Markov and Recursive Least Squares Methods for the Estimation of Data with Discontinuities

ROBERTO CRISTI, MEMBER, IEEE

Abstract—**An algorithm is presented for smoothing data piecewise modeled by linear equations within regions of a one-dimensional (1-D) or two-dimensional (2-D) field, from measurements corrupted by additive noise. Its main feature is the combination of Markov random field (MRF) models with recursive least squares (RLS) techniques in order to estimate the model parameters within the regions.**

Applications to 1-D and 2-D data are given, with particular emphasis on the segmentation of images with piecewise constant intensity levels.

I. Introduction

THE problem of restoring and segmenting data which can be described by models with piecewise constant parameters has received recent attention by researchers. Examples of applications are the segmentation of images into regions of similar textures and/or intensity levels, or speech signals into phonemes.

Because of the particular nature of the problem, stationary models based on classical linear Gaussian techniques are not adequate, and yield excessive smoothing across the boundaries of the regions.

An interesting approach is given by the doubly stochastic (DS) approach where local models (at the pixel, in two-dimensional (2-D), or sample, in one-dimensional (1-D), levels) and global models (at the region levels) are used to describe the process within the regions, and to model the regions themselves.

At the local level, Markov processes [1]–[3] and autogressive models [4], [5] have appeared in the literature as models for textures, intensity levels, or speech signals. At the global levels the compact regions have usually been modeled by Markov processes, in terms of causal Markov chains [5] or noncausal Markov random fields (MRF's) [1]–[4].

Estimation techniques based on these modeling assumptions have called for recursive algorithms by stochastic [2], [3], [6] or deterministic [7], [8] relaxation, and suboptimal techniques of line processing [9] or reduced update Kalman filtering [5]. In all these instances, once the model is formulated, the main difficulty is to cope with the large dimensions of the search space where to find the optimal solution [1]–[4], or with the growing memory required by the estimator [5].

A common assumption in most of the above mentioned approaches is the fact that the set of local models must be known *a priori*, which calls for parameter estimation on training data.

In this report an algorithm for filtering and segmentation of processes locally modeled by autoregressive equations is presented. This class of processes can be regarded as doubly stochastic (DS) in the sense that the observations Y_t are functions of a random process X_t. Namely

$$Y_t = f_t(X_t, W_t)$$

with $t \in Z^k$ for the $k-D$ problem (Z being the set of integers), W_t a white noise process, X_t a random process, constant within compact regions of Z^k and f_t a known mapping. In our case X_t is assumed to be modeled by a MRF.

The estimation by 2-D Kalman filtering techniques of a DS process with X_t a causal hidden 2-D Markov chain is shown in [11], as an extension of earlier results in 1-D [12]–[14].

Connections between the MRF approach to smoothing, and theories based on smoothness priors are also highlighted. In particular, the likelihood function associated with the MRF models with Gaussian disturbances has the same structure as the penalized likelihood at the basis of smoothing algorithms for nonstationary data found in Kitagawa and Gersch [15], and numerous references therein, going back to Whittaker [16]. In this light, the MRF model has the role of a smoothness prior which provides spatial continuity to the desired estimate of the data.

The problem is stated in Section II, together with relevant properties of MRF models. The estimation algorithm is presented in Sections III and IV, for two- and one-dimensional applications, respectively. Considerations on MRF models as smoothing priors are given in Section V, while applications of the technique are the subject of Section VI.

II. Modeling Assumptions

Let S be a rectangular lattice, subset of the integers Z (in 1-D) or Z^2 (in 2-D) of dimensions N or $N \times N$. On the lattice S we define the processes Λ, X, Y as follows:

a) $\Lambda = \{\Lambda_s, s \in S\}$, $\Lambda_s \in Z$ (or any countable set), the random field of *labels*. They mark different regions on S

Manuscript received December 23, 1987; revised November 20, 1989. This work was supported by the Office of Naval Research under NPS direct funding.

The author is with the Electrical and Computer Engineering Department, Naval Postgraduate School, Monterey, CA 93943.

IEEE Log Number 9038426.

which correspond to different objects (in 2-D) or different segments (in 1-D).

b) $X(\Lambda) = \{X_s(\Lambda), s \in S\}$, $X_s \in R^n$ a random field of *parameters* with realizations *constant within regions of equal label*. Namely

$$X_s(\Lambda) = X_t(\Lambda) \quad \text{if } \Lambda_s = \Lambda_t. \tag{2.1}$$

c) $Y(X, W) = \{Y_s(X, W), s \in S\}$, $Y_s \in R^m$ the process of *observations*, assumed dependent on X as

$$Y_s(X, W) = c_s^T X_s + W_s \tag{2.2}$$

with $c_s \in R^{n \times m}$ known for all s, and W_s a zero mean white Gaussian noise independent on X.

Throughout we assume scalar observations ($Y_s \in R$). Typical examples are images of objects with different light intensity levels and noisy measurements, or different textures. In the former case, the processes X_s, $Y_s \in R$ are associated with the true (X_s) and measured (Y_s) light intensities at point s in the lattice S, and in this case $c_s = 1$, for all $s \in S$. In the latter, each texture realization can be represented by an autoregressive model

$$y_s = \sum_{\tau \in A_s} x_{s,\tau} y_\tau + w_s$$

with $A_s \subset S$ and the coefficients $\{x_{s,\tau}, \tau \in A_s\}$ characterizing the texture at point s.

Since we consider S to be partitioned into compact regions (as in the case of image segmentation) corresponding to sizable objects, the model for the labels Λ is assumed to be a Markov random field [4], [6]–[10] having Gibbs distribution

$$\ln \mathcal{P}(\lambda) = \beta \sum_{s \in S} \gamma_s^h(\lambda) + \gamma_s^v(\lambda) - \ln Z \tag{2.3}$$

with Z a normalizing constant and γ_s^h, γ_s^v *potential functions* associated with the horizontal (γ_s^h) and vertical (γ_s^v) cliques. For each $s = (k, t)$ on the lattice define γ to be of the form

$$\gamma_{k,t}^h(\lambda) = \begin{cases} +1 & \text{if } \lambda_{k,t+1} = \lambda_{k,t} \\ -1 & \text{otherwise} \end{cases}$$

$$\gamma_{k,t}^v(\lambda) = \begin{cases} +1 & \text{if } \lambda_{k+1,t} = \lambda_{k,t} \\ -1 & \text{otherwise.} \end{cases} \tag{2.4}$$

The term β is a parameter and it is evident that (2.3) and (2.4), with $\beta > 0$, penalize transitions ($\gamma = -1$) between regions. Also the actual values and nature of the labels Λ have no particular significance and two realizations λ', λ'' with $\gamma_s(\lambda') = \gamma_s(\lambda'')$ for all $s \in S$ have the same probabilities.

About the process X its realizations are constant within regions as in (2.1) with values from a Gaussian population, namely

$$\mathcal{P}(X_s = x_s) = \mathfrak{N}(\mu_x, P_x) \tag{2.5}$$

$\mathfrak{N}(\cdot, \cdot)$ denoting normal distribution. Also, the values of X are assumed to be independent from region to region, in the sense that

$$\mathcal{P}(X_s = x_s \mid X_t = x_t, \lambda_s \neq \lambda_t) = \mathcal{P}(X_s = x_s) \tag{2.6}$$

for all $s \neq t \in S$, and any realization λ, $x(\lambda)$.

Observation: Previous approaches to segmentation using MRF models assume $X_s \in \mathfrak{F} = \{F_1, F_2, \cdots, F_M\}$ with $F_i \in R^n$ known *a priori*. This is the case of images composed of regions of known intensity levels and/or texture, for example determined from training data. By this approach the values of X can be considered as the labels themselves ($\Lambda = X$) and a MAP estimation follows from the likelihood

$$L(x \mid y) = \ln \mathcal{P}(y \mid x) + \ln \mathcal{P}(x) - \ln \mathcal{P}(y) \tag{2.7}$$

with $\mathcal{P}(y \mid x)$, $\mathcal{P}(x)$ readily computed from (2.2), (2.3). Approaches to the maximization of (2.5) with respect to x on the set $\mathfrak{F}^{N \times N}$ have been presented, based on relaxation (stochastic [7], deterministic [8], [9]) or dynamic programming [22].

Notation: In the sequel we make frequent use of the recursive least squares (RLS) update of the estimate $\hat{x}$ of the vector x based on observations $y_t = h_t x + w_t$ with h_t a matrix of appropriate dimensions, and w_t a realization of a white Gaussian noise process $\mathfrak{N}(0, R_t)$. By the notation

$$[x_1, P_1] = \text{RLS}\,[x_0, P_0 \mid h, y, R] \tag{2.8}$$

we denote the recursion

$$x_1 = x_0 + P_0 h [h P_0 h^T + R]^{-1} (y - h x_0)$$

$$P_1 = P_0 - P_0 h^T [h P_0 h^T + R]^{-1} h P_0. \tag{2.9}$$

III. Two-Dimensional Estimation and Segmentation

In this section we address the problem of estimating the processes X, Λ from the given observations y, a realization of the random variable Y.

On a MAP estimation framework we estimate the labels λ by maximizing the conditional probability

$$\ln \mathcal{P}(\lambda \mid y) = \ln \mathcal{P}(y \mid \lambda) + \ln \mathcal{P}(\lambda) - \ln \mathcal{P}(y). \tag{3.1}$$

The prior on λ is readily given by the Markov assumption given in the previous section

$$\ln \mathcal{P}(\lambda) = \beta \sum_{s \in S} \gamma_s^h(\lambda) + \gamma_s^v(\lambda) - \ln Z. \tag{3.2}$$

The combination of the term $\ln \mathcal{P}(y \mid \lambda)$ is addressed below. It is shown that, given the observed data y, a) the likelihood function can be computed recursively on a line-by-line basis, and b) it depends on the edges $\gamma(\lambda)$ only.

In order to proceed, call $X_k(t)$, $Y_k(t)$, $\cdots$ the random variables associated with the element $s = (k, t)$ of the lattice, and

$$Y_k = \{Y_k(t), 0 \leq t < N\}$$

the observation process associated with line k. Also define the random variables

$$\begin{aligned}\hat{X}_k(t \mid t-1) &= E[X_k(t) \mid Y_k(t-1), \cdots, Y_k(0), \Lambda] \\ \hat{X}_k(t) &= E[X_k(t) \mid Y_k, \cdots, Y_0, \Lambda] \\ \hat{X}_{k \mid k-1}(t) &= E[X_k(t) \mid \hat{X}_{k-1}(t), \cdots, \hat{X}_{k-1}(0), \Lambda] \\ \hat{X}_{k \mid k-1}(t \mid t-1) &= E[X_k(t) \mid Y_k(t-1), \cdots, Y_k(0), \hat{X}_{k-1}(t), \hat{X}_{k-1}(0), \Lambda]. \end{aligned} \tag{3.3}$$

Fig. 1 shows the region of support of the above variables. Also we denote by lower case $(\hat{x}_k(t \mid t-1) \cdots)$ their realizations corresponding to the given realizations y.

Then we can show the following:

Lemma: Given the observations y and the labels λ, the realizations of the random variables (3.3) and their respective error covariance matrices P can be recursively computed by filtering and smoothing operations as follows:

Filtering:

$$[\hat{x}_k(t+1 \mid t), P_k(t+1 \mid t)] = \begin{cases} \text{RLS}\,[\hat{x}_k(t \mid t-1), P_k(t \mid t-1) \mid c_{k,t}^T, y_k(t), \sigma^2] \\ \qquad \text{if } \gamma_{k,t+1}^h = +1 \\ [\mu_x, P_x] \quad \text{otherwise} \end{cases} \tag{3.4a}$$

$$[\hat{x}_{k \mid k-1}(t+1), P_{k \mid k-1}(t+1)] = \begin{cases} [\hat{x}_{k-1}(t+1), P_{k-1}(t+1)] \\ \qquad \text{if } \gamma_{k,t+1}^v = 1 \\ [\hat{x}_{k \mid k-1}(t), P_{k \mid k-1}(t)] \\ \qquad \text{if } (\gamma_{k,t+1}^h, \gamma_{k,t+1}^v) = (1, -1) \\ [\mu_x, P_x] \quad \text{otherwise} \end{cases} \tag{3.4b}$$

$$[\hat{x}_{k \mid k-1}(t+1 \mid t), P_{k \mid k-1}(t+1 \mid t)] = \text{RLS}\,[\hat{x}_{k \mid k-1}(t+1), P_{k \mid k-1}(t+1) \mid I, \hat{x}_k(t \mid t-1), P_k(t \mid t-1)] \tag{3.4c}$$

Smoothing:

$$[\hat{x}_k(t), P_k(t)] = \begin{cases} [\hat{x}_k(t+1), P_k(t+1)] \\ \qquad \text{if } \gamma_{k,t+1}^h = +1 \\ [\hat{x}_{k \mid k-1}(t \mid t-1), P_{k \mid k-1}(t \mid t-1)] \\ \qquad \text{if } \gamma_{k,t+1}^h = -1. \end{cases} \tag{3.4d}$$

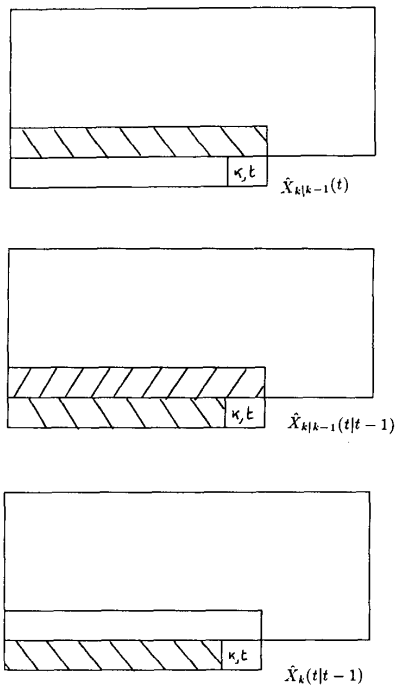


Fig. 1. Regions of support for definition (3.3).

Proof: Recursion a) is straightforward from simple 1-D considerations, and the fact

$$E[X_k(t) \mid y_k(t-1), \cdots, y_k(0), \lambda] = E[X_k(t) \mid y_k(t-1), \cdots, y_k(t-t_s)]$$

with the elements $(k, t), \cdots, (k, t-t_s)$ having all the same label.

Recursion b) comes from the fact that if $\gamma_{k,t+1}^v = +1$ (no vertical edge) then the elements $(k, t+1)$, $(k-1, t+1)$ have the same label and $\hat{x}_{k \mid k-1}(t+1) = \hat{x}_{k-1}(t+1)$ with the same error covariance matrices. If $\gamma_{k,t+1}^v = -1$ (vertical edge) then either $\hat{x}_{k \mid k-1}(t+1) = \hat{x}_{k \mid k-1}(t)$ (if no horizontal edge) or it is reinitialized.

For recursion c) consider the fact that, given the estimates $\hat{x}_{k-1}(0), \cdots, \hat{x}_{k-1}(t)$ and the labels λ, we can write

$$\begin{aligned} x_k(t) &= \hat{x}_{k \mid t-1}(t) + v_1 \\ \hat{x}_k(t \mid t-1) &= x_k(t) + v_2 \end{aligned}$$

where v_1, v_2 are realizations of Gaussian, zero mean, independent random variables V_1, V_2 with $E[V_1 V_2^T] = 0$ and covaraiance matrices $P_{k \mid k-1}(t)$, $P_k(t \mid t-1)$, respectively. Therefore (3.4c) follows from standard results on Kalman updating [21].

Finally, recursion d) comes from the fact that, on line k, the process X is constant within the regions and so is its smooth estimate $\hat{x}_k$, reinitialized at each of the transitions between adjacent regions. ■

Observation: Apart from the complication due to the fact that the process X changes value from region to region, recursions (3.4) are extensions to the 2-D case of the standard recursive least squares estimation algorithm. In fact, since a) the process X is assumed to be piecewise constant (say constant within a region $S_0 \subset S$) with distribution (2.5), and b) the observations are as in (2.2), the MAP estimate $\hat{x}_0$ on S_0 (in this case $\hat{x}_{k|k-1}(t+1 \mid t)$) can be obtained from MAP estimates $\hat{x}_1$, $\hat{x}_2$ on nonoverlapping partitions $S_0 = S_1 \cup S_2$ (respectively, $\hat{x}_{k|k-1}(t)$, $\hat{x}_k(t \mid t-1)$ in our case) as

$$\hat{x}_0 = \hat{x}_1 + P_1[P_1 + P_2]^{-1}(\hat{x}_2 - \hat{x}_1)$$

$$P_0 = P_1 - P_1[P_1 + P_2]^{-1}P_1$$

P_i being the respective covariance matrices. This is valid regardless of the dimensionality of the problem (1-D, 2-D, $\cdots$).

The estimates in the previous theorem can be used to compute the probability terms $\mathcal{P}(y \mid \lambda)$ in (3.1). In particular, we can show the following:

Theorem: Given the observations y, for any realization λ of the label process Λ we can compute $\mathcal{P}(y \mid \lambda)$ as

$$\ln \mathcal{P}(y \mid \lambda) = -\sum_{k,t=1}^{N-1} \ln\left(c_{k,t}^T P_{k|k-1}(t \mid t-1)c_{k,t} + \sigma^2\right) - \frac{\left|y_k(t) - c_{k,t}^T \hat{x}_{k|k-1}(t \mid t-1)\right|^2}{c_{k,t}^T P_{k|k-1}(t \mid t-1)c_{k,t} + \sigma^2} + C \tag{3.5}$$

with $\hat{x}_{k|k-1}(t \mid t-1)$, $P_{k|k-1}(t \mid t-1)$ as in (3.4), and C a constant.

Proof: First notice that

$$\ln \mathcal{P}(y \mid \lambda) = -\sum_{k=0}^{N-1} \ln P(y_k \mid y_{k-1}, \cdots, y_0, \lambda) \tag{3.6}$$

and consider two adjacent rows k, $k-1$. Since $y_k(t) = c_{k,t}^T x_k(t) + w_k(t)$ and the distribution $\mathcal{P}(x_k(t) \mid y_{k-1}, \cdots, y_0, \lambda)$ depends on the estimates $\{\hat{x}_{k-1}(\tau), 0 \le \tau < N\}$ only, we can write

$$\mathcal{P}(y_k \mid y_{k-1}, \cdots, y_0, \lambda) = \mathcal{P}(y_k(0), \cdots, y_k(N-1) \mid \hat{x}_{k-1}(0), \cdots, \hat{x}_{k-1}(N-1), \lambda) \tag{3.7}$$

with $\hat{x}_{k-1}(t)$ as in (3.3). From standard Bayes factorization we can write the recursion

$$\mathcal{P}(y_k(t), \cdots, y_k(0) \mid \hat{x}_{k-1}(t), \cdots, \hat{x}_{k-1}(0), \lambda) = \mathcal{P}(y_k(t) \mid y_k(t-1), \cdots, y_k(0), x_{k-1}(t), \cdots, x_{k-1}(0), \lambda) \times \frac{\mathcal{P}(x_{k-1}(t) \mid y_k(t-1), \cdots, y_k(0), x_{k-1}(t-1), \cdots, x_{k-1}(0), \lambda)}{\mathcal{P}(\hat{x}_{k-1}(t) \mid \hat{x}_{k-1}(t-1), \cdots, \hat{x}_{k-1}(0), \lambda)} \times \mathcal{P}(y_k(t-1), \cdots, y_k(0) \mid \hat{x}_{k-1}(t-1), \cdots, \hat{x}_{k-1}(0), \lambda). \tag{3.8}$$

Now: a) by definition (3.3) the random variable $\hat{X}_{k-1}(t)$ is independent on the observations on line k, therefore the middle factor of the right-hand side (RHS) of (3.8) is always equal to one; b) from the Gaussian assumptions on the parameters X and the noise W the first term on the RHS of (3.8) represents a Gaussian distribution $\mathcal{N}(c_{k,t}^T \hat{x}_{k|k-1}(t \mid t-1), c_{k,t}^T P_{k|k-1}(t \mid t-1)c_{k,t} + \sigma^2)$ with $\hat{x}$, P as in the previous lemma. The result then follows easily. ■

From the recursions in the lemma it is evident that the *a posteriori* probability $\ln \mathcal{P}(\lambda \mid y)$ depends on the transitions $\gamma = \{(\gamma_s^h, \gamma_s^v), s \in S\}$ associated with the labels λ. In this respect equations (3.2), (3.4), and (3.5) define a mapping $(\gamma(\lambda), y) \to L(\gamma(\lambda) \mid y)$ with L the likelihood

$$L(\gamma(\lambda) \mid y) = \ln \mathcal{P}(\lambda \mid y) + \ln \mathcal{P}(\lambda) + f(y) \tag{3.9}$$

and $f(y) = -\ln \mathcal{P}(y)$ indicating terms independent on λ.

An optimal MAP estimation of the labels λ (or its transitions $\gamma(\lambda)$) would require an exhaustive search over all possible transitions $\gamma(\lambda) \in \{-1, +1\}^{N \times N}$ for the one which maximizes the likelihood function L in (3.9). Since this is an unfeasible operation, suboptimal techniques have to be considered. In the next section a sequential, line-by-line, estimation procedure is presented.

A. Suboptimal Estimation

From (3.2), (3.4), (3.5) we see that the likelihood L can be written as

$$L(\gamma \mid y) = \sum_{(k,t) \in S} l(\gamma_{k,t}, \gamma_{k,t-1} \cdots \mid y) + C$$

where

$$l = -\ln\left(c_{k,t}^T P_{k|k-1}(t \mid t-1)c_{k,t} + \sigma^2\right) - \frac{\left|y_k(t) - c_{k,t}^T \hat{x}_{k|k-1}(t \mid t-1)\right|^2}{c_{k,t}^T P_{k|k-1}(t \mid t-1)c_{k,t} + \sigma^2} + \beta(\gamma_{k,t}^h + \gamma_{k,t}^v).$$

The dependence on $\gamma_{k,t}, \gamma_{k,t-1}, \cdots$ is implied in the estimates $\hat{x}_{k|k-1}(t \mid t-1)$ and respective covariance matrices $P_{k|k-1}(t \mid t-1)$. A suboptimal estimation of $\gamma_{k,t}$ can be obtained by the maximization

$$\hat{\gamma}_{k,t} = \operatorname{argmax}_{\gamma_{k,t}} l(\gamma_{k,t}, \hat{\gamma}_{k,t-1}, \cdots \mid y)$$

with the maximum over the set of all possible transitions $(\gamma_{k,t}^h, \gamma_{k,t}^v) \in \{-1, +1\}^2$ which are consistent with the previous estimates $(\hat{\gamma}_s^h, \hat{\gamma}_s^v)$.

IV. One-Dimensional Estimation

The arguments presented for the 2-D case can be easily extended to the simpler 1-D case. In particular, given the labels λ with transitions $\gamma(\lambda)$ define the random variable

$$\hat{X}(t \mid t-1) = E[X(t) \mid Y(t-1) \cdots Y(0), \Lambda].$$

It is now easy to show that we can write the updating as

$$[\hat{x}_{t+1}, P_{t+1}] = \begin{cases} \text{RLS}\,[\hat{x}_t, P_t \mid c_t^T, y_t, \sigma^2 I] & \text{if } \gamma_{t+1} = +1 \\ [\mu_x, P_x) & \text{otherwise.} \end{cases}$$

From this, the likelihood function becomes

$$l(\gamma \mid y) = \sum_{0 \le t < N} -\ln \sigma_t(\gamma) - \frac{\left| y_t - c_t^T \hat{x}_t(\gamma) \right|^2}{2\sigma_t(\gamma)^2} + \beta \gamma_t$$

with $\sigma_t(\gamma) = \sigma^2 + c_t^T P_t(\gamma) c_t$ and the estimation of γ follows the same arguments presented in the previous section.

V. Choice of the Parameter in the MRF

A crucial point in the effective application of the filtering algorithm presented above is the choice of the model parameter appearing in the likelihood (2.8).

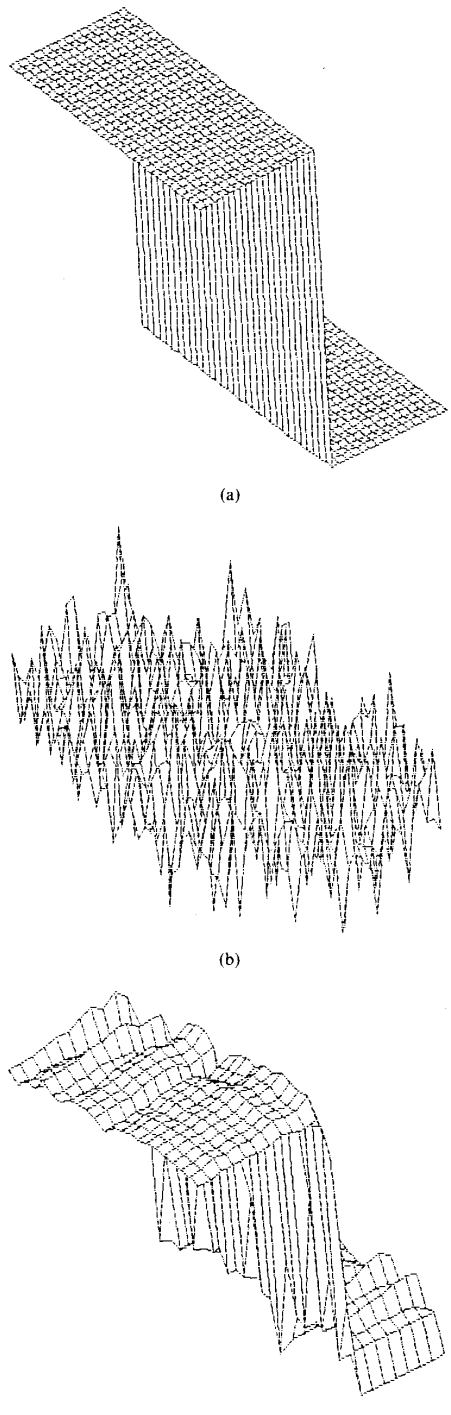
In the framework of this paper and in most previous works along similar lines [9], [20], [23] the Markov random field models the original data, and therefore is a characteristic of the class of data we want to reconstruct. Procedures to estimate the model parameters account for the "coding scheme" of Besag [4] where subsets of data are used to determine conditional probabilities. For the case of MRF as models of textures a more rigorous technique by Derin and Cole [24] has been effectively used. A different approach is taken by Geman and Geman's [7] simulated annealing, where stochastic relaxation is combined with a monotonic increasing parameter.

In spite of the above mentioned results in the model parameter estimation, the MRF parameter β in (2.3) is set mostly by trial and error until a reasonable filtering is reached. Furthermore, extensive simulation results in this and other related works ([20] for example) have shown that satisfactory values for β are strongly influenced by the signal-to-noise ratio of the data.

A better understanding of the role of MRF in estimation can be reached by relating the likelihood function (2.8) to the general form of a penalized likelihood as

$$-\sum_{t \in L} \left(y_t - f_t(x_t) \right)^2 + \lambda \phi(x) \qquad (5.1)$$

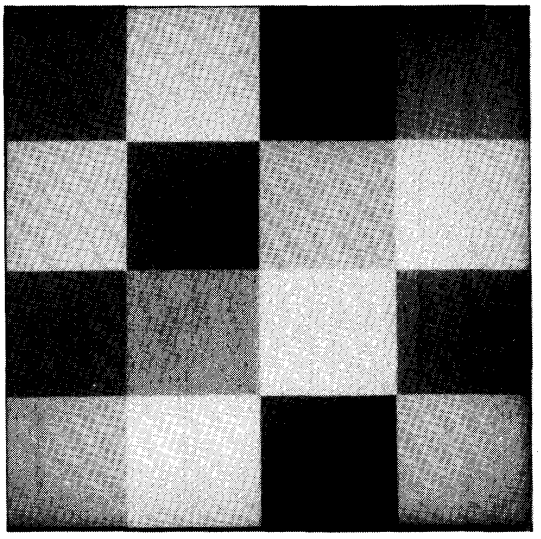
where $\phi(x) \ge 0$ for all x is a function chosen so to reward "smooth" realizations. The parameter λ, often called the *hyperparameter* in the Bayesian literature [25], is a factor weighting deviation from the observations y (left most term in (5.1)) and smoothness $\phi(x)$. A great amount of work has appeared on estimating data from penalized likelihood (see, for example, Kitagawa and Gersch [15] with references).



(a) (b) (c)

Fig. 2. (a) Original 32 × 32 data. (b) Noisy data. (c) Estimated data.

However, a fundamental difference exists in the interpretation given to the parameter λ in (5.1). By regarding the rightmost term in (5.1) as a "smoothness prior" the



(a)

70	140	65	110
180	60	90	190
50	120	175	55
100	200	75	130

(b)

Fig. 3. (a) Original chessboard data (128 × 128). (b) Intensity levels (black = 0, white = 255).

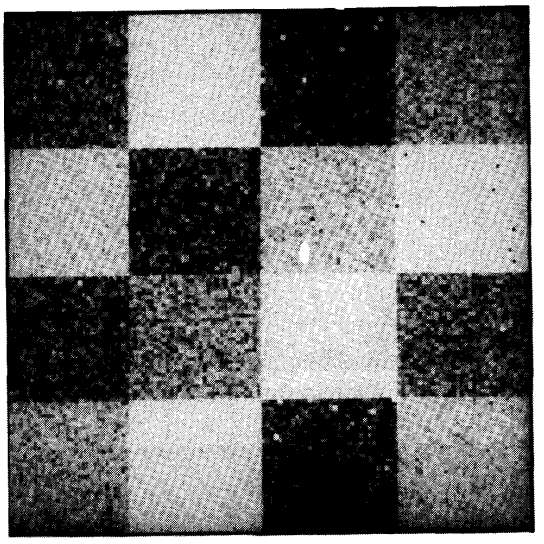


Fig. 4. Noisy data: $\sigma = 20.0$.

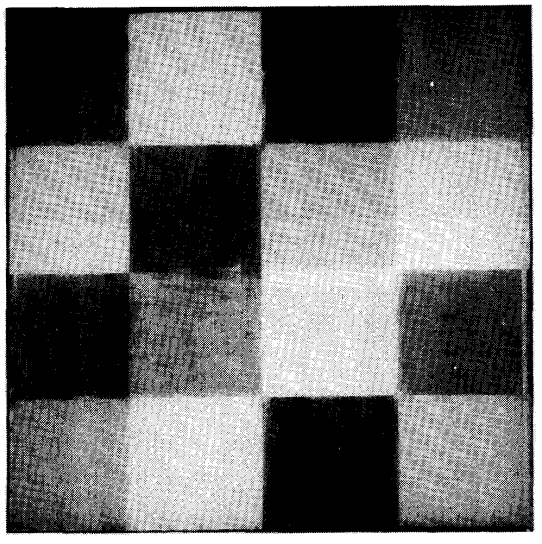


Fig. 5. Filtered data.

hyperparameter λ is disengaged from being numerically set *a priori* and becomes, as Shiller states it in [26] (also in Kitagawa and Gersch [15]) the measure of stiffness of a "flexible rule."

In this respect, optimal values of the parameters λ (and therefore β) can be computed on line (as in [15]) by including this parameter in the likelihood function. For example in Gull and Daniel [27] the parameter λ becomes a Lagrange multiplier in the maximization of the smoothing function $\phi(x)$ subject to the known standard deviation of the noise.

Research is presently conducted on this problem. Numerous simulation results strongly indicate that optimal values of the parameter β do not change for data subject to the same degradation.

VI. Application Examples

The estimation technique presented in the previous sections has been tested on both synthetic and underwater images having compact regions of different intensity levels. This case corresponds to the assumed model (2.1), (2.2) with $X_s \in R$ representing the actual (piecewise constant) intensity levels, and $Y_s = X_s + W_s$ the noisy observations (i.e., $c_s = 1$ for all $s \in S$).

On Fig. 2 an example of application on a two-dimensional set of data (32 × 32 points) is shown. The data set is characterized by two different values of 50 and 100 units, shown in Fig. 2(a). The same set of data with Gaussian noise added ($\sigma = 25.0$ units) is shown in Fig. 2(b), with the estimated data in Fig. 2(c).

Images investigated (128 × 128 pixels) have been processed on a VAX 11/785 using standard FORTRAN 77 with a computing time of 45 s.

The examples shown include a chessboard of 16 different intensity levels, corrupted by a random Gaussian noise (Fig. 3-5), and an underwater scene (Figs. 6-8).

In the case of the chessboard, the improvement on the SNR is in excess of 6 dB. The significant feature, however, is the fact that this improvement has been obtained without smoothing the edges of the image.

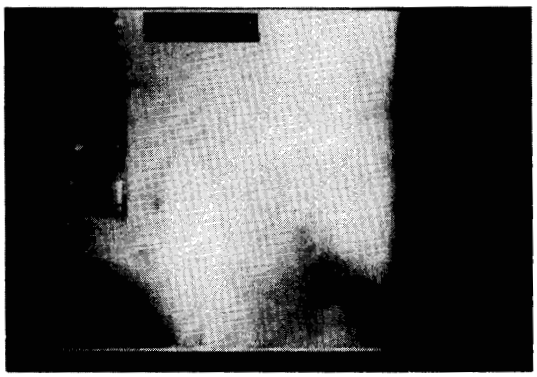


Fig. 6. Original underwater data (128 × 128).

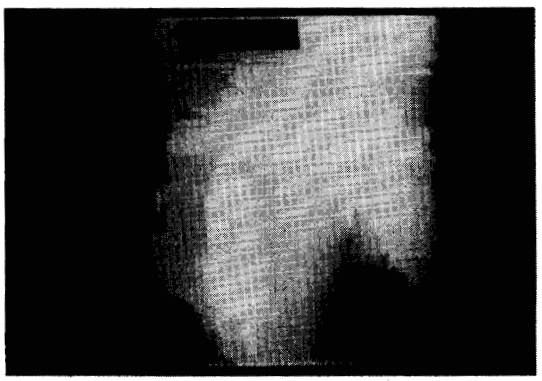


Fig. 7. Filtered underwater data.

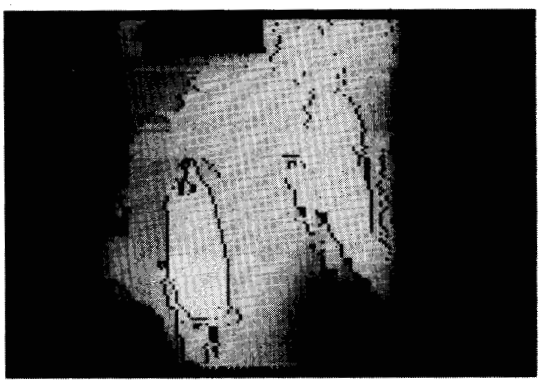


Fig. 8. Filtered underwater data showing estimated transitions.

The underwater scenery data (Fig. 6) is naturally corrupted by disturbances due to particles and sand, as well as the glare caused by lighting equipment. The effect of filtering is shown in Figs. 7 and 8, which highlights the different regions of the pictures. In particular, Fig. 8 shows the detected transitions.

A realization of a 1-D random process with a piecewise

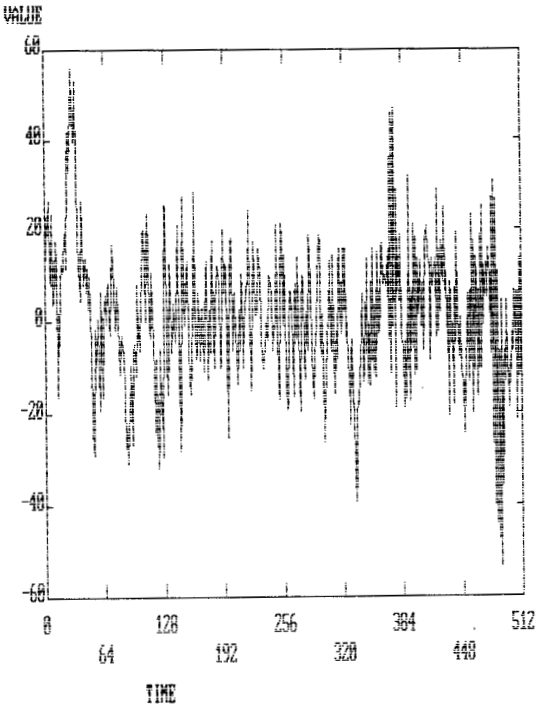


Fig. 9. Original autoregressive data.

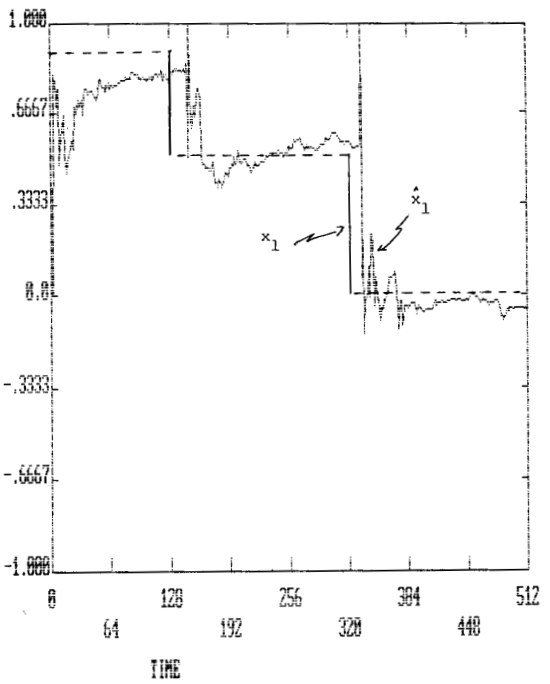


Fig. 10. True (x_1) and estimated ($\hat{x}_1$) AR parameter.

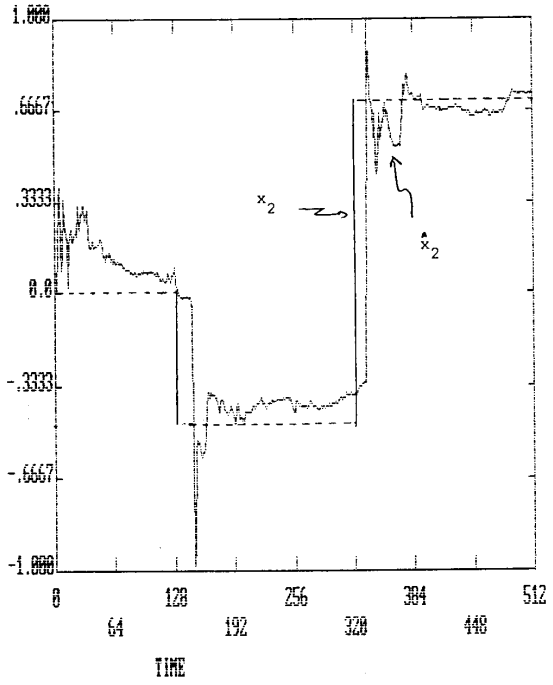


Fig. 11. True (x_2) and estimated ($\hat{x}_2$) AR parameter.

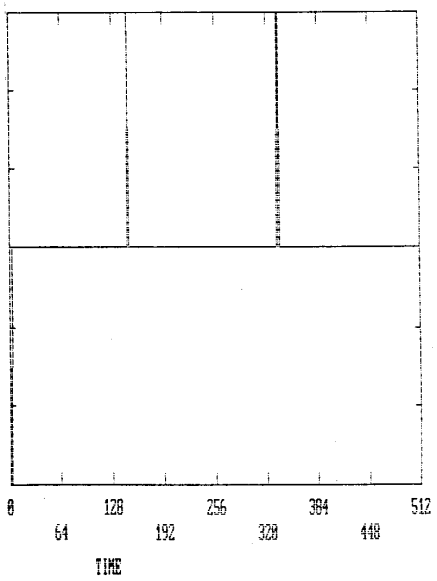


Fig. 12. Estimated transitions.

constant autoregressive model is shown in Fig. 9. A second-order model is assumed with parameters x_1 and x_2 shown in Fig. 10 and 11. The estimates $\hat{x}_1$ and $\hat{x}_2$ are also shown, with the estimated transition vector in Fig. 12.

VII. Conclusions

An algorithm for estimation and segmentation of data which are piecewise described by autoregressive models has been presented. The particular feature of the algorithm is that it searches for the best sequence of edges in order to maximize a suitable likelihood function.

The issues of on line estimation of the Markov random fields parameters and efficient implementations using parallel processing techniques are under consideration.

References

[1] K. Abend, T. Harley, and L. N. Kanal, "Classification of binary random patterns," *IEEE Trans. Inform. Theory*, vol. IT-11, 1965.

[2] L. N. Kanal, "Markov mesh models," in *Image Modelling*. New York: Academic, 1980.

[3] C. W. Therrien, "The analysis of multichannel two-dimensional random signals," NPS Tech. Rep. Elec. Comput. Eng. Dep., Naval Postgraduate School, Monterey, CA, Oct. 1986.

[4] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. B*, vol. 36, 1974.

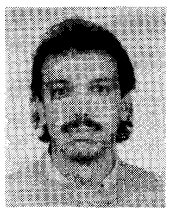
[5] H. Derin, M. Elliott, R. Cristi, and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-6, Nov. 1984.

[6] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-9, Jan. 1987.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-6, Nov. 1984.

[8] R. Cristi, M. Shridhar, and M. V. Prasada Rao, "Segmentation of multilevel images using Gibbs distribution," in *Proc. IEEE ICASSP*, 1985.

[9] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, no. 3, 1986.
[10] C. W. Therrien, "An estimation theoretic approach to terrain image segmentation," *Comput. Vision, Graph., Image Processing*, vol. 22, June 1983.
[11] J. W. Woods, S. Dravida, and R. Mediavilla, "Image estimation using doubly stochastic Gaussian random field models," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-9, Mar. 1987.
[12] G. A. Ackerson and K. S. Fu, "On state estimation in switching environments," *IEEE Trans. Automat. Contr.*, vol. AC-15, Feb. 1970.
[13] J. K. Tugnait and A. H. Hadad, "A detection estimation scheme for state estimation in switching environments," *Automatica*, vol. 15, 1979.
[14] M. Basseville, "Detecting changes in signals and systems—a survey," *Automatica*, vol. 24, May 1988.
[15] G. Kitagawa and W. Gersch, "A smoothness priors time—varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans. Automat. Contr.*, vol. AC-30, Jan. 1985.
[16] E. T. Whittaker, "On a new method of graduation," *Proc. Edinborough Math. Soc.*, vol. 41, 1923.
[17] R. Kinderman and J. L. Snell, *Markov Random Fields and their Applications*. Providence, RI: American Math Society, 1980.
[18] E. Ising, *Z. Physik*, vol. 31, 1925.
[19] H. Derin, "Estimating components of univariate Gaussian mixtures using Prony's method," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-9, Jan. 1987.
[20] H. Elliott, H. Derin, R. Cristi, and D. Geman, "Application of the Gibbs distribution to image segmentation," in *Proc. 1984 IEEE ASSP* (San Diego, CA), Mar. 1984.
[21] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[22] R. Cristi, "Application of Markov random fields to smoothing and segmentation of noisy pictures," in *Proc. ICASSP 1988* (New York, NY), Apr. 1988.
[23] F. S. Cohen and D. B. Cooper, "Simple parallel hierarchical and relaxation algorithms for segmentating noncausal Markovian random fields," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. PAMI-9, Mar. 1987.
[24] M. Derin and W. S. Cole, "Segmentation of textured images using Gibbs random fields," *Comput. Vision, Graph., Image Processing*, vol. 35, 1986.
[25] D. V. Lindsey and A. F. Smith, "Bayes estimates for the linear model," *J. Roy. Stat. Soc. B*, vol. 34, 1972.
[26] R. Shiller, "A distributed lag estimator derived from smoothness priors," *Econometrica*, vol. 41, 1973.
[27] S. F. Gull and G. J. Daniel, "Image reconstruction from incomplete and noisy data," *Nature*, 1978.



Roberto Cristi (S'82–M'82) was born in Ferrara, Italy, on September 3, 1952. He graduated maxima cum laude in electrical engineering from the University of Padua, Italy, in 1977, received the M.Sc. degree from UMIST, Manchester, and Brunel University, Uxbridge, the U.K., and the Ph.D. degree from the University of Massachusetts—Amherst in 1983, all in electrical engineering.

He has been with the University of Padua, Italy; the General Electric Company, Leicester, U.K.; and the University of Michigan, Dearborn. Also he has been a consultant for the Ford Motor Company. At the present time he is an Associate Professor in the Electrical and Computer Engineering Department of the Naval Postgraduate School in Monterey, CA. His professional interests are in adaptive control, image processing, and signal processing.

Dr. Cristi is a member of Tau Beta Pi and Sigma Chi.