Unsupervised Image Restoration and Edge Location Using Compound Gauss–Markov Random Fields and the MDL Principle

Mário A. T. Figueiredo, *Member, IEEE,* and José M. N. Leitão, *Member, IEEE*

Abstract—Discontinuity-preserving Bayesian image restoration typically involves two Markov random fields: one representing the image intensities/gray levels to be recovered and another one signaling discontinuities/edges to be preserved. The usual strategy is to perform joint *maximum a posteriori* (MAP) estimation of the image and its edges, which requires the specification of priors for both fields. In this paper, instead of taking an edge prior, we interpret discontinuities (in fact their locations) as deterministic unknown parameters of the compound Gauss–Markov random field (CGMRF), which is assumed to model the intensities. This strategy should allow inferring the discontinuity locations directly from the image with no further assumptions. However, an additional problem emerges: The number of parameters (edges) is unknown. To deal with it, we invoke the *minimum description length* (MDL) principle; according to MDL, the best edge configuration is the one that allows the shortest description of the image and its edges. Taking the other model parameters (noise and CGMRF variances) also as unknown, we propose a new unsupervised discontinuity-preserving image restoration criterion. Implementation is carried out by a continuation-type iterative algorithm which provides estimates of the number of discontinuities, their locations, the noise variance, the original image variance, and the original image itself (restored image). Experimental results with real and synthetic images are reported.

I. INTRODUCTION

A. Discontinuity-Preserving Image Restoration and Compound Gauss–Markov Random Fields

EARLIER approaches to image restoration/reconstruction used continuity and smoothness restrictions as a means of overcoming the ill-posed nature of the problem [1], [5], [6], [51] (see also [17] and the references therein). Independently of their formal setting (e.g., statistical estimation or regularization theory), these methods typically looked for compromises between fidelity to the observed data and global smoothness; possible discontinuities in the original images were ignored and thus not preserved by the restoration/reconstruction procedures. However, discontinuities (edges) are key visual features [36]; this is testified to by the amount of research dedicated

to the edge detection problem, ever since the early days of digital image processing and computer vision, [5], [25], [36], [42], [52].

The incorporation of discontinuity[1] detection and preservation into restoration/reconstruction procedures has been proposed by several authors. This perspective, in which image restoration and discontinuity detection are jointly sought, is different from edge detection as a separate and autonomous problem. Some fundamental references that adopt statistical tools are those of Geman and Geman [21], Marroquin, Mitter and Poggio [37], Geiger and Girosi [20], Jeng and Woods [26], [27], and Leclerc [34]; in a nonstatistical framework, the works of Blake and Zisserman [7], Terzopoulos [50], Grimson and Pavlidis [22], and Mumford and Shah [40] are also basic references. Independently of their theoretical foundations, the majority of these formulations are modifications of previous techniques; typically, *everywhere* (continuity/smoothness) constraints are weakened becoming *almost-everywhere* constraints, the exceptions being at the locations of discontinuities. A common feature of most approaches to discontinuity-preserving restoration/reconstruction is that they lead to very difficult optimization problems; some authors have advocated the use of stochastic techniques (e.g., [21], [37]) while others have proposed (faster) suboptimal deterministic schemes (e.g., [7], [15], [20], [34], [56]). In addition to the optimization issues, and arguably even more important than these, difficulties in dealing with the involved parameters also arise.

Building on previous work on Markov random fields (MRF's) for Bayesian image restoration (e.g., [6], [21], [54]), Jeng and Woods have introduced the compound Gauss–Markov random field (CGMRF) model that allows for edge-preserving Bayesian restoration with a continuous (Gauss–Markov) *a priori* statistical model for the intensity field together with discrete (binary) hidden edge variables signaling discontinuities [26], [27]. The commonly adopted strategy to image restoration using CGMRF's is to follow Geman and Geman's approach [21]: The edge variables are interpreted as a random field, with some prior, and joint *maximum a posteriori* (MAP) estimation is performed. Typical priors try to penalize edges in a way that depends on

The authors are with the Instituto de Telecomunicaçoes, and Departamento de Engenharia Electrotécnica e de Computadores. Instituto Superior Técnico, 1096 Lisboa Codex, Portugal (e-mail: mtf@it.ist.utl.pt).

---

[1] When dealing with images, or visual surfaces, defined on discrete domains, the term *discontinuity* must not be understood in its common mathematical sense but rather as *abrupt* change.

parameters defined in a more or less *ad hoc* manner [21], [26], [27], [48], [49], [56]; consequently, the amount of signaled edges in these methods is highly dependent on the definition of these priors. Like most edge-preserving restoration techniques, the use of CGMRF's leads to a hard optimization problem (for which stochastic [26] and deterministic [48], [49], [56] algorithms have been applied) and raises the issue of how to define the involved parameters.

B. Proposed Approach

1) Locating Discontinuities as a Parameter Estimation Problem: In CGMRF's, the edges play the role of parameters; in fact, if the discontinuities are somehow known, we have a fully specified CGMRF model. Accordingly, locating the edges of a CGMRF is equivalent to estimating its parameters based on a (possibly noisy) observed realization of it.

The discontinuities of a CGMRF are best described by their locations, since their number is usually by far smaller than the number of their possible locations; in other words, it is more natural to write down the locations of a certain discontinuity set than to specify, for each possible location, the presence or absence of edge (i.e., a binary variable). Therefore, we take discontinuity locations as the parameters of the intensity field prior CGMRF; this is a change of perspective from the classical formulation of Geman and Geman [21] where the binary variables are themselves elements of a random field to be estimated. Locating discontinuities is then formulated as a parameter estimation problem. The advantage of this perspective is that it allows inferring the edges directly from the observed image without assuming any Bayesian prior. However, it raises a problem: The number of parameters (discontinuities/edges) is not known *a priori*.

2) The Minimum Description Length Approach: Since the goal is to estimate an unknown number of parameters (discontinuities), it is impossible to use the *maximum likelihood* (ML) criterion. Here, we resort to Rissanen's approach: "Take the parameters so that the model they define allows the shortest representation of the data;" formally, the *representation lengths* are Shannon code lengths, and the corresponding criterion is the *minimum description length* (MDL) principle [43], [44], [46]. It includes ML as a special case, gives it a coding theoretical meaning, and, more importantly, it can be used when the number of parameters is unknown, whereas ML can not [43], [44], [46]. According to the MDL criterion, the best edge configuration is the one that allows the shortest description of the image (including the edge locations).

3) Unsupervised Estimation: In image restoration/reconstruction problems, only a noisy version of the original image is observed. However, the optimization problem resulting from the proposed criterion does depend on the original image, on the noise variance, and (in addition to the number of discontinuities and their locations) on the global parameters of the original CGMRF, which are all unknown. Hence, these quantities will also be considered as unknowns to be estimated from the observed image. What we propose is an MDL-based *unsupervised* (also called *adaptive*, or *blind*)

discontinuity-preserving restoration criterion, built upon a CGMRF model.

4) Implementation of the Proposed Criterion: The resulting optimization problem has to be simultaneously solved with respect to all the above mentioned unknowns. To this end, we propose an algorithm with two nested loops. The inner loop is basically an iterative restoration/parameter-estimation scheme. This type of procedure has been applied by several authors in different areas, under different names; e.g., Mendel's *block component search algorithm* [38], the *adaptive segmentation algorithm* of Lakshmanan and Derin [31], [53], and Besag's adaptive version of the *iterated conditional modes* (ICM) scheme [6] (which we have also previously used [14]). It is also conceptually related to (and in special cases coincides with) the *expectation-maximization* (EM) algorithm of Dempster *et al.* [10], recently used by several authors for blind image segmentation and restoration problems [30], [32], [33], [57]. Under certain conditions, this scheme is shown to converge to a *partial optimal solution* (POS) [31], [53]. A POS corresponds to a weaker optimality criterion, when compared to the original one.

The outer loop of the algorithm is a continuation-type scheme similar to the *simulated tearing* algorithm we have proposed in [15]; the function to be minimized is embedded in a family of functions (depending on a control parameter) of which the first member is easy to minimize; this minimum is then tracked along the family (by varying the control parameter) until the desired solution is reached. By including the inner loop into the outer continuation-type loop, a sequence of POS's of increasing quality (i.e., decreasing values of the original objective function) is obtained.

The *mean field annealing* (MFA) algorithm is another continuation scheme that could be considered for this application [20], [56]; however, MFA is heavily supported on the interpretation of the edge variables as elements of a random field, which we are abandoning.

C. Previous Related Work

Leclerc pioneered in using the MDL principle for image partitioning [34]; the main difference between his work and ours is that we model the true image as a sample of an MRF while in [34] it is deterministically modeled as piecewise polynomial. Keeler uses MDL in an MRF-type context pursuing a different goal: to build "*complexity*"-based priors for discontinuity configurations [29]. Recently, MDL and related ideas have seen several applications in the fields of image processing and computer vision: segmentation [9], [29], [32]; line and curve detection [47]; contour estimation [18]; motion and displacement segmentation and/or estimation [11], [23], [59]; and shape description [35].

To the authors' knowledge, not much has been done in unsupervised restoration using noncausal CGMRF's. The technique developed in [49] (and also used in [56]) assumes known noise variance, fixed edge penalty, and availability of the original image. In [57], noise and also an unknown blur are considered; however, the parameters of the original CGMRF and of the edge field are *a priori* fixed.

Nadabar and Jain [41] resort to *computer aided design* (CAD) models of the objects in the scene (assumed available) to estimate parameters of the discontinuity field; this is computationally very demanding and not applicable to image restoration problems.

For other MRF-type models, parameter estimation and/or unsupervised estimation has been considered in, e.g., [3], [12], [14], [28], [30]–[32], [53], and [58]. See also [2] for an elucidative review of some joint image-restoration/parameter-estimation methods.

D. Paper Overview

In Section II, the set of tools underlying our work is described; in particular, compound Gauss–Markov random fields and the MDL principle are briefly reviewed. Section III formally introduces the proposed approach. The algorithm designed to implement the estimation criterion is presented in Section IV. Finally, Section V reports experimental results obtained with synthetic and natural images, and Section VI contains some concluding remarks.

II. UNDERLYING TOOLS

A. Compound Gauss–Markov Random Fields (CGMRF)

Let the image $\boldsymbol{x} = \{x_{ij} \in \mathbb{R}: i = 1, 2, \cdots, M; j = 1, 2, \cdots, N\}$, containing $M \times N$ real valued pixels, be a sample of a GMRF $\boldsymbol{X} = \{X_{ij}: i = 1, 2, \cdots, M; j = 1, 2, \cdots, N\}$ defined on an $M \times N$ lattice. The probability density function (pdf) of $\boldsymbol{X}$ is

$$p(\boldsymbol{x}) = \frac{|\boldsymbol{A}|^{1/2}}{(2\pi)^{(MN)/2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}\right\} \qquad (1)$$

where $\boldsymbol{x}$ here stands for a vector containing the lexicographically ordered pixel values, $\boldsymbol{A}$ is the inverse of the covariance matrix (termed *potential matrix* [39]), and $|\boldsymbol{A}|$ its determinant [6]. Since we are in the presence of a Markov field, the conditional probability densities verify

$$p(x_{ij}|\{x_{kl}, (k,l) \neq (i,j)\}) = p(x_{ij}|\{x_{kl}, (k,l) \in N_{ij}\})$$

where N_{ij} is the *neighborhood* of pixel (i,j) [6], [21]. The Gibbs joint pdf of the MRF $\boldsymbol{X}$ can be recognized in (1); for a GMRF, the neighborhood system and the clique potentials are completely determined by the potential matrix [39] (see [6] or [21], for precise definitions).

Proposed by Jeng and Woods [26], [27], (see also [48], [49], [56]) compound GMRF's are simply GMRF's in which the potential matrix is parametrized by a collection of binary (0/1) edge variables. Each of these, when set to one, breaks the direct bond between neighbor pixels. Specifically, let $\boldsymbol{h} = \{h_{ij} \in \{0,1\}, i = 2, 3, \cdots, M; j = 1, 2, \cdots, N\}$ and $\boldsymbol{v} = \{v_{ij} \in \{0,1\}, i = 1, 2, \cdots, M; j = 2, 3, \cdots, N\}$ be the sets of, respectively, horizontal and vertical edge variables. The probability density function of a zero-mean first-order[2]

---

[2]I.e., in which the neighborhood is the set of four nearest neighbors of (i,j): $N_{ij} = \{(i, j-1), (i, j+1), (i-1, j), (i+1, j)\}$; this is the simplest CGMRF. Higher order models involve larger neighborhoods; here, without loss of generality, we consider the first-order case for which $\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})$ is block tridiagonal with tridiagonal blocks [39].

CGMRF $\boldsymbol{X}$, given edge configuration $(\boldsymbol{h}, \boldsymbol{v})$, is (see, e.g., [48], [49])

$$p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})$$
$$= \frac{1}{Z_1(\boldsymbol{h}, \boldsymbol{v})} \exp\left\{-\frac{\mu}{2} \sum_{i=1}^{M} \sum_{j=2}^{N} \omega_v (1 - v_{ij})(x_{ij} - x_{ij-1})^2\right.$$
$$-\frac{\mu}{2} \sum_{i=2}^{M} \sum_{j=1}^{N} \omega_h (1 - h_{ij})(x_{ij} - x_{i-1j})^2$$
$$\left. -\frac{\mu}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} (1 - 2(\omega_v + \omega_h))x_{ij}^2 \right\} \qquad (2)$$

where μ is the "global smoothness" (i.e., inverse of the "global variance") of the CGMRF, while ω_h and ω_v control its relative vertical and horizontal "smoothness". In (2), $Z_1(\boldsymbol{h}, \boldsymbol{v})$ is the normalizing constant called *partition function* [21]. It is clear (in the exponent of (2)) that each difference between neighbor pixel values is quadratically penalized unless the associated line variable is set to one. This is the feature of CGMRF's that makes them suitable priors for edge preserving Bayesian restoration.

In vector notation, (2) can be written as

$$p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v}) = \frac{|\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})|^{1/2}}{(2\pi)^{(MN)/2}} \exp\left\{-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})\boldsymbol{x}\right\} \qquad (3)$$

where the dependence of $\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})$ on parameters μ, ω_v, and ω_h is not explicitly indicated. The factor multiplying the exponential in (3) is the reciprocal of the *partition function* $Z_1(\boldsymbol{h}, \boldsymbol{v})$, which depends on the edge configuration $(\boldsymbol{h}, \boldsymbol{v})$ via the determinant of the potential matrix $\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})$.

Assume that only a noisy version $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ is observed, where $\boldsymbol{n}$ is a sample of a white Gaussian homogeneous noise field of variance σ^2; the probabilistic observation model is then

$$p(\boldsymbol{y}|\boldsymbol{x}) = (2\pi\sigma^2)^{-MN/2}$$
$$\cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{M} \sum_{j=1}^{N} (y_{ij} - x_{ij})^2\right\}. \qquad (4)$$

The goal of image restoration is to estimate $\boldsymbol{x}$ from the observed $\boldsymbol{y}$. Fixing $\boldsymbol{h}$ and $\boldsymbol{v}$, the *a posteriori* probability density function $p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v}, \boldsymbol{y}) \propto p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})p(\boldsymbol{y}|\boldsymbol{x})$, considering (4) and (3), is convex with respect to $\boldsymbol{x}$ and its maximizer (the MAP estimate) is simply

$$\hat{\boldsymbol{x}}_{\mathrm{MAP}}(\boldsymbol{h}, \boldsymbol{v}) = (\sigma^2 \boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v}) + \boldsymbol{I})^{-1}\boldsymbol{y}. \qquad (5)$$

Equation (5) can be solved by any technique suitable for extremely large but highly sparse and structured matrices (see, e.g., [17] and references therein).

The condition $(1 - 2(\omega_v + \omega_h)) > 0$ is sufficient for having $|\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})| \neq 0$ for any $(\boldsymbol{h}, \boldsymbol{v})$, i.e., for that $Z_1(\boldsymbol{h}, \boldsymbol{v}) < \infty$; in isotropic models, the ones herein considered, $\omega_v = \omega_h \equiv \omega$ and the condition is simply $\omega < 1/4$. Of course, anisotropy could be considered; however, since here ω_v and ω_h are assumed known, there is no point in introducing this additional complication. *Weak membrane*-type models lack the $(1 - 4\omega)x_{ij}^2$ term [20]; as a consequence, matrix $\boldsymbol{A}(\boldsymbol{h}, \boldsymbol{v})$ is singular

for many edge configurations. In Bayesian estimation parlance, $p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})$ becomes an *improper* (or *nonnormalizable*) prior [4]. For MAP estimation, this is not a difficulty because $A(\boldsymbol{h}, \boldsymbol{v})$ is added to an identity matrix [see (5)]; however, in parameter estimation problems the partition function (and consequently $|A(\boldsymbol{h}, \boldsymbol{v})|$) must be explicitly used and difficulties emerge.

B. The Usual Approach

As stated above, the common strategy consists in interpreting the edge configuration also as a realization of a random field and to perform joint MAP estimation of $\boldsymbol{x}, \boldsymbol{h}$, and $\boldsymbol{v}$, given $\boldsymbol{y}$,

$$
\begin{aligned}
(\hat{\boldsymbol{x}}, \hat{\boldsymbol{h}}, \hat{\boldsymbol{v}})_{\mathrm{MAP}} &= \arg \max_{\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v}} \{p(\boldsymbol{x}, \boldsymbol{h}, v|\boldsymbol{y})\} \\
&= \arg \max_{\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v}} \{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v})p(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v})\} \quad (6) \\
&= \arg \max_{\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v}} \{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})p(\boldsymbol{h}, \boldsymbol{v})\} \quad (7)
\end{aligned}
$$

where $p(\boldsymbol{h}, \boldsymbol{v})$ is the prior probability function of the edge fields. Going from (6) to (7) relies on the fact that $\boldsymbol{y}$ is only an observation of $\boldsymbol{x}$, i.e. $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v}) = p(\boldsymbol{y}|\boldsymbol{x})$; $\boldsymbol{h}$ and $\boldsymbol{v}$ are termed *hidden fields*.

In general, $p(\boldsymbol{h}, \boldsymbol{v})$ is not explicitly written; instead, a joint prior $p(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v})$ is defined. The most common one (see [20], [21], [26], [27], [48], [56]) is obtained by adding some discontinuity penalty function $U(\boldsymbol{h}, \boldsymbol{v})$ to the exponent of (2); this leads to

$$
\begin{aligned}
&p(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v}) \\
&= \frac{1}{Z_2} \exp \left\{ -\frac{\mu}{2} \sum_{ij} [\omega(1 - v_{ij})(x_{ij} - x_{ij-1})^2 \right. \\
&\quad + \omega(1 - h_{ij})(x_{ij} - x_{i-1j})^2 + (1 - 4\omega)x_{ij}^2] \\
&\quad \left. - U(\boldsymbol{h}, \boldsymbol{v}) \right\}
\end{aligned} \quad (8)
$$

where Z_2 is the partition function, and the summation $\sum_{ij}$ is simply a shorthand (to be used throughout this paper) for the more detailed notation in (2). Inserting (2) and (8) into Bayes law, $p(\boldsymbol{h}, \boldsymbol{v}) = p(\boldsymbol{x}, \boldsymbol{h}, \boldsymbol{v})/p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})$, reveals that (8) implicitly includes the following prior for the edge field:

$$
p(\boldsymbol{h}, \boldsymbol{v}) = \frac{1}{Z_3} Z_1(\boldsymbol{h}, \boldsymbol{v}) \exp\{-U(\boldsymbol{h}, \boldsymbol{v})\}. \quad (9)
$$

The presence of $Z_1(\boldsymbol{h}, \boldsymbol{v})$ makes the meaning of (9) not at all obvious since its dependence on $\boldsymbol{h}$ and $\boldsymbol{v}$ is complex and unclear; as pointed out in [27], this may not even be an MRF. Finally, notice that in the sequel we will not write $p(\boldsymbol{h}, \boldsymbol{v})$ since edges will be considered as deterministic parameters.

C. The Minimum Description Length Principle (MDL)

MDL is an information-theoretical principle proposed by Rissanen that allows the generalization of ML estimation to cases where not only the parameters but also their number are unknown [43], [44], [46]. The ML estimate of a k-dimensional parameter vector $\boldsymbol{\theta}_{(k)}$, given observation $\boldsymbol{x}$, is defined as

$\hat{\boldsymbol{\theta}}_{(k)\mathrm{ML}} = \arg \max \{p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})\}$ (subscript (k) indicates that a vector is k-dimensional). Underlying MDL is a coding theoretic interpretation of ML estimation. If, based on the probabilistic model $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})$, one builds a Shannon-type code, the length[3] of the code word for the observed data $\boldsymbol{x}$ is

$$
L(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}) = -\log p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}) \quad (10)
$$

with rounding effects neglected [8], [45]. Accordingly, given observation $\boldsymbol{x}$, looking for $\hat{\boldsymbol{\theta}}_{(k)\mathrm{ML}}$ is the same as looking for the Shannon-type code in which $\boldsymbol{x}$ has the shortest code word; in fact, from (10)

$$
\hat{\boldsymbol{\theta}}_{(k)\mathrm{ML}} = \arg \max_{\boldsymbol{\theta}_{(k)}} \{p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})\} = \arg \min_{\boldsymbol{\theta}_{(k)}} L(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}). \quad (11)
$$

Of course, it can be argued that only discrete data can have finite code lengths. However, as Rissanen recently noted [45], even the negative log of densities can be seen as code lengths; finite values may be obtained by truncating to finite precision and replacing the densities with the resulting probabilities. Abuse of the term "code length" is convenient and harmless, since the precision itself is not important.

When k is unknown, ML cannot be used; the MDL principle stipulates that one should still look for the shortest description (code length) of the data, which in fact must also include the parameters themselves. The total length of the optimal code for $\boldsymbol{x}$, given k-dimensional $\boldsymbol{\theta}_{(k)}$, is

$$
L(\boldsymbol{x}, \boldsymbol{\theta}_{(k)}, k) = L(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}) + L(\boldsymbol{\theta}_{(k)}) + L(k)
$$

where $L(\boldsymbol{\theta}_{(k)})$ is the code length for a k-dimensional $\boldsymbol{\theta}_{(k)}$, and $L(k)$ is the code length for k itself (usually a constant, independent of k). The MDL estimate of k and $\boldsymbol{\theta}_{(k)}$ is then [after dropping $L(k)$]

$$
(\hat{k}, \hat{\boldsymbol{\theta}}_{(k)})_{\mathrm{MDL}} = \arg \min_{k, \boldsymbol{\theta}_{(k)}} \{-\log p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}) + L(\boldsymbol{\theta}_{(k)})\}. \quad (12)
$$

Notice that, if $L(\boldsymbol{\theta}_{(k)})$ only depends on k, then for fixed k the MDL and ML estimates coincide. As a corollary, the MDL estimate of $\boldsymbol{\theta}_{(k)}$ coincides with its ML estimate given the MDL estimate of k.

III. PROPOSED FORMULATION

A. Discontinuity Locations as Parameters

Examining the CGMRF pdf given by (2) reveals that the discontinuities/edges play the role of (possibly unknown) parameters. This perspective avoids (in fact precludes) the specification of a Bayesian prior for discontinuity estimation, since we will be interpreting them as deterministic (albeit unknown). Notice that the true difficulty consists in estimating the edges; as stated above, if $\boldsymbol{h}$ and $\boldsymbol{v}$ (and the other parameters) are somehow known and fixed, the image estimate is simply given by (5).

In general, the active (i.e., equal to one) edge variables are a small minority; then, it is obviously more efficient to write down their locations than to exhaustively enumerate all

---

[3] In bits, if base 2 logarithms are used, or in *nats*, if natural logarithms are adopted [8].

binary variables. Formally, let $\boldsymbol{\theta}_{(k)} = [\theta_1, \theta_2, \cdots, \theta_k]^T$ be a k-dimensional parameter vector expressing the locations of k discontinuities; $\boldsymbol{\theta}_{(k)}$ is just a compact description of $\boldsymbol{h}$ and $\boldsymbol{v}$, exploiting the fact that only the elements that are equal to one need to be specified. Each component of $\boldsymbol{\theta}_{(k)}$ is a triplet (i, j, b), with $i \in \{1, 2, \cdots, M\}, j \in \{1, 2, \cdots, N\}$, and $b \in \{0, 1\}$, indicating that $h_{ij} = 1$, if $b = 1$, or, that $v_{ij} = 1$, if $b = 0$

$$(h_{ij} = 1) \Leftrightarrow \exists_{\theta_n = (i,j,1)}, \qquad n = 1, 2, \cdots, k$$
$$(v_{ij} = 1) \Leftrightarrow \exists_{\theta_n = (i,j,0)}, \qquad n = 1, 2, \cdots, k.$$

In words, h_{ij} (resp. v_{ij}) is equal to one if and only if there exists one element of $\boldsymbol{\theta}_{(k)}$ which is equal to $(i, j, 1)$ (resp. to $(i, j, 0)$). To clarify this notation, consider as example an edge configuration with $h_{7\,9} = 1, h_{12\,15} = 1, h_{13\,5} = 1, h_{47\,19} = 1, v_{2\,8} = 1, v_{13\,5} = 1, v_{31\,18} = 1$, and all other h_{ij}'s and v_{ij}'s equal to zero; then, $k = 7$ and $\boldsymbol{\theta}_{(7)} = [(7, 9, 1), (12, 15, 1), (13, 5, 1), (47, 19, 1), (2, 8, 0), (13, 5, 0), (31, 18, 0)]^T$. Now, given that $i \in \{1, \cdots, M\}, j \in \{1, \cdots, N\}$, and $b \in \{0, 1\}$, we have

$$L(\boldsymbol{\theta}_{(k)}) = k(\log M + \log N + \log 2) = k\log(2MN). \quad (13)$$

Different description length functions could be obtained if other ways of encoding the discontinuity configuration were used (see, e.g., [24], [25]). For example, chain-coding [19], [25] (as used in [34]) will favor long connected sequences of discontinuities; this would be a direction to explore.

Finally, notice that since $\boldsymbol{\theta}_{(k)}$ contains the same information as the pair $(\boldsymbol{h}, \boldsymbol{v})$, writing $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)})$ is exactly the same as writing $p(\boldsymbol{x}|\boldsymbol{h}, \boldsymbol{v})$, as given by (2) or (3).

B. The Global Parameters μ, σ^2, and ω

Two other parameters play a fundamental role in the problem at hand: the inverse of the CGMRF variance, μ, and the noise variance σ^2. In a realistic setting, these parameters are unknown and must be estimated from the noisy observed data. In the sequel, we explicitly include μ and σ^2, writing $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$ and $p(\boldsymbol{y}|\boldsymbol{x}, \sigma^2)$ to emphasize that they will be considered unknown parameters.

Parameter ω is a *technical* parameter whose only role is preventing the prior $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$ from becoming improper. As long as we keep $\omega \lesssim 1/4$, the results are practically independent of ω. In all examples presented in Section V, we set $\omega = 0.2499$.

C. An MDL Criterion with Incomplete Data

If $\boldsymbol{x}$ was directly observed, MDL could be immediately used by introducing (3) and (13) into (12). Since only $\boldsymbol{y}$, a noisy version of $\boldsymbol{x}$, is observed, we use $p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$ (which is equal to $p(\boldsymbol{y}|\boldsymbol{x}, \sigma^2)p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$), instead of $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$, and interpret the original image $\boldsymbol{x}$ as missing data, in the sense of [10]. Parameters $k, \boldsymbol{\theta}_{(k)}, \mu$, and σ^2 could (in principle) be estimated from the *complete data* set $\{\boldsymbol{x}, \boldsymbol{y}\}$, but only $\boldsymbol{y}$ is observed; $\boldsymbol{x}$ also has to be estimated. The criterion then

becomes

$$\begin{aligned}(\hat{\boldsymbol{x}}, \hat{k}, \widehat{\boldsymbol{\theta}_{(k)}}, \hat{\mu}, \widehat{\sigma^2})_{\text{MDL}} \\ = \arg\min_{(\boldsymbol{x}, k, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)} \{L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)\}\end{aligned} \quad (14)$$

where

$$\begin{aligned}L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2) \\ = -\log p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{\theta}_{(k)}, \mu, \sigma^2) + L(\boldsymbol{\theta}_{(k)}) \\ = -\log p(\boldsymbol{y}|\boldsymbol{x}, \sigma^2) - \log p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu) + k\log(2MN).\end{aligned}$$
$$(15)$$

The lengths $L(\mu)$ and $L(\sigma^2)$ are not included, since they are independent of k; i.e., with respect to μ and σ^2, this is simply an ML criterion. The criterion (14)–(15) can also be interpreted as an MDL version of the (Bayesian) *joint mode approach* to simultaneous image restoration and parameter estimation [2].

D. The Pseudolikelihood Approximation

Before addressing the issue of solving (14), notice the huge difficulty in computing (15). The obstacle lies in $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$, as given by (3), which involves the determinant of a very large matrix $(MN \times MN)$. This is the well-known problem of computing the partition function, arising in parameter estimation in MRF models [3], [6], [31], [53], [57]. Here, we resort to Besag's pseudolikelihood approximation [6] which, omitting the parameters, states that

$$p(\boldsymbol{x}) \simeq \prod_{ij} p(x_{ij}|\{x_{kl}, (k, l) \in N_{ij}\}). \quad (16)$$

Exact formulae proposed in [3], or better approximations, such as the one presented in [55], can only be applied to homogeneous fields; CGMRF's are not homogeneous.

Since X is a Gaussian MRF, the conditional pdf's are also Gaussian

$$p(x_{ij}|\{x_{ij-1}, x_{ij+1}, x_{i-1j}, x_{i+1j}\}, \boldsymbol{\theta}_{(k)}, \mu) = \mathcal{N}(\eta_{ij}, \lambda_{ij}^2) \quad (17)$$

where $\mathcal{N}(\nu, \psi^2)$ denotes a Gauss function with mean ν and variance ψ^2. The local means and variances, η_{ij} and λ_{ij}^2, which obviously depend on $\boldsymbol{x}, \boldsymbol{\theta}_{(k)}$ (i.e. on $\boldsymbol{h}$ and $\boldsymbol{v}$) and μ, are given by

$$\begin{aligned}\eta_{ij}(\boldsymbol{x}, \boldsymbol{\theta}_{(k)}, \mu) = \lambda_{ij}^2(\boldsymbol{\theta}_{(k)})\mu\omega[\overline{h}_{ij}x_{i-1j} + \overline{v}_{ij}x_{ij-1} \\ + \overline{h}_{i+1j}x_{i+1j} + \overline{v}_{ij+1}x_{ij+1}]\end{aligned} \quad (18)$$

where $\overline{h}_{ij} \equiv (1 - h_{ij})$ and $\overline{v}_{ij} \equiv (1 - v_{ij})$, and

$$\lambda_{ij}^2(\boldsymbol{\theta}_{(k)}, \mu) = \frac{1}{\mu[1 - \omega(h_{ij} + h_{i+1j} + v_{ij} + v_{ij+1})]}. \quad (19)$$

Introducing (13), (4), and the pseudolikelihood approximation (16), together with (17), (18) and (19), into (15) leads (after dropping additive constants and common factors) to

$$
\begin{aligned}
L(&\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2) \\
&\simeq 2k \log(2MN) + MN \log \sigma^2 + \sum_{ij} \log \lambda_{ij}^2(\boldsymbol{\theta}_{(k)}, \mu) \\
&+ \sum_{ij} \frac{(x_{ij} - \eta_{ij}(\boldsymbol{x}, \boldsymbol{\theta}_{(k)}, \mu))^2}{\lambda_{ij}^2(\boldsymbol{\theta}_{(k)}, \mu)} + \frac{1}{\sigma^2} \sum_{ij} (x_{ij} - y_{ij})^2.
\end{aligned}
\tag{20}
$$

Some manipulation allows rewriting (20) as

$$
\begin{aligned}
L(&\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2) \\
&\simeq 2k \log(2MN) + MN \log \sigma^2 + \sum_{ij} \log \lambda_{ij}^2(\boldsymbol{\theta}_{(k)}, \mu) \\
&+ 2\mu\omega \sum_{ij} [\overline{v}_{ij}(x_{ij} - x_{ij-1})^2 + \overline{h}_{ij}(x_{ij} - x_{i-1j})^2] \\
&+ \mu \sum_{ij} (1 - 4\omega)x_{ij}^2 + \frac{1}{\sigma^2} \sum_{ij}(x_{ij} - y_{ij})^2.
\end{aligned}
\tag{21}
$$

As a final observation, notice that resorting to the pseudolikelihood approximation is not necessary in one-dimensional (1-D) problems. In 1-D, each discontinuity breaks the potential matrix into two independent subblocks and the determinant can be factored into two subdeterminants; as a result, an exact expression for $p(\boldsymbol{x}|\boldsymbol{\theta}_{(k)}, \mu)$ can be obtained (see [13], for details). In 2-D, only configurations of discontinuities which split the image into two disjoint areas lead to a similar factorization; this, of course, can not be guaranteed in general.

IV. IMPLEMENTATION

A. A Weaker Optimality Criterion

The joint minimization (14) is extremely hard to implement; alternatively, we consider (as in [14]) a weaker criterion, the *partial optimal solution* (POS) [31], [53], which is defined as a joint solution of

$$
\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}_{(k)}}, \hat{\mu}, \widehat{\sigma^2})
\tag{22}
$$

$$
(\hat{k}, \widehat{\boldsymbol{\theta}_{(k)}}) = \arg\min_{(k, \boldsymbol{\theta}_{(k)})} L(\hat{\boldsymbol{x}}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \hat{\mu}, \widehat{\sigma^2})
\tag{23}
$$

$$
\hat{\mu} = \arg\min_{\mu} L(\hat{\boldsymbol{x}}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}_{(k)}}, \mu, \widehat{\sigma^2})
\tag{24}
$$

$$
\widehat{\sigma^2} = \arg\min_{\sigma^2} L(\hat{\boldsymbol{x}}, \boldsymbol{y}, \widehat{\boldsymbol{\theta}_{(k)}}, \hat{\mu}, \sigma^2).
\tag{25}
$$

Although POS's are, of course, not necessarily global minima of the objective function (21), they are almost always local minima exhibiting the following obvious properties:

- $\hat{\boldsymbol{x}}$ is the MAP estimate given $\boldsymbol{y}, \hat{\mu}, \widehat{\sigma^2}, \hat{k}$ and $\widehat{\boldsymbol{\theta}_{(k)}}$;
- $\hat{k}$ and $\widehat{\boldsymbol{\theta}_{(k)}}$ are the MDL estimates given $\hat{\boldsymbol{x}}$ and $\hat{\mu}$;
- $\hat{\mu}$ is the ML estimate given $\hat{\boldsymbol{x}}, \hat{k}$, and $\widehat{\boldsymbol{\theta}_{(k)}}$;
- $\widehat{\sigma^2}$ is the ML estimate given $\hat{\boldsymbol{x}}$ and $\boldsymbol{y}$.

It is easy to solve (22), (24), and (25), since $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$ is convex with respect to $\boldsymbol{x}, 1/\mu$, and σ^2, leading to

$$
\hat{\boldsymbol{x}} = (\widehat{\sigma^2} \boldsymbol{A}(\hat{\boldsymbol{h}}, \hat{\boldsymbol{v}}, \hat{\mu}) + \boldsymbol{I})^{-1} \boldsymbol{y}
\tag{26}
$$

$$
\hat{\mu} = \left(\frac{2}{MN} \sum_{ij} \left[\omega\overline{v}_{ij}(\hat{x}_{ij} - \hat{x}_{ij-1})^2 + \omega\overline{h}_{ij}(\hat{x}_{ij} - \hat{x}_{i-1j})^2 \right. \right.
$$
$$
\left. \left. + \frac{1 - 4\omega}{2}\hat{x}_{ij}^2 \right] \right)^{-1}
\tag{27}
$$

$$
\widehat{\sigma^2} = \frac{1}{MN} \sum_{ij} (\hat{x}_{ij} - y_{ij})^2.
\tag{28}
$$

Recall that $\overline{v}_{ij}$ stands for $(1 - \hat{v}_{ij})$ and $\overline{h}_{ij}$ stands for $(1 - \hat{h}_{ij})$; the pair $(\hat{\boldsymbol{h}}, \hat{\boldsymbol{v}})$ is equivalent to $\boldsymbol{\theta}_{(k)}$.

It is not possible to obtain a closed-form solution for (23). Alternatively, we adopt the ICM algorithm; i.e., $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$ is successively minimized with respect to each line variable. By computing the difference between the objective function for $h_{ij} = 1$ and $h_{ij} = 0$, keeping all other variables constant, we get the decision criterion

$$
h_{ij} = \begin{cases} 1 \Leftarrow (x_{ij} - x_{i-1j})^2 > \frac{1}{2} \log(4(MN)^2 t_{ij}^h) \\ 0 \Leftarrow \text{otherwise} \end{cases}
\tag{29}
$$

where we obtain (30), shown at the bottom of the page, which is simply a comparison with a threshold depending on a set of surrounding line variables (an analogous criterion is obtained for each v_{ij}). Notice that we were led to edge interactions, not explicitly included *a priori*, but rather implicitly contained in the partition function of the CGMRF.

B. Algorithm

The proposed algorithm is composed of two nested loops. The inner loop is basically an iterative restoration/parameter-estimation scheme. Formally, it consists in cyclicly solving (22)–(25). As shown in [14] and [53], the only stationary points of this scheme are POS's; more specifically, they are POS's where (22), (24), and (25) are global maxima [see (26)–(28)], while (23) is a coordinate-wise maximum (characteristic of ICM). The order by which (22)–(25) are solved is irrelevant to this fact; nevertheless, different orders can lead to slightly different POS's. Experimentally, we found that these differences are negligible.

$$
t_{ij}^h = \frac{[1 - \omega(h_{i-1j} + v_{i-1j} + v_{i-1j+1})][1 - \omega(h_{i+1j} + v_{ij} + v_{ij+1})]}{[1 - \omega(1 + h_{i-1j} + v_{i-1j} + v_{i-1j+1})][1 - \omega(1 + h_{i+1j} + v_{ij} + v_{ij+1})]}
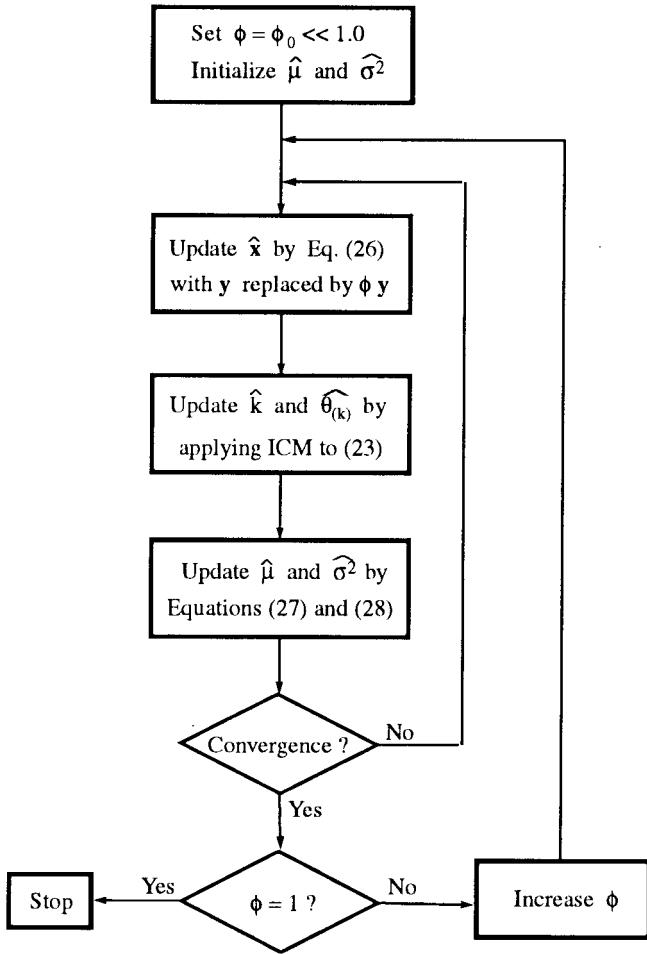\tag{30}
$$

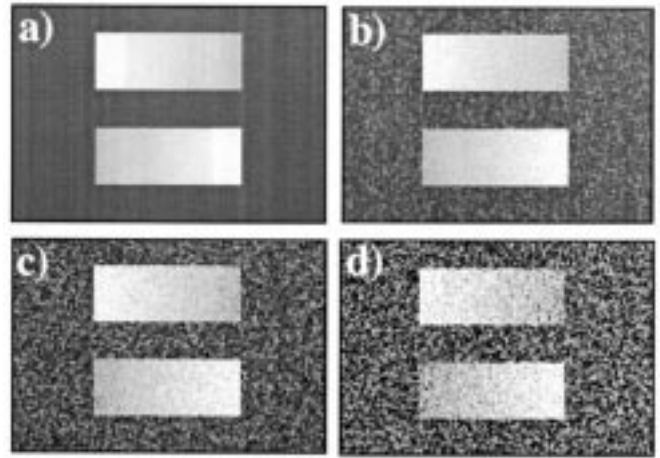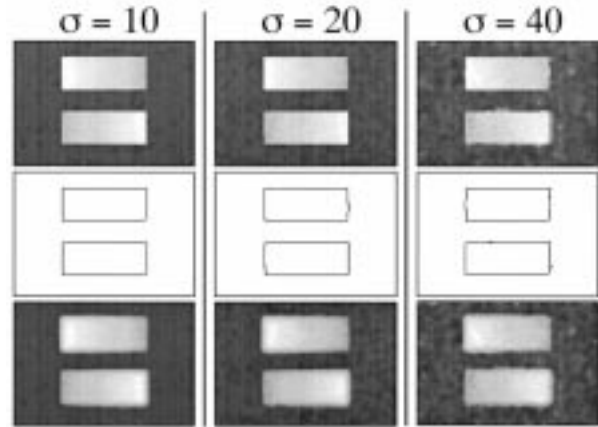Fig. 1. Structure of the proposed algorithm.

Fig. 2. Synthetic image: (a) original and three noisy versions with (b) $\sigma = 10.0$, (c) $\sigma = 20.0$, and (d) $\sigma = 40.0$.

Fig. 3. Results obtained from the noisy images of Fig. 2; first, second and third columns correspond to Fig. 2 (b), (c), and (d), respectively. The first row contains the images restored by the proposed algorithm, the second row shows the corresponding discontinuity configurations, and the third row displays the images resulting from not taking discontinuities into account.

The outer loop is a continuation-type scheme similar to the *simulated tearing* algorithm we have proposed in [15]. The function to be minimized $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$ is embedded into a family

$$\{L_\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2), \phi \in]0, 1]\} \quad (31)$$

indexed by the control parameter ϕ. The family members are defined by $L_\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2) \equiv L(\boldsymbol{x}, \phi\boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$, i.e., by multiplying the observed image by ϕ. For $\phi \simeq 0$, we clearly have $\hat{k} = 0$, thus $\boldsymbol{\theta}_{(k)}$ is an empty (zero-dimensional) vector, and we only have to solve (22), (24), and (25), by applying (26)–(28), which are global maxima (in the POS sense) easy to obtain. These solutions are then tracked by slowly varying ϕ up to $\phi = 1$, i.e., up to the original objective function $L(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}_{(k)}, \mu, \sigma^2)$. As will be seen in the experiments presented ahead, this continuation scheme provides a sequence of POS's of increasing quality, i.e., decreasing values of the objective function.

The complete structure of the algorithm is depicted in Fig. 1. Equation (26) is solved with the Gauss–Seidel method [17]. The ICM scheme used to solve (23) consists in cyclically applying (29) to all edge variables until a stationary configuration is reached (typically two to four sweeps). The convergence condition of the inner loop is verified when the differences between consecutive estimates of μ and σ^2 both fall below some threshold (2% in all examples presented ahead); of course, the number of iterations spent in the inner loop is highly dependent on this threshold. Parameter ϕ evolves linearly according to $\phi = \phi_0 + (1.0 - \phi_0)(t - 1)/(Q - 1)$, where t is the iteration number of the outer loop (starting at $t = 1$, and going up to $t = Q$), ϕ_0 is the initial value, and Q is the total number of iterations of the outer loop; clearly, for $t = Q$ we have $\phi = 1.0$.

Initial estimates of σ^2 and μ are obtained by the following *ad hoc* procedures. The initial estimate of σ^2 is taken as the variance of the difference between the observed image and a median filtered version of it. Concerning μ: i) horizontal and vertical first-order difference images are obtained from the noisy image; ii) these images are squared and median filtered; iii) the means of the two images are computed; iv) finally, the initial estimate of $1/\mu$ is set to the average value of these two means.
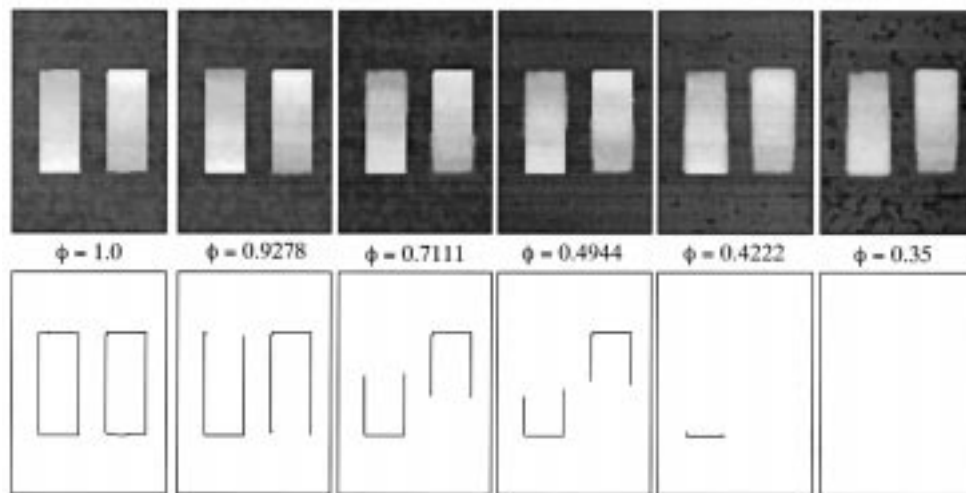
Fig. 4. Some image estimates (with the gray levels multiplied by $1/\phi$) and associated edge configurations obtained along the algorithm for the indicated values of parameter ϕ, from the noisy image of Fig. 2 (c). The final estimate corresponds to $\phi = 1.0$.
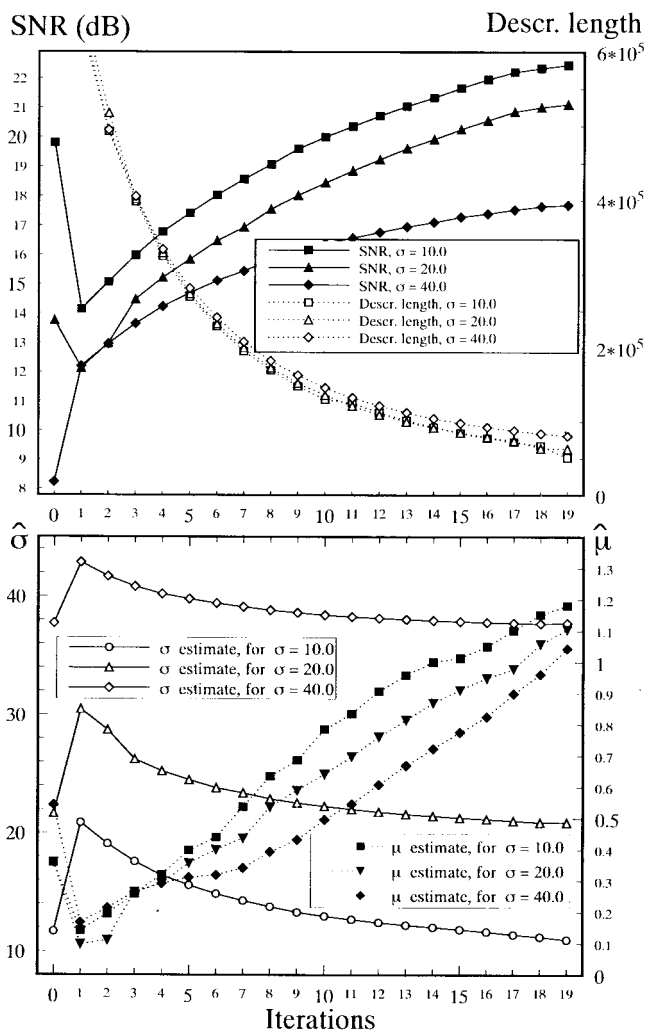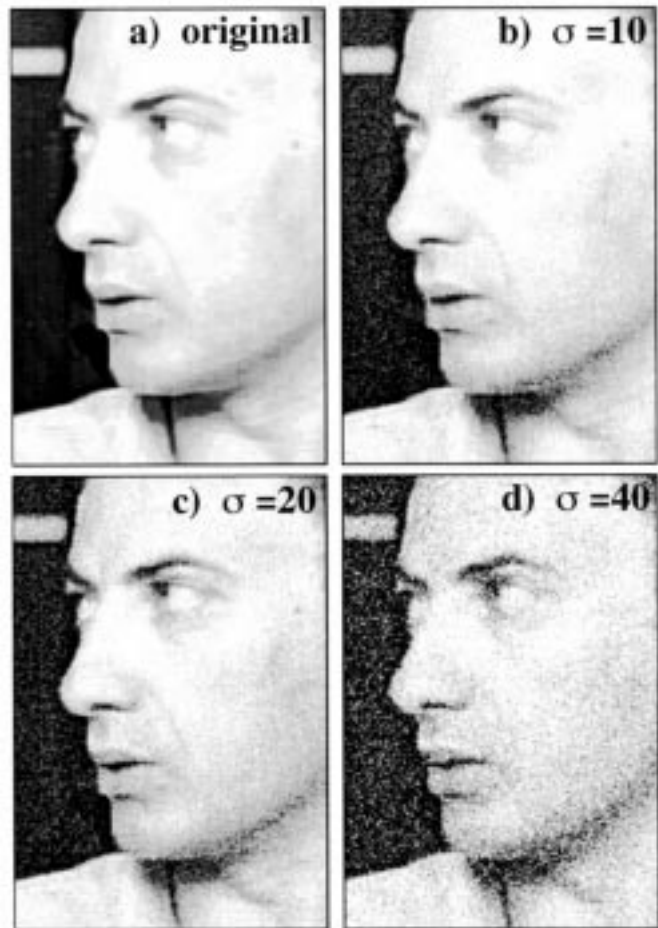

Fig. 5. Description length, SNR, and the estimates of σ and μ, versus iteration number, for the examples of Fig. 3.

Fig. 6. Natural image: (a) original and three noisy versions with (b) $\sigma = 10.0$, (c) $\sigma = 20.0$, and (d) $\sigma = 40.0$.

The computational load of the algorithm is, of course, highly dependent on the image size, on the convergence threshold of the inner loop, and on the total number of iterations of the outer loop (Q). From the involved expressions (basically, (27), (28), (30) and the Gauss–Seidel updates for (26)), a rough (over)estimate can be obtained; for an $M \times N$ image

$$n_{\text{FLOP's}} \simeq (15n_{\text{GS}} + 25n_{\text{ICM}} + 25)n_{\text{IL}}MNQ \qquad (32)$$

Fig. 7. Results obtained from the noisy images of Fig. 6; first, second and third columns correspond to Fig. 6 (b), (c), and (d), respectively. The first row contains the images restored by the proposed algorithm, the second row shows the corresponding discontinuity configurations, and the third row displays the images resulting from not taking discontinuities into account.

Fig. 8. Description length, SNR, and the estimates of σ and μ, versus iteration number, for the examples of Fig. 7.

where $n_{\text{FLOP's}}$ denotes the total number of *floating point operations* (FLOP's), n_{GS} is the number of iterations of the Gauss–Seidel scheme used to solve (26), n_{ICM} is the number of iterations spent in the ICM algorithm, and n_{IL} the (average) number of inner loop iterations. For typical values, $n_{\text{GS}} = 15, n_{\text{ICM}} = 4, n_{\text{IL}} = 5, Q = 20, M \times N = 10^4$, we get $n_{\text{FLOP's}} \simeq 350 \ 10^6 = 350$ MFLOP's (millions of FLOP's). On a reasonably fast workstation, this represents a few seconds.

V. EXPERIMENTAL EXAMPLES

A. Introduction

This section reports a set of experimental results. The initial parameter of the outer loop was always set to $\phi_0 = 0.35$; this value is sufficiently low to assure that the first estimate of k is 0. As in any continuation method, the quality of the results depends strongly on Q; however, some simple tests showed that there is no point in going above $Q = 15 \sim 20$ (see [15]). We took $Q = 19$, but very similar results would be obtained with, e.g., $Q = 16$ or $Q = 20$; some serious degradation
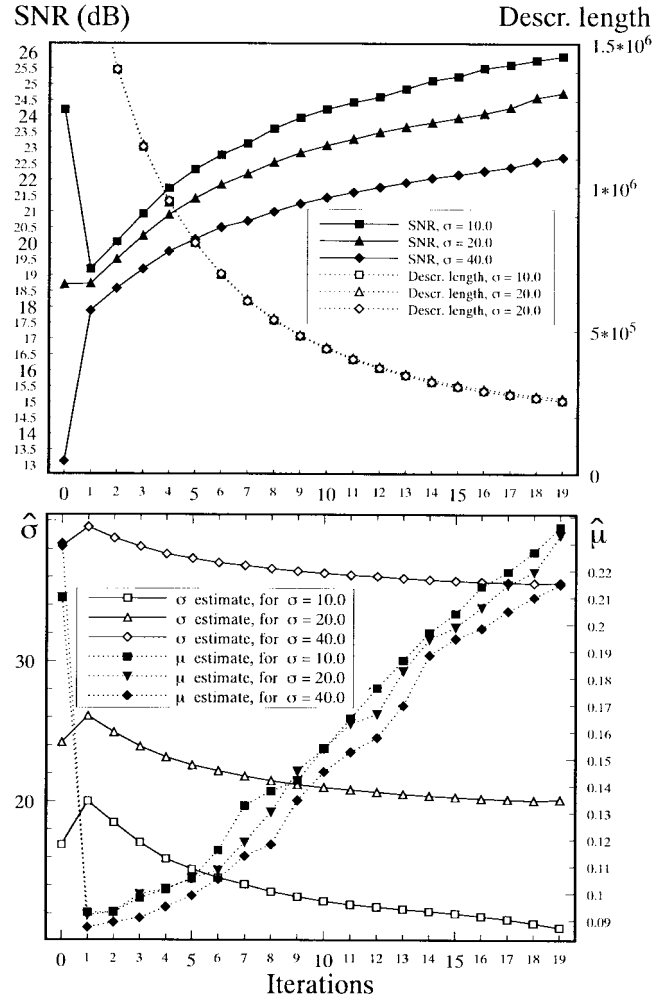
appears for values below $Q = 6 \sim 9$. As mentioned above, all tests were performed with $\omega = 0.2499$.

B. Synthetic Images

Fig. 2 presents the original image and three noisy versions of it (with $\sigma = 10.0, \sigma = 20.0$, and $\sigma = 40.0$) used in the first set of experiments. In Fig. 3, the results obtained from these noisy images are displayed: the first row shows the final image estimates obtained by the proposed scheme, while the second row exhibits the corresponding discontinuity configurations; the images in the third row were obtained without considering discontinuities (but taking the parameter estimates provided by the proposed scheme); notice how edges are unduly smoothed.

With the purpose of illustrating the working of the algorithm, some intermediate image and edge estimates are shown in Fig. 4 (the image estimates presented were multiplied by $1/\phi$ to have a similar mean gray level). Fig. 5 plots the evolutions of the description length, the signal-to-noise ratio (SNR[4]), and the estimates of μ and σ versus the iteration

[4]The definition used is $\text{SNR} = 10 \log_{10} (s_{\text{orig}}^2 / s_{\text{dif}}^2)$, where s_{orig}^2 is the average of squares of the original (undegraded) image and s_{dif}^2 is the average of squares of the difference between the considered image and the undegraded one [25].
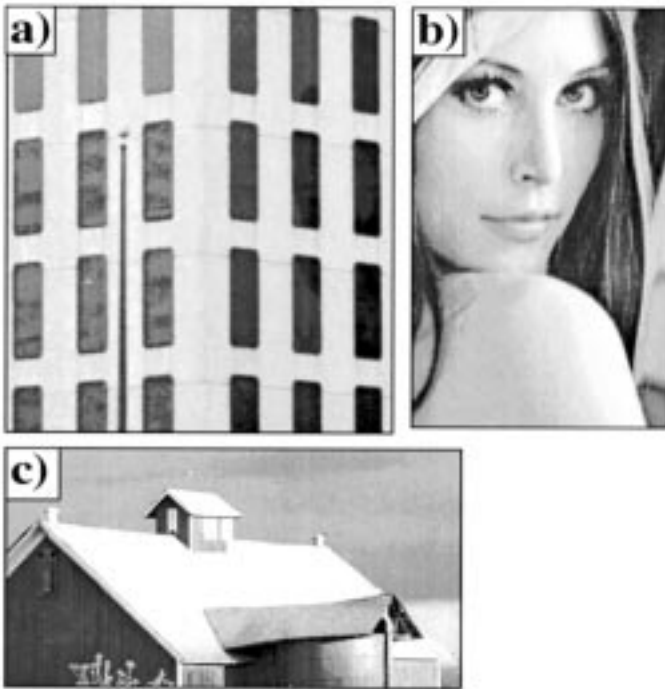
Fig. 9.   Natural images (original versions). (a) Building. (b) Lena. (c) farm.

Fig. 10.   Example based on the image of Fig. 9(a). (a) Noisy version. (b) Restored image. (c) Corresponding edge configuration. (d) Smoothed image (no edges considered).
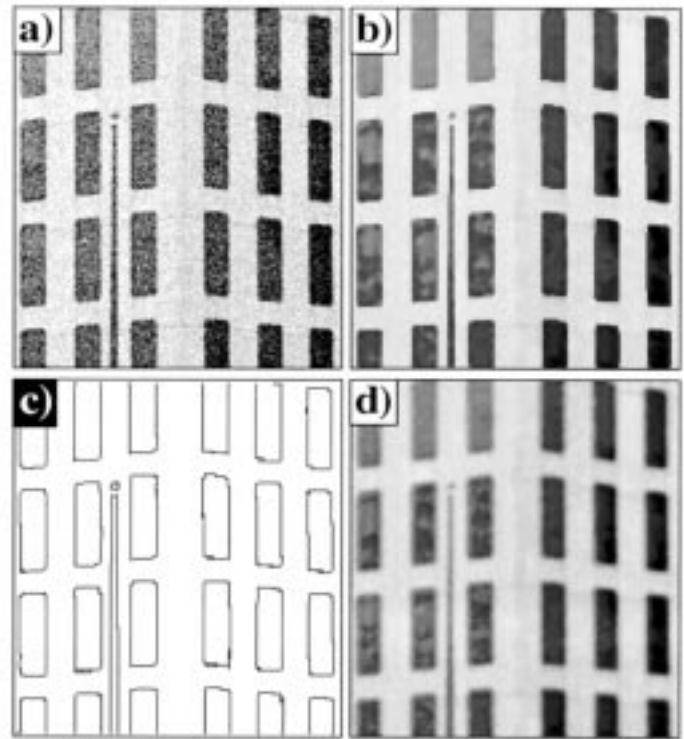

Fig. 11.   Example based on the image of Fig. 9(b). (a) Noisy version. (b) Restored image. (c) Corresponding edge configuration. (d) Smoothed image (no edges considered).

number of the outer loop. Abscissa 0 corresponds to the initial values of the description length and SNR (i.e., for the noisy images), and to the initial estimates of σ and μ (obtained as described above). When computing the description length and the SNR, x is multiplied by $1/\phi$ to cancel the effect of having multiplied y by ϕ and, thus, obtain meaningful plots; for similar reasons, the plotted values of $\hat{\sigma}$ and $\hat{\mu}$ are multiplied by $1/\phi$ and ϕ^2, respectively. Note how the SNR in the first iterations is even lower than the SNR of the observed noisy image; this is due to the fact that the first estimates are oversmoothed and obtained with parameter estimates still far from the final values. See also how the first estimates obtained by the algorithm (iteration 1) are quite different from the initial ones; this demonstrates robustness with respect to initialization. We present a final observation: The continuation procedure does, in fact, provide a sequence of estimates with decreasing values of the objective function (the description length), as is apparent in Fig. 5.

The final estimates of σ are 10.9, 20.8, and 37.7; the errors are below 10%, which can be considered as reasonable. Underestimation in the higher noise situation is caused by saturation effects, i.e., although noise with $\sigma = 40$ was added to the original image, the result is bounded to the interval [0, 255], and thus the resulting noise has smaller effective standard deviation. The similarity among the final μ estimates, under the three different noise conditions, reveals a good degree of consistency.

C. Natural Images

The first set of tests with natural images is parallel to the one performed with synthetic images described in the previous subsection. Figs. 6–8 report the results obtained; all explanations and comments of the preceding subsection also apply to these experiments. The final estimates of σ are now 11.0, 20.1, and 35.7; the first two are again good estimates, with errors below 10%. Underestimation in the higher noise situation was already justified in the previous subsection. Again, a good consistency among the three μ estimates is apparent.

The next set of tests uses the natural images of Fig. 9 (building, Lena, and farm), with $\sigma = 20.0$. Figs. 10–13 show the results. The apparent loss of texture is typical of CGMRF-type priors which are surface models rather than texture models. Notice also how the images restored without discontinuities are oversmoothed. The plots in Fig. 13 were obtained as the previous ones; the final σ estimates

Fig. 12. Example based on the image of Fig. 9(c). (a) Noisy version. (b) Restored image. (c) Corresponding edge configuration. (d) Smoothed image (no edges considered).
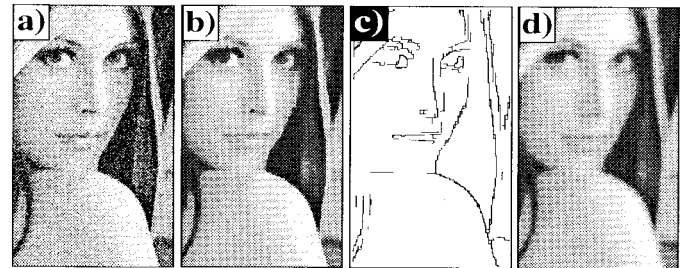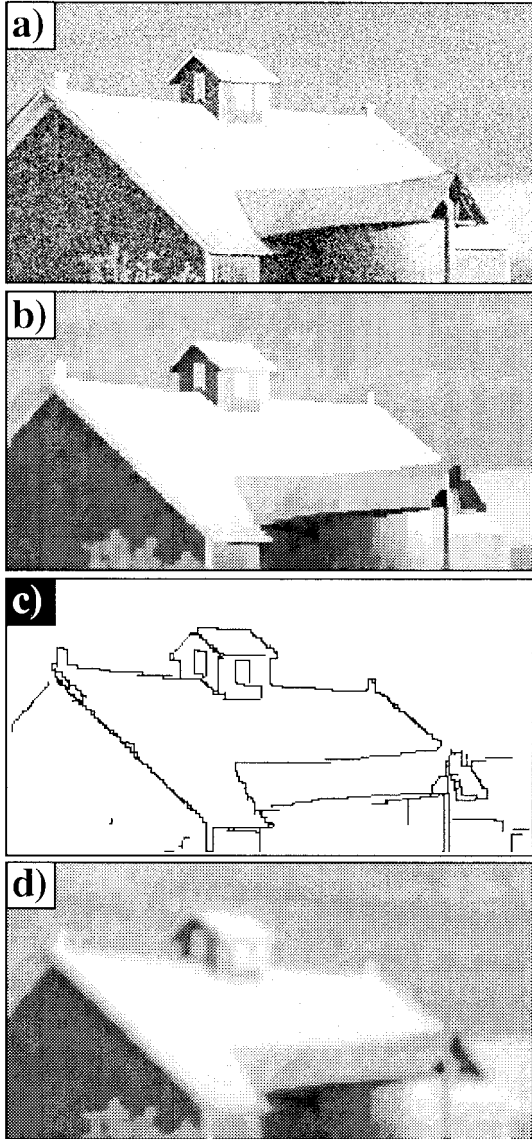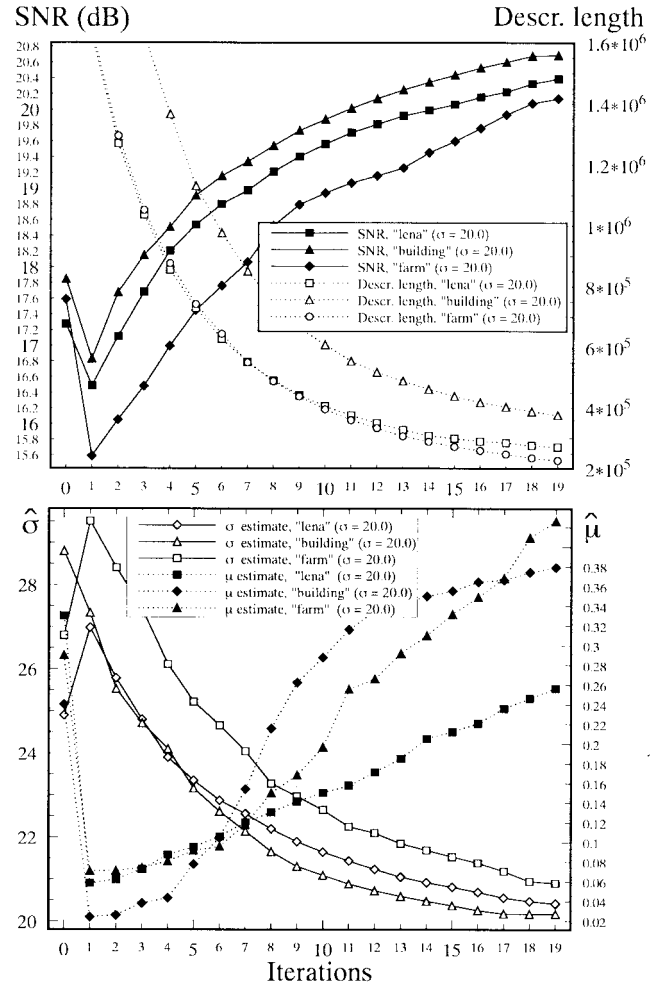

Fig. 13. Description length, SNR, and the estimates of σ and μ, versus iteration number, for the examples of Figs. 10–12.

TABLE I
SUMMARY OF QUANTITATIVE RESULTS IN TERMS OF SNR

Noisy image	Signal-to-noise ratio (SNR)		
	initial	proposed algorithm	without edges
Fig. 2 (b)	19.8 dB	22.5 dB	17.1 dB
Fig. 2 (c)	13.7 dB	21.1 dB	16.3 dB
Fig. 2 (d)	8.2 dB	17.7 dB	16.0 dB
Fig. 6 (b)	24.2 dB	25.9 dB	23.8 dB
Fig. 6 (c)	18.7 dB	24.7 dB	22.6 dB
Fig. 6 (d)	13.1 dB	22.7 dB	21.4 dB
Fig. 10 (a)	17.8 dB	20.7 dB	18.0 dB
Fig. 11 (a)	17.2 dB	20.4 dB	19.0 dB
Fig. 12 (a)	17.6 dB	20.1 dB	17.4 dB

are now 20.4, 20.2, and 20.9, respectively, i.e., with errors below 5%.

Table I summarizes the results in terms of SNR, comparing the noisy images with the restored images obtained with and without discontinuities. A conclusion that can be drawn from these results is that the SNR obtained without considering edges is always worse than the one obtained with the edge-preserving technique; sometimes, it is even worse than the SNR of the noisy image itself.

D. MAP Estimates Obtained by the MFA Algorithm

The most closely related edge-preserving image restoration method is the MAP estimator (see Section III-B) with a linear edge penalty function $U(\boldsymbol{h}, \boldsymbol{v}) = \gamma \Sigma_{ij} (h_{ij} + v_{ij})$, as, e.g., in [20], [48], [49], and [56]. This is a nonadaptive

criterion that has to be hand tuned; naturally, with an adequate choice of parameters, the results will be extremely close to
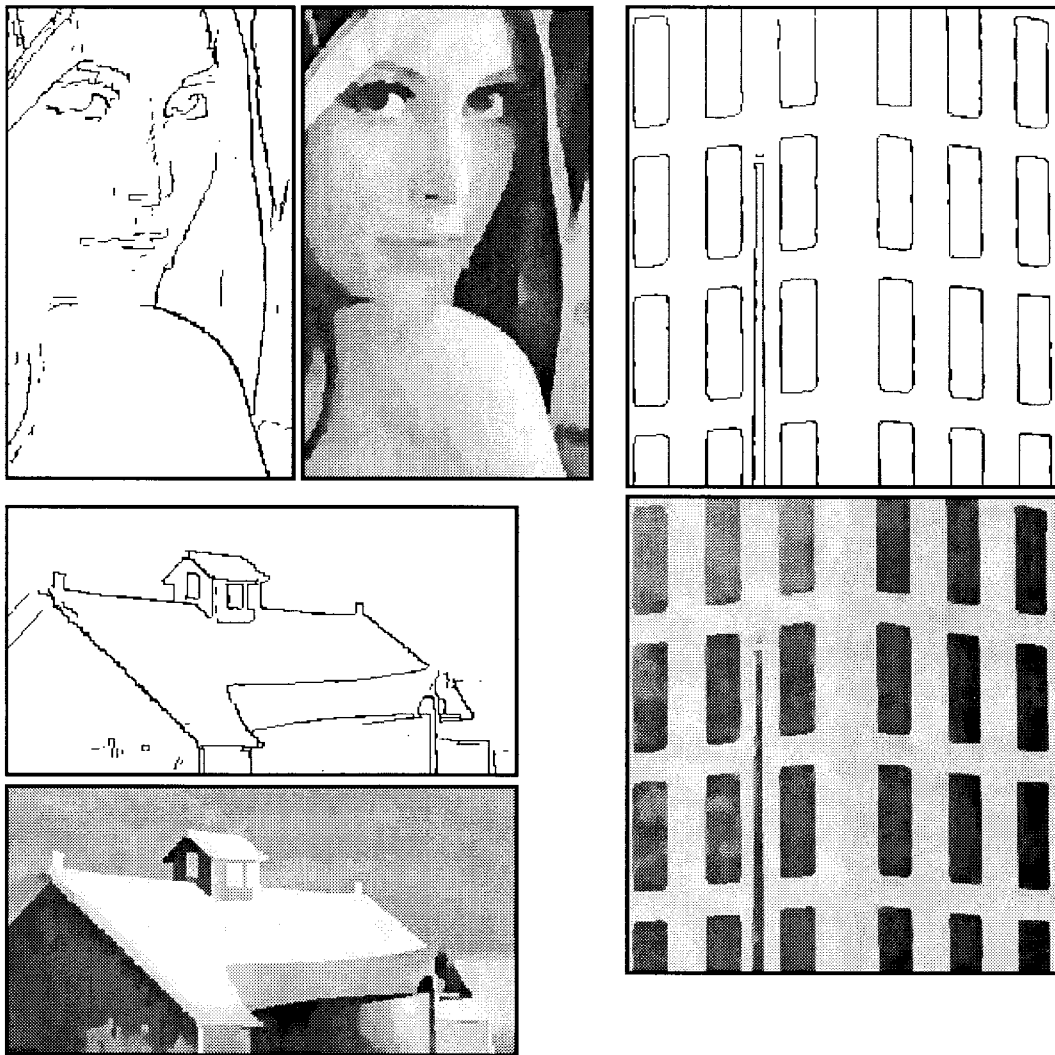
Fig. 14.   Results obtained with the (user supervised) MFA algorithm.

those obtained with the proposed adaptive scheme (since the underlying prior is the same). This fact was confirmed by means of the MFA algorithm (as described in [56]) with the noisy images of Figs. 10–12. The true σ^2 and the estimated (by our method) μ values were used; parameter γ was tuned to yield the best possible results. Fig. 14 shows the resulting discontinuity configurations and restored images, which are in fact very similar to those in Figs. 10–12. The SNR values are 20.8 dB, 20.3 dB, and 20.0 dB, which are also very close to those obtained with our method (see Table I). We stress that these results could only be obtained with user interaction and not in an unsupervised manner as in the method proposed in this paper.

VI. CONCLUDING REMARKS

In this paper we have introduced a new approach to unsupervised discontinuity-preserving image restoration using compound Gauss–Markov random fields and the minimum description length principle. The main features are that i) no Bayesian prior concerning the edges/discontinuities is needed

and ii) no knowledge of the hyperparameters (noise and original field variances) is required. To deal with the resulting optimization problem, a deterministic continuation-type iterative technique was developed.

Experiments were carried out with synthetic and natural images, contaminated by noise with different variances. Our algorithm was able to achieve edge-preserving restoration while simultaneously correctly estimating the noise standard deviation and the global variance of the original field. Particularly relevant are the good noise standard deviation estimates (errors always below 10%, sometimes well below 5%) testifying for the good performance of the technique. The advantage of using a discontinuity-preserving approach, over a nondiscontinuity-preserving one, was evidenced both visually and quantitatively (in terms of SNR).

REFERENCES

[1] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[2] G. Archer and D. Titterington, "On some Bayesian/regularization methods for image restoration," *IEEE Trans. Image Processing*, vol. 43, pp. 989–995, 1995.

[3] N. Balram and J. Moura, "Noncausal Gauss-Markov random fields: Parameter structure and estimation," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1333–1355, 1993.

[4] J. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 1980.

[5] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, pp. 869–889, 1988.

[6] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.

[7] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[8] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] T. Darrel and A. Pentland, "Cooperative robust estimation using layers of support," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 474–487, 1995.

[10] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–38, 1977.

[11] J. Dengler, "Estimation of discontinuous displacement vector fields with the minimum description length criterion," in *Proc. IEEE Conf. Computer Vision and Pattern Recovery—CVPR'91*, Maui, HI, 1991, pp. 276–282.

[12] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 39–55, 1987.

[13] M. Figueiredo, "Image restoration with and without discontinuities: computational and modeling aspects," Ph.D. dissertation, Instituto Superior Técnico, Lisbon, Portugal, 1994.

[14] M. Figueiredo and J. Leitão, "Bayesian estimation of ventricular contours in angiographic images," *IEEE Trans. Med. Imag.*, vol. 11, pp. 416–429, 1992.

[15] ——, "Simulated tearing: An algorithm for discontinuity preserving visual surface reconstruction," in *Proc. IEEE Conf. Computer Vision and Pattern Recovery—CVPR'93*, New York, 1993, pp. 28–33.

[16] ——, "Adaptive discontinuity location in image restoration," in *Proc. IEEE Int. Conf. Image Processing—ICIP'94*, Austin, TX, 1994, vol. II, pp. 665–669.

[17] ——, "Sequential and parallel image restoration: Neural network implementations," *IEEE Trans. Image Processing*, vol. 3, pp. 789–801, 1994.

[18] ——, "Unsupervised contour estimation," in *Proc. IEEE Int.. Conf. Image Processing—ICIP'96*, Lausanne, Switzerland.

[19] H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comput.*, vol. EC-10, pp. 260–268, 1961.

[20] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Machne Intell.*, vol. 13, pp. 401–412, 1991.

[21] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.

[22] W. Grimson and T. Pavlidis, "Discontinuity detection for visual surface recognition," *Comput. Vis., Graph., Image Processing*, vol. 30, pp. 316–330, 1985.

[23] H. Gu, M. Asada, and Y. Shirai, "The optimal partition for moving edge segments," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition—CVPR'93*, New York, pp. 367–372.

[24] C. Huang, "Coding of two-tone images," *IEEE Trans. Commun.*, vol. COMM-25, pp. 1406–1424, 1977.

[25] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[26] F. Jeng and J. Woods, "Image estimation by stochastic relaxation in the compound Gaussian case," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing—ICASSP'88*, New York, pp. 1016–1019.

[27] ——, "Compound Gauss-Markov random fields for image estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 683–697, 1991.

[28] Z. Kato, J. Zerubia, and M. Berthod, "Unsupervised parallel image classification using a hierarchical Markovian model," in *Proc. Int. Conf. Computer Vision—ICCV'95*, Cambridge, MA.

[29] K. Keeler, "Map representations and coding-based priors for segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition—CVPR'91*, Hawaii, pp. 420–425.

[30] R. Lagendijk, J. Biemond, and D. Boekee, "Identification and restoration of noisy blurred images using the expectation-maximization algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1180–1191, 1990.

[31] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs fields using simulated annealing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 799–813, 1989.

[32] D. Langan, J. Modestino, and J. Zhang, "Cluster validation for unsupervised stochastic model-based image segmentation," in *Proc. IEEE Int. Conf. Image Processing—ICIP'94*, Austin, TX, vol. II, pp. 197–201.

[33] K. Lay and A. Katsaggelos, "Blur identification and image restoration based on the EM algorithm," *Opt. Eng.*, vol. 29, pp. 436–445, 1990.

[34] Y. Leclerc, "Constructing simple stable descriptions for image partitioning," *Int. J. Comput. Vis.*, vol. 3, pp. 73–102, 1989.

[35] M. Li, "Minimum description length based 2D shape description," in *Proc. Int. Conf. Computer Vision—ICCV'93*, Berlin, Germany, pp. 512–517.

[36] D. Marr, *Vision*. New York: Freeman, 1979.

[37] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, pp. 76–89, 1987.

[38] J. Mendel, *Maximum-Likelihood Deconvolution: A Journey into Model-Based Signal Processing*. New York: Springer-Verlag, 1990.

[39] J. Moura and N. Balram, "Recursive structure of noncausal Gauss–Markov random fields," *IEEE Trans. Inform. Theory*, vol. 38, pp. 334–354, 1992.

[40] D. Mumford and J. Shah, "Boundary detection by minimizing functionals," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition—CVPR'85*, pp. 22–26.

[41] S. Nadabar and A. Jain, "Parameter estimation in MRF line process models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 326–329, 1996.

[42] R. Nevatia and K. Babu, "A survey of edge detection techniques," *Comput. Graph. Image Processing*, vol. 13, pp. 257–269, 1980.

[43] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Stat.*, vol. 11, pp. 416–431, 1983.

[44] ——, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 629–636, 1984.

[45] ——, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 40–47, 1996.

[46] ——, *Stochastic Complexity in Stastistical Inquiry*. Singapore: World Scientific, 1989.

[47] J. Sheinvald, B. Dom, W. Niblack, and S. Banerjee, "Combining edge pixels into parametrized curve segments using the MDL principle and the Hough transform," in *Advances in Image Analysis*, Y. Mahdavieh and R. Gonzalez, Eds. SPIE, 1992, pp. 503–536.

[48] T. Simchony, R. Chellappa, and Z. Lichtenstein, "Graduated nonconvexity algorithm for image estimation using compound Gauss Markov field models," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing—ICASSP'89*, Glasgow, Scotland, pp. 1417–1420.

[49] ——, "Pyramid implementation of optimal-step conjugate-search algorithms for some low-level vision problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, pp. 1408–1425, 1990.

[50] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 413–424, 1986.

[51] A. Tikhonov, A. Goncharsky, and V. Stepanov, "Inverse problems in image processing," in A. Tikhonov and A. Goncharsky, Eds., *Ill-Posed Problems in the Natural Sciences*. Moscow, Russia: Mir, 1987, pp. 220–232.

[52] V. Torre and T. Poggio, "On edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 147–163, 1986.

[53] C. Won and H. Derin, "Unsupervised segmentation of noisy and textured images using Markov random fields," *Comput. Vision, Graph., Image Processing: Graph. Models Image Processing*, vol. 54, pp. 308–328, 1992.

[54] J. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232–240, 1972.

[55] Z. Wu and R. Leahy, "An approximate method of evaluating the joint likelihood function for first-order GMRF's," *IEEE Trans. Image Processing*, vol. 2, pp. 520–522, 1993.

[56] J. Zerubia and R. Chellappa, "Mean field annealing using compound Gauss-Markov random fields for edge detection and image estimation," *IEEE Trans. Neural Networks*, vol. 4, pp. 703–709, 1993.

[57] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Processing*, vol. 2, pp. 27–40, 1993.

[58] Y. Zhao, X. Zhuang, L. Atlas, and L. Anderson, "Parameter estimation and restoration of noisy images using Gibbs distribuitions in hidden Markov models," *Comput. Vis., Graph., Image Processing: Graph. Models Image Processing*, vol. 54, pp. 189–197, 1992.

[59] J. Zheng and S. Bolstein, "Motion-based object segmentation and estimation using the MDL principle," *IEEE Trans. Image Processing*, vol. 4, pp. 1223–1235, 1995.

Mário A. T. Figueiredo (S'87–M'95) was born in Luanda, Angola, in 1962. He received the E.E., M.S., and Ph.D. degrees in electrical and computer engineering, all from Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 1985, 1990, and 1994, respectively.

He is currently Assistant Professor at the Department of Electrical and Computer Engineering, IST, and Researcher with the Communication Theory and Pattern Recognition Group, Institute of Telecomunications. His research interests include image processing, low-level computer vision, medical imaging, information theory and coding.

José M. N. Leitão (M'95) was born in Aldeia Velha-Avis, Portugal, in 1946. He received the E.E. and Ph.D. degrees in electrical engineering in 1970 and 1983, respectively, both from Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal. He received the "Agregado" degree in electrical and computer engineering, also from IST, in 1992.

He was with the Laboratory of Physiology of the Instituto Gulbenkian de Ciência, Oeiras, Portugal, from 1970 to 1972. After spending three years at the University of Tübingen, Germany, he joined the faculty of IST in 1976, where he is currently Full Professor with the Department of Electrical and Computer Engineering. He is also the coordinator of the Communication Theory and Pattern Recognition Group, Institute of Telecommunications. His main research interests are communication theory, pattern recognition, and signal and image processing.