weights to sites t for which the differences $x_s - x_t$, $t \in N_s$ are small (large). Thus neighboring sites t belonging to the same region has s will be preferred for interpolation.

We now consider updating the missing colors at site s of the image. Assume that G was observed at site s and that (R, B) are the missing colors [i.e., $x_{1s} = (R, B)$ and $x_{2s} = G$ in Sec. 5]. In the conditional updating step, the color attributes (R, G, B) at all neighboring sites are given. Assume initially that Γ is the identity matrix in (35). Then the multivariate pairwise difference prior can be written as

$$\pi(x) = \frac{1}{(2\pi)^{3d/2}} \cdot |U|^{3/2} \exp\left\{ - \sum_{c=\{R,G,B\}} \sum_{\substack{(s,t) \\ \text{neighbors}}} w_{st}(x_{c,s} - x_{c,t})^2 \right\}, \tag{A.6}$$

where $x_{c,s}$ represents the intensity of color c at site s.

Estimates of $x_{R,s}$ and $x_{B,s}$ are given by the local weighted average of intensities,

$$\hat{x}_{c,s} = \frac{\sum_{t \in N_s} w_{st} x_{c,t}}{\sum_{t \in N_s} w_{st}}, \tag{A.7}$$

for $c = \{R, B\}$ if the weights w_{st} are known. However, because they are unknown, we estimate the weights w_{st} using the marginal distribution of x_G and the prior in (A.4). This entails replacing the x_u's in (A.2) and (A.5) by $x_{G,u}$. The estimates of $x_{R,s}$ and $x_{B,s}$ are then given by the estimated local means

$$x_{R,s} = \frac{\sum_{t \in N_s} \hat{w}_{st} x_{R,t}}{\sum_{t \in N_s} \hat{w}_{st}} \quad \text{and} \quad x_{B,s} = \frac{\sum_{t \in N_s} \hat{w}_{st} x_{B,t}}{\sum_{t \in N_s} \hat{w}_{st}}. \tag{A.8}$$

For a general Γ, we use the conditional distribution of $(x_{R,s}, x_{B,s})$ given realizations at all other sites and $x_{G,s}$, instead of (A.6). In this case, (A.7) turns out to be the local cokriging equations given in (33), whereas the unknown weights w_{st} are estimated from the marginal distribution of x_G, as before.

[Received April 2001. Revised June 2002.]

REFERENCES

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussions), *Journal of the Royal Statistical Society*, Ser. B, 36, 192–236.

——— (1986), "On the Statistical Analysis of Dirty Pictures" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 48, 259–302.

Bowsher, J. E., Johnson, V. E., Turkington, T. G., Jaszczak, R. J., Floyd, C. E., and Coleman, R. E. (1996), "Bayesian Reconstruction and Use of Anatomical a priori Information for Emission Tomography," *Transactions in Medical Imaging*, 12, 673–686.

Chellappa, R., and Jain, A. (eds.) (1991), *Markov Random Fields, Theory and Application*, New York: Academic Press.

Chu, C. K., Glad, I. K., Godtliebsen, F., and Marron, J. S. (1998), "Edge-Preserving Smoothers for Image Processing" (with discussion), *Journal of American Statistical Association*, 93, 526–556.

Cressie, N. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

Dass, S. C. (2000), "On the Propriety of Posteriors for Gaussian Markov Random Fields for Reconstructing Images With Missing Observations," technical report, University of Michigan, Dept. of Statistics.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Derin, H., and Elliot, H. (1987), "Modelling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9, 39–55.

Geman, S., and 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., and McClure, D. E. (1987), "Statistical Methods for Tomographic Image Reconstruction," in *Proceedings of the 46th Session of the International Statistical Institute, Vol. 4 (Tokyo 1987)*, pp. 5–21.

Guyon, X. (1995), *Random Fields on a Network*, New York: Springer-Verlag.

Holst, G. C. (1998), *Sampling, Aliasing, and Data Fidelity for Electronic Imaging Systems, Communications, and Data Acquistion.*, SPIE Optical Engineering Press.

Jeng, F. C., and Woods, J. W. (1991), "Compound Gauss-Markov Random Fields for Image Estimation," *IEEE Transactions in Signal Processing*, 39, 683–691.

Le, N. D., Sun, W., and Zidek, J. V. (1997), "Bayesian Multivariate Spatial Interpolation With Data Missing by Design," *Journal of the Royal Statistical Society*, Ser. B. 59, 501–510.

Le, N. D., and Zidek, J. V. (1992), "Interpolation With Uncertain Spatial Covariances: A Bayesian Alternative to Kriging," *Journal of Multivariate Analysis*, 43, 351–374.

Long, A. E., and Myers, D. E., (1997), "A New Form of the Cokriging Equations," *Mathematical Geology*, 29, 685–703.

Mardia, K. V. (1988), "Multidimensional Multivariate Gaussian Markov Random Fields with Applications to Image Processing," *Journal of Multivariate Analysis*, 24, 264–285.

McLachlan, G., and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.

Potts, R. B., and Ward, J. C. (1955), "The Combinatorial Method and the Two-Dimensional Ising Model," *Programming in Theoretical Physics*, 13, 38–46.

Strauss, D. J. (1975), "A Model for Clustering," *Biometrika*, 62, 467–475.

——— (1977), "Clustering on Coloured Lattices," *Journal of Applied Probability*, 14, 135–143.

Ver Hoef, J. M., and Barry, R. P. (1998), "Constructing and Fitting Models for Cokriging and Multivariable Spatial Prediction," *Journal of Statistical Planning and Inference*, 69, 275–294.

Ver Hoef, J. M., and Cressie, N. (1993), "Multivariable Spatial Prediction," *Mathematical Geology*, 25, 219–240.

Winkler, G. (1995), *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Berlin: Springer-Verlag.