ELSEVIER

Image and Vision Computing 17 (1999) 779–793

image AND vision COMPUTING

Unsupervised edge-preserving image restoration via a saddle point approximation

Luigi Bedini*, Anna Tonazzini, Salvatore Minutoli

Istituto di Elaborazione della Informazione, Consiglio Nazionale delle Ricerche, Via S. Maria, 46, I-56126 Pisa, Italy

Accepted 15 July 1998

Abstract

This paper proposes a fast method to estimate the Gibbs hyperparameters of an MRF image model with explicit lines during the restoration process. It consists of a mixed-annealing algorithm for the maximization of the posterior distribution with respect to the image field, periodically interrupted to compute, via ML estimation, a new set of parameters. We first consider the weak membrane model and show that, by adopting a saddle point approximation for the partition function, these new parameters are defined as those that maximize the conditional prior distribution of the lines given the intensities, evaluated on the current estimate of the whole image field. In this way the computation of the expectations involved in the ML estimation can be performed by analytical summation over the binary line elements alone, with a strong reduction of the computational complexity. The approach can be extended to the general case of self-interacting line models, by substituting the analytical computations with a binary, short-range Gibbs sampler.

Keywords: Edge-preserving regularization; Gibbs parameter estimation; Unsupervised image restoration

1. Introduction

In images of real scenes, the intensity discontinuities are often related to features, such as object boundaries, occlusions or textures, which are fundamental for subsequent classification and recognition tasks. Any method for solving visual reconstruction problems should thus preserve them correctly, by exploiting available a priori information about their local properties of connectivity, thinness and smoothness. The most effective way to describe the local behaviour of both the intensity field f and the discontinuity field l (line elements) has so far been to use Markov random field (MRF) models, in the form of prior Gibbs distributions. In the Bayesian framework, the reconstruction problem can then be reformulated as a maximum a posteriori (MAP) estimation or, equivalently, as the minimization of a posterior energy function, or cost function, $E(f,l)$, given the data $\mathbf{g}$ [1,2]. As usual, we consider $E(f,l) = E_g(f) + U(f,l)$, where $E_g(f)$ is the data consistency term and $U(f,l)$ is a sum of various terms expressing a number of constraints on the local smoothness of the intensity process, on the possible configurations of the line process, and on the mutual interactions between the two processes. Each of these terms is weighted by a free parameter that expresses the relative confidence on the corresponding constraint.

Despite its better performance with respect to maximum likelihood estimation or standard regularization, the approach described above presents two significant difficulties. The first is related to the parametric nature of the image models considered. Indeed, although the functional form of the prior distribution is usually known, the a priori information are not enough to determine the vector $\mathbf{w}$ of the free parameters (also known as hyperparameters or Gibbs parameters) of the model. On the other hand, the values of these parameters have been proved to be critical, since they have a strong effect on the reconstructions. The second difficulty originates from the fact that the cost function is always non-convex. This implies that standard descent methods do not guarantee convergence to the global minimum.

Although as yet none conclusive, a great deal of work has been done about the definition of efficient and fast non-convex optimization algorithms. In particular, both deterministic and stochastic algorithms have been proposed. The stochastic ones, based on simulated annealing techniques, ensure asymptotic convergence to the global optimum but present formidable computational complexity, due to the dimension of the problems treated and the number of iterations required [1,3]. The deterministic methods, such as the graduated non-convexity (GNC) algorithm [4], the

* Corresponding author. Tel: + 39 50 593400; fax: + 39 50 554342.
E-mail address: bedini@iei.pi.cnr.it (L. Bedini)

0262-8856/99/$ - see front matter
PII: S0262-8856(98)00154-1

mean-field annealing [5,6], the iterated conditional modes (ICM) algorithm [7], and the generalized expectation-maximization (EM, GEM) technique [8–10], allow for some reduction of the computation time but are sub-optimal.

With respect to the parameter estimation problem, there is as yet no general effective approach, so currently the most common strategy is to select them empirically, by trial and error. Nevertheless, many efforts have been made to find criteria that lead to objective and optimal estimators. The supervised techniques assume that the parameters can be estimated off-line from training data, consisting of one or more clean samples (i.e. realizations) of the underlying MRF model. When the hyperparameters have to be estimated for reconstruction purposes, e.g. to restore a given degraded image, training data is often not available, so unsupervised techniques have to be adopted, which attempt to estimate the parameters adaptively during reconstruction, directly from the observed data.

In the supervised techniques, maximum likelihood (ML) estimators are applied, which are based on the prior computed on the available samples. If the parameters are identifiable, the log-prior is a strictly concave function with respect to the parameters, and this estimate is asymptotically consistent [11,12]. The main computational difficulty with this approach is that the partition function Z of the distribution considered is also a function of the hyperparameters, and must be taken into consideration when maximizing the distribution itself. Since the calculation of Z involves all possible configurations of the MRF, the direct maximization of the distribution becomes intractable even for small size problems. Thus, in practice, various approximations are adopted. Younes [11,13] demonstrated the convergence of a stochastic gradient ascent algorithm to maximize the original log-prior, when the expectation in the gradient is approximated by the value computed on the last configuration of a single-sweep-long Markov chain. Nevertheless, the most popular approach makes use of the maximum pseudo-likelihood (MPL) estimation [14], which is an extension of the coding method [14]. These approximations are based on assuming probabilistic independence of subsets (codings) of elements in the field and then factorizing the distribution. Although the factorized distribution is not a true likelihood, in many cases of parameter estimation, MPL can be considered a good approximation of the ML estimate, and it is also consistent, in the sense that it converges in probability to the ML estimate when the size of the image increases [7]. A similar approach is based on deriving a mean field approximation [14] for the distribution, assuming that the fluctuations around the mean values of the field are small. To study the effectiveness of MRF models for textured images, Cross and Jain [15] used the coding method to estimate the model parameters from natural texture samples. Derin and Guler [16] studied the properties of the MRF model for the explicit line process alone. They tested an MRF line model which incorporates various constraints on the local line geometry, and computed an MPL estimate of the parameters from synthetic and real edge maps.

In a previous work [17], for application to MRF models for piecewise smooth images with explicit constrained discontinuities, we proposed using the learning algorithm of a Generalized Boltzmann Machine (GBM) with multiple interactions [18,19]. The GBM defined has pixels and line elements as units, and has the parameters of the prior Gibbs energy as weights for cliques of interacting neurons. The learning algorithm for this GBM minimizes the Kullback distance (or divergence measure) between the internal distribution of the network and a theoretical distribution from which the training examples are ideally drawn [20,21]. During a sequence of alternating clamped and unclamped example presentations, the weights are updated on the basis of expectations computed on the free units. It is straightforward to observe that this learning algorithm behaves like an ML estimator from realizations of the MRF. In particular, the rule for updating the weights amounts exactly to a gradient ascent over the MRF log-prior. Again, also in this case, the computational difficulty is related to the computation of the expectations involved, which would require summation over the continuous intensity field and the binary line field. Thus, we considered the special case where the pixel units, which may be continuous, are always clamped by the environment, while the binary line units evolve freely according to a stochastic dynamics that recalls the well known Gibbs sampler. This amounts to computing the ML estimation of the parameters on the basis of the prior probability distribution of the line elements, conditioned on the intensity elements, and can lead to a significant reduction in the computational demand of the learning algorithm, since we only need to compute expectations over binary units. Our tests showed that excellent convergence to the true parameters can be obtained when a large number of samples, constituting intensity plus discontinuity maps, are drawn from a given distribution and used as examples [22].

As already said, when the parameter estimation is for reconstruction purposes, samples of the MRF model are usually unavailable. Thus, one can rely solely on the data, which are the only available observation (though incomplete and unclean) of the underlying, unobservable MRF. Moreover, from our experience, we found that the parameters estimated from clean realizations of the MRF are not always suitable for getting good reconstructions. In other words, the characteristics of the degradation process must be taken into account. An attempt in this direction could be to base the estimation over degraded versions of the available clean realization. We applied this strategy with some success, for image restoration via a mixed-annealing algorithm augmented with the off-line learning algorithm for Boltzmann Machines already described. The examples we used were an undegraded image with the same characteristics of the one that had to be restored, plus a few

degraded versions of it, coupled with the ideal map of the line elements [17].

Unsupervised techniques have been successfully applied in the case where **w** reduces to the single smoothness parameter λ, as in standard regularization [23] or when the discontinuities are implicitly addressed [4,24]. Several statistical methods have been used and compared to estimate λ from the data, while reconstruction is being performed. The most popular methods are the chi-squared, the generalized cross-validation, the equivalent degrees of freedom [25–27], and the ML estimate [28]. In the general case of the Gibbs parameters, one could attempt to maximize, with respect to **w**, the marginal distribution of the data **g** conditioned on the parameters themselves. This naturally leads to an EM algorithm, when the observations are viewed as incomplete data. Nevertheless, the marginal distribution is no longer concave, and no convergence proof can be given [11]. Moreover, also in this case an impractical summation over the MRF variables is required, with a consequent need for approximations. Alternatively, an iterative MAP–ML procedure can be adopted for simultaneous image reconstruction and parameter estimation [7,29]. More precisely, given an initial set of parameters, the MAP reconstruction is periodically interrupted, and the current estimate is interpreted as a realization of the MRF and then used to compute, via ML estimation, a new set of parameters. The same considerations made for the case of supervised estimation apply to this case, so that practical algorithms require the adoption of some form of approximation.

For example, Lakshmanan and Derin [29] applied MPL for unsupervised image segmentation. They used an external simulated annealing scheme for the MAP segmentation of the discrete-valued intensity elements *f*, which is periodically interrupted to estimate the parameters **w** through an internal simulated annealing algorithm. For image restoration with implicit lines, Besag proposed carrying out a single cycle of ICM to approximate the MAP solution, followed by MPL to estimate the parameters, at each iteration [7]. On the other hand, Zhang proposed using the mean field approximation and the EM algorithm both in image segmentation [30], where the MRF is discrete-valued, and in image restoration [31], where the MRF is given by a continuous-valued intensity process, coupled with a binary line process. In the second case, he addressed the problem of estimating the blur parameters (blind restoration), and assumed that the MRF model parameters were known.

Recently, Monte Carlo Markov Chain (MCMC) techniques and the importance sampling theorem [32] have been proposed to approximate the ratio of two parametric partition functions taken at two different values of the parameters themselves. This result has been exploited for the ML estimation of the Gibbs parameters, in the context of unsupervised image reconstruction. Higdon et al. applied MCMC techniques for the iterative and simultaneous estimation of the image and the parameters in emission tomography, via stochastic sampling [33]. Descombes et al. applied the same technique for the unsupervised segmentation of images, where the ML parameter estimation step is performed via gradient ascent over the log-prior [34].

In the present paper, we deal with the unsupervised restoration of piecewise smooth images with explicit, constrained discontinuities. We propose an iterative mixed-annealing procedure for the MAP restoration, to be periodically interrupted to estimate the parameters from the current values (f^k, l^k) of both the intensity and the line process via ML estimation. Owing to the presence of the explicit line process, we look for the vector **w** which maximizes the probability of having the line configuration l^k, given the intensity configuration f^k, rather than looking for the vector **w** which maximizes the joint probability of having the couple (f^k, l^k). It is easy to see that in this way the parameter estimation results in an ML estimation based on the conditional prior distribution. The computation of the expectations involved in a gradient ascent technique for the maximization of the log conditional prior thus requires summation over binary variables alone. In particular, when the line process is non-interacting, the expectations can be computed analytically. They can otherwise be computed through a cheap Gibbs sampler over binary variables. This technique has close relationship with an MCMC-based method, which is fully developed in a parallel paper by the present authors [35].

Basing the estimate of the conditional prior is feasible in our case, since all the parameters of the MRF model considered herein are related to local configurations of the line process and hence they are all present in the conditional prior itself. In fact, we consider the line elements as an auxiliary process, whose configurations are sufficient to describe the local features of the coupled field. As an additional argument for the value of the method, we show that it can be derived by taking a saddle point approximation for the partition function of the joint posterior distribution, when we regard it as a function of both the image field and the parameters.

The paper is organized as follows. In Section 2 we recall the basic concepts underlying the techniques for the simultaneous MAP restoration and ML estimation, and revise the mixed-annealing algorithm adopted for the MAP restoration. In Section 3 we derive the method proposed herein for the ML estimation of the parameters, in the case of non-interacting lines (weak membrane model), and show how it is straightforward to extend it to the general case of self-interacting, constrained lines. In Section 4 the overall, practical MAP–ML procedure is established, and finally, in Section 5, the experimental results are discussed.

2. Simultaneous MAP restoration and ML estimation of the Gibbs parameters

In the Bayesian framework based on MRF models with explicit discontinuities, the original $N \times N$ image to be

recovered is regarded as a pair (f,l), where f is the matrix of random variables associated with the observable pixel intensities, and $l=(h,v)$ represents a couple of matrices of binary random variables associated with the unobservable interpixel discontinuities (horizontal and vertical line elements). The Gibbsian prior probability of the MRF is defined according to:

$$P(f,l)=\frac{1}{Z}\exp(-U(f,l)) \tag{1a}$$

$$U(f,l)=\sum_{C\in\mathscr{C}} V_C(f,l) \tag{1b}$$

In Eq. (1b) the potentials $V_C(f,l)$ are functions supported on the elements (cliques) of the clique system $\mathscr{C}$ uniquely associated with the neighbourhood system $\mathscr{G}$ of the MRF. For an isotropic and homogeneous MRF it is

$$U(f,l)=\sum_r w_r V_r(f,l) \tag{2}$$

where each V_r is the sum of the local potentials over the set of homogeneous cliques and w_r is a parameter expressing the relative weight of these potentials. Depending on the particular problem, the order of the neighbourhood system and the potential functions should be chosen to capture the desired statistical relationship among the elements in the field. For instance, a typical model for piecewise smooth images must express the smoothness constraint on the pixel intensities, the conditions under which this constraint can be broken to create a discontinuity, and the self-interactions between line elements, so as to describe the admitted or forbidden local configurations of the discontinuities.

Assuming a linear model for the formation of the data and additive Gaussian noise with a known variance σ^2, the posterior probability is:

$$P(f,l|\mathbf{g})=\frac{1}{Z}\exp\left(-\frac{\|\mathbf{g}-H\mathbf{f}\|^2}{2\sigma^2}-U(f,l)\right) \tag{3}$$

where $\mathbf{f}$ is the vector of the lexicographically ordered notation [36] for f, $\mathbf{g}$ is the vector of the data, H is the matrix associated with a linear operator and Z is the normalizing constant. The set of all the parameters w_r constitutes herein the already mentioned hyperparameter vector $\mathbf{w}$. Assuming that their values are known, the MAP estimate is defined as the minimizer of the posterior energy function associated with the posterior probability Eq. (3), which is given by:

$$E(f,l|\mathbf{g})=\frac{\|\mathbf{g}-H\mathbf{f}\|^2}{2\sigma^2}+U(f,l) \tag{4}$$

Nevertheless, as already highlighted, the a priori information is usually not enough to precisely determine the values of the hyperparameters. On the other hand, these values are very critical for the quality of the reconstruction. The complexity of trial-and-error procedures for the selection of these values is very high, since they require a complete solution of the reconstruction problem at each attempt. Moreover, these parameters are strongly correlated. In previous works [37,38], we have found that, although the addition of new constraints on the line configurations always improves image quality, in many practical applications a line continuation constraint, which enforces hysteresis of the boundary edges, and a thinness constraint, which enforces non-maximum suppression, seem to be the most important ones. Neglecting some of the parameters greatly reduces the cost of trial-and-error procedures for their determination. Nevertheless, this entails a finer tuning of the remaining parameters, and reduces considerably the robustness of the methods with respect to small variations in them. For all these reasons, studying efficient techniques for parameter estimation plays a crucial role in image restoration through MRF models.

One natural way to deal with the problem of parameter estimation is to consider $\mathbf{w}$ itself as a variable of the reconstruction problem. By explicitly introducing the dependence of all the functions defined in Eqs. (1a), (1b), (2), (3), (4) from $\mathbf{w}$ [e.g. $U(f,l)\equiv U(f,l|\mathbf{w})$], a fully Bayesian approach to reconstruct the image and estimate the 'best' Gibbs parameters for a given problem is to assume a prior distribution for $\mathbf{w}$ and then simultaneously compute (f,l) and $\mathbf{w}$ by maximizing some joint distribution of both. Such a strategy should of course satisfy the requirement that the solution for the image field is still a MAP estimate. Assuming a uniform prior for $\mathbf{w}$, the distribution most adopted in the literature is $P(f,l,\mathbf{g}|\mathbf{w})$, which leads to a MAP estimate for (f,l) and an ML estimate for $\mathbf{w}$, based on the prior of (f,l). We adopt instead the posterior distribution $P(f,l|\mathbf{g},\mathbf{w})$, now regarded as a function of $\mathbf{w}$ too. On this basis, (f,l) and $\mathbf{w}$ should be given by the solution of the following problem:

$$\max_{f,l,\mathbf{w}} P(f,l|\mathbf{g},\mathbf{w}) \tag{5}$$

Since the joint maximization Eq. (5) is a very difficult task, we exploit the following sub-optimal iterative procedure:

$$(f^k,l^k)=\arg\max_{f,l} P(f,l|\mathbf{g},\mathbf{w}^k), \tag{6a}$$

$$\mathbf{w}^{k+1}=\arg\max_{\mathbf{w}} P(f^k,l^k|\mathbf{g},\mathbf{w}). \tag{6b}$$

Starting from an initial guess $\mathbf{w}^0$ and iterating steps Eq. (6a) and Eq. (6b), a sequence $(f^k,l^k,\mathbf{w}^k)$ which converges to a local maximum of $P(f,l|\mathbf{g},\mathbf{w})$ can be obtained. In this sense, problem Eqs. (6a), (6b) is weaker than problem Eq. (5). Nevertheless, if $(f^*,l^*,\mathbf{w}^*)$ is the solution of Eqs. (6a), (6b), then by definition (f^*,l^*) is the MAP estimate of (f,l) based on $\mathbf{g}$ and $\mathbf{w}^*$, and $\mathbf{w}^*$ is the ML estimate of $\mathbf{w}$ based on the posterior of the image computed in (f^*,l^*). The same considerations hold at each stage of the iterative procedure Eqs. (6a), (6b). Thus, to obtain $(f^*,l^*,\mathbf{w}^*)$ several MAP estimates for (f,l) and several ML estimates for $\mathbf{w}$ need to be computed.

To solve Eq. (6a) any of several existing algorithms can

be adopted, for example simulated annealing (SA), ICM, GNC or GEM, depending on the form of the posterior distribution. Here, owing to the presence of an explicit binary line process, we propose the mixed-annealing algorithm [39] in which the continuous variables are updated in a deterministic way, while the binary variables are updated in a stochastic way. This strategy is based on the fact that the energy $E(f,l|\mathbf{g},\mathbf{w}^k)$ is a convex function with respect to f, so that its minimum $f^*(l)$, for any fixed configuration of l, can be computed by using deterministic schemes, such as gradient descent techniques. The problem is thus solved by minimizing the function $E(f^*(l),l|\mathbf{g},\mathbf{w}^k)$ with respect to l. This minimization entails using simulated annealing techniques, but only with respect to the line variables, and hence with a greatly reduced complexity. In fact, as the neighbourhood size is generally modest and the line process is binary, the computational cost of drawing a sample of the line process is low.

The scheme of the annealing algorithm for the binary line variables alone consists of an order for visiting all the line elements, a rule for decreasing a temperature parameter T, and the time spent at each temperature, that is the length of the Markov chains. This scheme could be either sequential, where at each time a single line element is updated, or parallel, where all the line elements are updated independently of each other. In the latter case, however, asymptotic convergence is only ensured for asynchronous updates, i.e. independent but not simultaneous updates [1,18]. The line element $l_{i,j}$ at site (i,j), horizontal or vertical, is updated using the Gibbs sampler [1], according to the local conditional posterior:

$$P_T(l_{i,j}=1|l_{m,n}(m,n)\neq(i,j),\mathbf{g},\mathbf{w}^k)=\frac{1}{1+\exp[-\Delta E/T]} \quad (7)$$

where

$$\Delta E = E(f^*(l^1),l^1|\mathbf{g},\mathbf{w}^k)-E(f^*(l^0),l^0|\mathbf{g},\mathbf{w}^k) \quad (8)$$

and l^1 and 1^0 are line configurations which only differ in terms of the element $l_{i,j}$ which is '1' in the first and '0' in the second.

The practical implementation of the overall mixed-annealing scheme is still very expensive, because $f^*(l)$ must be computed whenever an update for a line element is proposed. In a previous paper [40], we showed that the availability of an analogue Hopfield-type neural network [41] would permit the computation of $f^*(l)$ in almost real time, thus making the mixed-annealing algorithm very efficient. Nevertheless, even in the absence of such an architecture, we found that, from an experimental point of view, good results can also be obtained by using an approximation of the theoretical scheme. This approximation is based on the local nature of the intensity process and on its usual short range interactions with the line process, so that, under the reasonable hypothesis that $f^*(l)$ undergoes localized, small variations for mild perturbations in l, $f^*(l)$ can be computed, for instance, only after one or more updates of the whole line process, or even once per each Markov chain. If the value for $f^*(l)$ is kept fixed along the updating of the line process at each given temperature, Eq. (8) reduces to:

$$\Delta E = U(f^*,l^1|\mathbf{w}^k)-U(f^*,l^0|\mathbf{w}^k) \quad (9)$$

which means that the Gibbs sampler draws samples according to the prior probability, $P_T(l|f^*,\mathbf{w}^k)$, of the lines conditioned on the intensity process, rather than according to the conditional posterior.

The mixed-annealing algorithm, whether theoretical or approximated, produces an estimate (f^k,l^k) of the image, according to the current value of $\mathbf{w}^k$. This will be used to solve problem Eq. (6b), i.e. the parameter estimation step.

The difficulty with Eq. (6b) is related to the presence in the posterior distribution of the normalizing constant which depends on $\mathbf{w}$. As already said, one common solution is to adopt pseudo-likelihood or mean field approximations [14] for this distribution. Nevertheless, it must be pointed out that these approximations have mostly been applied to problems where the image pixels can assume only a few values, and the line process is not considered, at least not explicitly. In our opinion, introducing the discontinuities into the MRF model considerably complicates the problem of parameter estimation, in that intensity and line elements are highly correlated, thus preventing the pseudo-likelihood from being a good approximation of the original distribution. In addition, considering a many-levels or even continuous-valued intensity field makes the computation of the pseudo-likelihood very heavy. We will thus investigate a different solution, which allows for a fast ML estimation of the parameters using the posterior distribution of the lines, conditioned on some given value for the intensities, rather than on the joint posterior in Eq. (6b). This approach will be derived and discussed in Section 3.

3. ML hyperparameter estimation

As stated in Eq. (6b), in this paper we propose to perform the ML estimation of $\mathbf{w}$ with respect to $P(l^k,f^k|\mathbf{g},\mathbf{w})$. As a function of $\mathbf{w}$, $P(l^k,f^k|\mathbf{g},\mathbf{w})$ is given by:

$$P(f^k,l^k|\mathbf{g},\mathbf{w})=\frac{P(\mathbf{g}|f^k,l^k,\mathbf{w})P(f^k,l^k|\mathbf{w})}{P(\mathbf{g}|\mathbf{w})} \quad (10)$$

where

$$P(\mathbf{g}|\mathbf{w})=\int_f\sum_l P(f,l,\mathbf{g}|\mathbf{w})\mathrm{d}f$$

$$=\int_f\sum_l P(\mathbf{g}|f,l,\mathbf{w})P(f,l|\mathbf{w})\mathrm{d}f \quad (11)$$

Substituting in Eqs. (10), (11) the expressions that holds

in our case for $P(\mathbf{g}|f,l,\mathbf{w})$ and $P(f,l|\mathbf{w})$ we get:

$$P(f^k, l^k|\mathbf{g}, \mathbf{w}) = \frac{\exp\left[-\frac{\|\mathbf{g} - H\mathbf{f}^k\|^2}{2\sigma^2}\right]\exp[-U(f^k, l^k|\mathbf{w})]}{\int_f \sum_l \exp\left[-\frac{\|\mathbf{g} - H\mathbf{f}^k\|^2}{2\sigma^2}\right]\exp[-U(f, l|\mathbf{w})]df} \quad (12)$$

Since it is

$$\arg\max_{\mathbf{w}} P(f^k, l^k|\mathbf{g}, \mathbf{w}) = \arg\min_{\mathbf{w}} -\log P(f^k, l^k|\mathbf{g}, \mathbf{w}) \quad (13)$$

we found it more convenient to reformulate the ML parameter estimation problem as the one that minimizes the negative log-posterior. This function can easily be shown to be convex with respect to $\mathbf{w}$, so that a possible criterion to get its minimum is to look for the zero of its gradient, i.e. by solving the following system of equations:

$$\frac{\partial}{\partial w_r}\left\{-\log P(f^k, l^k|\mathbf{g}, \mathbf{w})\right\} = \frac{\partial U(f^k, l^k|\mathbf{w})}{\partial w_r} + \frac{1}{Z(\mathbf{w})}\frac{\partial Z(\mathbf{w})}{\partial w_r}$$

$$= V_r(f^k, l^k) - E_{\mathbf{w}}[V_r(f, l)] = 0 \quad (14)$$

where the expectations are computed with respect to the posterior distribution $P(f,l|\mathbf{g},\mathbf{w})$ and $Z(\mathbf{w})$ is the partition function. Eq. (14) highlights a well known result in statistical mechanics and probability theory which states that all mean values of the system described by a Gibbs distribution can be obtained from the partition function.

Hence, a gradient descent technique could be applied to update the parameters vector $\mathbf{w}$, according to the following iterative scheme:

$$w_r^{t+1} = w_r^t - \eta\frac{\partial}{\partial w_r}\left\{-\log P(l^k, f^k|\mathbf{g}, \mathbf{w})\right\}|_{\mathbf{w}^t}$$

$$= w_r^t + \eta\left\{E_{\mathbf{w}^t}[V_r(f, l)] - V_r(f^k, l^k)\right\} \quad (15)$$

where η is a positive, small parameter, to be suitably chosen to ensure convergence. The physical meaning of this updating rule is to look for those parameters which make the various potentials, computed in the MRF realization (f^k,l^k), equal to the their averages, when (f,l) is allowed to assume all the possible configurations. In other words, the parameter updating rule aims to determine the probability under which the frequencies of occurrence of the various local configurations in (f^k,l^k) are equal to their statistical expectations. It is important to recognize in this updating rule the learning algorithm for the weights of a Generalized Boltzmann Machine, with mixed continuous and binary units distributed over a single layer. One particularity is that these units are associated with the data, the pixels and the line elements, but, while the units f and l are free to evolve in the unclamped phase and are kept fixed to the single example (f^k,l^k) in the clamped phase, the units associated with the data are always clamped to $\mathbf{g}$. Moreover, the interconnection weights among the data units and the pixel units are fixed, and related to the degradation model through the elements of matrix H.

Starting with $\mathbf{w}^0 = \mathbf{w}^k$, at the end of the iterative procedure Eq. (15) we would get the new estimate $\mathbf{w}^{k+1}$ of the parameters, to be used in Eq. (6a) for computing a new update (f^{k+1},l^{k+1}) of the image.

Nevertheless, the computational charge of the iterative scheme Eq. (15) as it stands is insurmountable, due to the computation of the expectations, or equivalently of the partition function, which requires summation over binary variables and integration over continuous variables. The summation with respect to the binary variables can be estimated, in the expectations, via a relatively cheap Gibbs sampler and, in the case of non-interacting lines, can even be performed analytically. On the contrary, there is no way to reduce the cost of the summation in f, so that some form of approximation needs to be adopted. Our approach to derive this approximation acts on the level of the partition function, and uses a well known results of the statistical mechanical theory, i.e. the saddle point approximation [42]. The expectations in the gradient are then computed by derivatives of the approximated partition function.

Let us consider first the following simplified form for the isotropic, homogeneous prior energy $U(f,l|\mathbf{w})$:

$$U(f, l|\lambda, \alpha) = \lambda\sum_{i,j}(f_{i,j} - f_{i+1,j})^2(1 - h_{i,j}) + \lambda\sum_{i,j}(f_{i,j} - f_{i,j+1})^2$$

$$\times(1 - v_{i,j}) + \alpha\sum_{i,j}h_{i,j} + \alpha\sum_{i,j}v_{i,j} \quad (16)$$

This choice for $U(f,l|\mathbf{w})$ corresponds to the weak membrane energy first proposed by Blake and Zissermann [4]. Prior Eq. (16) favours solutions which present discontinuities where the absolute value of the horizontal or vertical gradient is greater than the threshold value $\sqrt{\alpha/\lambda}$, and vary smoothly where there are no discontinuities. Considered individually, λ tunes the strength of the smoothness constraint, while α determines the cost to be paid whenever a discontinuity is created, thus preventing the creation of too many discontinuities. For the weak membrane model the partition function $Z(\lambda,\alpha)$ of the posterior of Eq. (12) is:

$$Z(\lambda, \alpha) = \int_f \sum_{h,v}\exp\Bigg[-\frac{\|\mathbf{g} - H\mathbf{f}\|^2}{2\sigma^2} - \lambda\sum_{i,j}(f_{i,j} - f_{i+1,j})^2$$

$$\times(1 - h_{i,j}) + (f_{i,j} - f_{i,j+1})^2(1 - v_{i,j})$$

$$- \alpha\sum_{i,j}(h_{i,j} + v_{i,j})\Bigg]df \quad (17)$$

In Eq. (17), the contribution of the line process can be computed exactly. Indeed, distributing the summation over h and v, and exploiting the independence of the single line elements from each other, one can integrate out the

variables h and v thus getting:

$$Z(\lambda,\alpha)=\int_f \exp\left[-\frac{\|\mathbf{g}-H\mathbf{f}\|^2}{2\sigma^2}\right] \times\prod_{i,j}\left\{\exp\left[-\lambda(f_{i,j}-f_{i+1,j})^2\right]+\exp(-\alpha)\right\} \times\prod_{i,j}\left\{\exp\left[-\lambda(f_{i,j}-f_{i,j+1})^2\right]+\exp(-\alpha)\right\}df$$

$$=\int_f \exp\left[-\frac{\|\mathbf{g}-H\mathbf{f}\|^2}{2\sigma^2}\right] \times\exp\left[\sum_{i,j}\log\left\{\exp\left[-\lambda(f_{i,j}-f_{i+1,j})^2\right]+\exp(-\alpha)\right\}\right] \times\exp\left[\sum_{i,j}\log\left\{\exp\left[-\lambda(f_{i,j}-f_{i,j+1})^2\right]+\exp(-\alpha)\right\}\right]df \tag{18}$$

which can be interpreted as the partition function of a new posterior distribution depending on f only. Geiger and Girosi [5] call the Gibbsian prior energy associated with this distribution the effective potential. Indeed, it describes the effect of the interaction of the line field with the intensity field, by suitably modifying the interaction of the field f with itself. To overcome the need for integrating with respect to f, one could argue that a good approximation for Z is given by the value which is predominant in the function to be integrated. In fact, by neglecting the statistical fluctuations of the field f, the partition function of Eq. (18) can be approximated through the saddle point approximation [42] as follows:

$$Z(\lambda,\alpha) \approx C\exp\left[-\frac{\|\mathbf{g}-H\hat{\mathbf{f}}\|^2}{2\sigma^2}\right] \times\exp\left[\sum_{i,j}\log\left\{\exp\left[-\lambda(f_{i,j}-\hat{f}_{i+1,j})^2\right]+\exp[-\alpha]\right\}\right] \times\exp\left[\sum_{i,j}\log\left\{\exp\left[-\lambda(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2\right]+\exp[-\alpha]\right\}\right] \tag{19}$$

where C is a constant and $\hat{f}$ satisfies

$$\nabla\left\{\frac{\|\mathbf{g}-H\mathbf{f}\|^2}{2\sigma^2} -\sum_{i,j}\log\left[\exp(-\lambda(f_{i,j}-f_{i+1,j})^2)+\exp(-\alpha)\right] -\sum_{i,j}\log\left[\exp(-\lambda(f_{i,j}-f_{i,j+1})^2)+\exp(-\alpha)\right]\right\}=0 \tag{20}$$

Assuming this approximation for the partition function, the explicit forms for the expectations appearing in the gradient of Eq. (14) are:

$$E_{\lambda,\alpha}\left[\sum_{i,j}(f_{i,j}-f_{i+1,j})^2(1-h_{i,j})+\sum_{i,j}(f_{i,j}-f_{i,j+1})^2(1-v_{i,j})\right] =-\frac{1}{Z(\lambda,\alpha)}\frac{\partial Z(\lambda,\alpha)}{\partial\lambda} \approx\sum_{i,j}\frac{(\hat{f}_{i,j}-\hat{f}_{i+1,j})^2\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i+1,j})^2]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i+1,j})^2]+\exp[-\alpha]} +\sum_{i,j}\frac{(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2]+\exp[-\alpha]} \tag{21a}$$

and

$$E_{\lambda,\alpha}\left[\sum_{i,j}(h_{i,j})+\sum_{i,j}(v_{i,j})\right]=-\frac{1}{Z(\lambda,\alpha)}\frac{\partial Z(\lambda,\alpha)}{\partial\alpha}\approx \sum_{i,j}\frac{\exp[-\alpha]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i+1,j})^2]+\exp[-\alpha]} +\sum_{i,j}\frac{\exp[-\alpha]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2]+\exp[-\alpha]} \tag{21b}$$

respectively. It is straightforward to observe that these expectations can be considered as computed with respect to the conditional posterior $P(l|\hat{f},\mathbf{g},\lambda,\alpha)$, which coincides with the conditional prior $P(l|\hat{f},\lambda,\alpha)$. In particular, from Eq. (21b), it results that the two terms:

$$\bar{h}_{i,j}=\frac{\exp[-\alpha]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i+1,j})^2]+\exp[-\alpha]} \tag{22a}$$

and

$$\bar{v}_{i,j}=\frac{\exp[-\alpha]}{\exp[-\lambda(\hat{f}_{i,j}-\hat{f}_{i,j+1})^2]+\exp[-\alpha]} \tag{22b}$$

represent the mean value for the variables $h_{i,j}$ and $v_{i,j}$, respectively, under the same probability.

Geiger and Girosi [5] applied the above saddle point

approximation to derive a set of deterministic equations for the solution of an image reconstruction problem from sparse and noisy data. In particular, they proposed a gradient descent method as a fast, parallel and iterative scheme to obtain the solution of Eq. (20).

In our application, as $\hat{f}$ depends on $\mathbf{w}$, Eq. (20) should in principle be solved for each step of the parameter updating rule of Eq. (15). A further reduction in the computational charge can be obtained by assuming that mild changes in the values of the parameters produce slight modifications in $\hat{f}$, so that this can be kept fixed along scheme Eq. (15) to the value corresponding to $\mathbf{w}^k$. A second, possible level of approximation is thus based on assuming $\hat{f} \approx f^k$. This assumption is justified by the fact that, when $\hat{f}$ is computed with respect to $\mathbf{w}^k$, as in our case now, it actually converges to f^k in the limit to zero of a temperature parameter T to be inserted as usual in the posterior probability [5]. This argument is the basis for the convergence of a practical algorithm, described in Section 4, where this approximation is used in the context of a mixed-annealing for MAP restoration, periodically interrupted to update the parameters according to rule Eq. (15).

The use of the saddle point approximations described above leads to the result that the estimation is actually based on the conditional prior $P(l^k|f^k,\lambda,\alpha) \equiv P(l^k|f^k,\mathbf{g},\lambda,\alpha)$ rather than on the joint posterior $P(f^k,l^k|\mathbf{g},\lambda,\alpha)$. This can be easily understood by verifying that the gradient of the negative logarithm of the conditional prior with respect to λ and α is identical to the one obtained by considering Eqs. (21a), (21b) with $\hat{f} \equiv f^k$. It is straightforward to observe that in this case the practical computation of the expectations is very fast, since f^k is already available from step Eq. (6a). Hence, the computational cost of the iterative scheme of Eq. (15) is very low.

This result can be extended to the general case of an MRF model where the line process is left to self-interact, in order to constrain the local geometry of the discontinuities. In this case, the line process can no longer be integrated out from the partition function. Nevertheless, the expectations appearing in the gradient can be computed, under the conditional prior $P(l|f^k,\mathbf{w})$, through time averages, by means of a Gibbs sampler that is still cheap. In this general case the hyperparameter updating rule of Eq. (15) becomes:

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \eta \nabla(-\log P(l^k|f^k,\mathbf{w}))|_{\mathbf{w}^t} \tag{23}$$

As a remark, the purpose for the parameter updating rule is now the search for that prior probability under which the potentials computed in (f^k,l^k) are equal to the corresponding potentials computed in $(f^k,\bar{l})$, where $\bar{l}$ is the expected value of the line process for the given intensity f^k. In the formalism of the Boltzmann Machine this is equivalent to a learning algorithm from the single example (f^k,l^k), where the input units, i.e. the intensities, are kept fixed to the environment both in the clamped and unclamped phase. This special case of learning is referred as "strict classification without hidden units" [21].

So far, we have shown that the explicit presence in the MRF model of a line process enables alternative criteria to be used to define a hyperparameter estimator which can be implemented through very fast and efficient algorithms. In fact, the approximations derived above can be interpreted as different choices for the probability distribution upon which to base the estimation. From a mathematical point of view, these choices are justified by the convergence properties of the estimators themselves, due to the concavity of the related log-probabilities. The physical point of view is that the fluctuation of l, with respect to sample l^k and conditioned on some relevant value for f, are assumed to give almost the same information about the parameters as the fluctuations of both fields around (f^k,l^k).

4. The overall practical procedure

The exact implementation of the overall MAP–ML estimation problem defined by Eqs. (6a), (6b) would require the iterative application of steps Eq. (6a) and Eq. (6b) in sequence. Step Eq. (6a) consists of the MAP estimate of the image field (f, l) given $\mathbf{g}$ and $\mathbf{w}^k$. As seen in Section 2, this can be performed by a mixed-annealing algorithm, carried out for a decreasing sequence of temperatures. During each iteration of the mixed-annealing algorithm the temperature is kept fixed at its current value and two main tasks are executed, based on the posterior distribution

$$P_T(f, l|\mathbf{g}, \mathbf{w}^k) = \frac{1}{Z} \exp\left[\frac{-E(f, l|\mathbf{g}, \mathbf{w}^k)}{T} \right] \tag{24}$$

i.e. the updating of the line elements and the minimization of the related posterior energy with respect to f. The line elements are updated according to the Gibbs sampler; all the line sites are repeatedly visited and their values are set by drawing samples from the local prior distributions. Due to the small size of the neighbourhood system adopted and the fact that the line elements are binary, the Gibbs sampler can be considered a cheap algorithm in this case. Due to the convexity in f of the posterior energy and the local nature of the intensity field, the minimization of $E(f,l|\mathbf{g},\mathbf{w}^k)$ can be performed deterministically, through parallel algorithms such as Successive Over Relaxation (SOR) [4]. Nevertheless, in our case this minimization constitutes the most expensive part of step Eq. (6a), one reason being that it has to be performed at each of a usually very long sequence of temperatures. Step Eq. (6b) consists of the ML estimation of the hyperparameters $\mathbf{w}$, given (f^k,l^k). As seen in Section 3, this can be performed by a gradient descent algorithm (Eq. 23) which, at each iteration, entails using a relatively cheap Gibbs sampler to drawn samples of the binary line process from the conditional prior $P(l|f^k,\mathbf{w})$ and create a Markov chain to compute the time averages.

It is worth noting that the iterative algorithms, both deterministic and stochastic, used in the steps Eq. (6a) and Eq. (6b), can be made fully parallel. From a practical point of

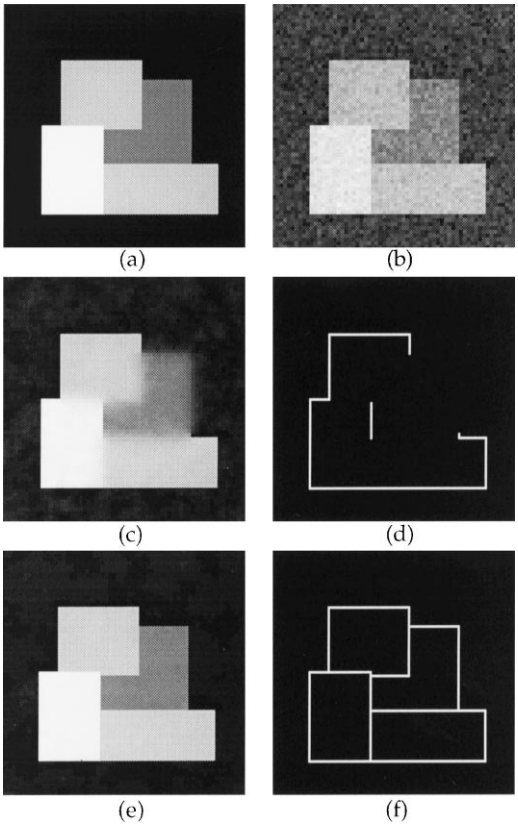

Fig. 1. 64 × 64 synthetic image: (a) piecewise constant original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 10$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.01$, $\alpha = 25$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.01$, $\alpha_i = 25$; final parameters: $\lambda_f = 0.035$, $\alpha_f = 1.075$); (f) related edge map.

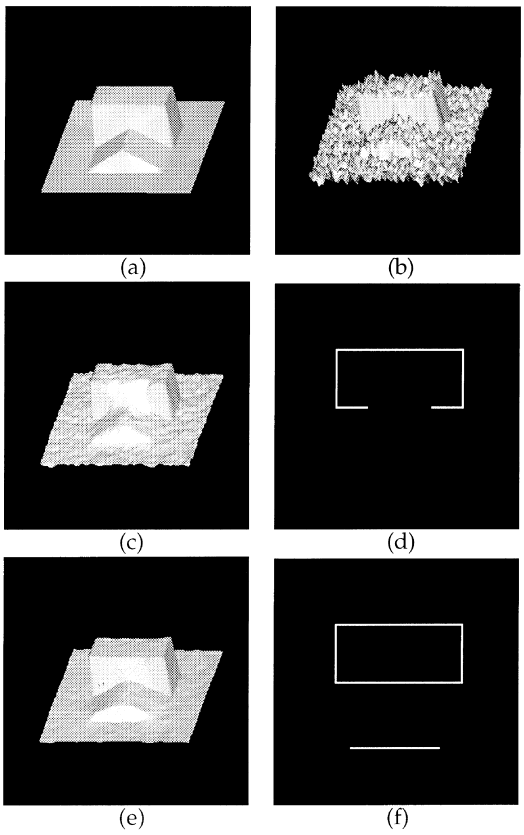

Fig. 2. 64 × 64 synthetic image: (a) piecewise planar original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 10$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.01$, $\alpha = 25$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.01$, $\alpha_i = 25$; final parameters: $\lambda_f = 0.04$, $\alpha_f = 2.4$); (f) related edge map.

view, we derived a procedure in which an external mixed-annealing algorithm for the MAP reconstruction of (f,l) is periodically interrupted in order to compute a new set $\mathbf{w}$ of parameters. These parameters are obtained by a single step of the iterative scheme Eq. (23), so as to increase the value that the probability assumes on the current image estimate. The scheme of the procedure is the following:

1. set $k = 0$, $\mathbf{w}^k$, (f^k,l^k), T_k
2. set $t = 1$, $l^t = l^k$; $f^t = f^k$

   for $t = 1, L$ execute:

   a. $f^{t+1} = \arg\min_f E(f,l^t|\mathbf{g},\mathbf{w}^k)$
   b. compute l^{t+1} according to $P_{T_k}(l|f^{t+1},\mathbf{w}^k)$
3. set $f^{k+1} = f^L$, $l^{k+1} = l^L$
4. $\mathbf{w}^{k+1} = \mathbf{w}^k - \eta\nabla(-\log P_{T_k}(l^{k+1}|f^{k+1},\mathbf{w}))|_{\mathbf{w}^k}$
5. set $k = k + 1$; go to step 2 until a termination criterion is satisfied

The convergence of the above procedure is ensured since, as T_k approaches zero, $P_{T_k}(l|f^k,\mathbf{w}^k)$ becomes peaked around l^k, and hence the gradient in step 4 approaches zero as well. This scheme is an approximated version of the iterative scheme proposed in Eqs. (6a), (6b), since, as described in Section 3, for updating $\mathbf{w}$ (step 4) a saddle point approximation of the posterior distribution has been adopted, and it has been assumed $\hat{f} \approx f^k$. However, it is well known that, as T_k goes to zero, the contribution to the partition function due to $\hat{f}$ becomes more and more predominant, so that the saddle point approximation becomes better and better as the algorithm proceeds. Moreover, it has been shown [5] that $\hat{f}$ and f^k both converge to the MAP estimate, and so f^k approaches $\hat{f}$ as well. As in the case of the scheme Eqs. (6a), (6b), the procedure above produces a sequence $(f^k,l^k,\mathbf{w}^k)$ converging to a final point $(f^*,l^*,\mathbf{w}^*)$, where (f^*,l^*) is the MAP estimate of (f,l) given $\mathbf{g}$ and $\mathbf{w}^*$. On the basis of the considerations above reported, $\mathbf{w}^*$ can be considered a good approximation of the ML estimate of $\mathbf{w}$, based on the posterior of the image computed in (f^*,l^*). The good quality of this approximated estimate is confirmed by the experimental results, as shown in Section 5. These results also highlight the strong reduction of the computational costs of the procedure when compared with the exact implementation of the scheme Eqs. (6a), (6b).

5. Experimental results

To evaluate the performance of the MAP–ML procedure

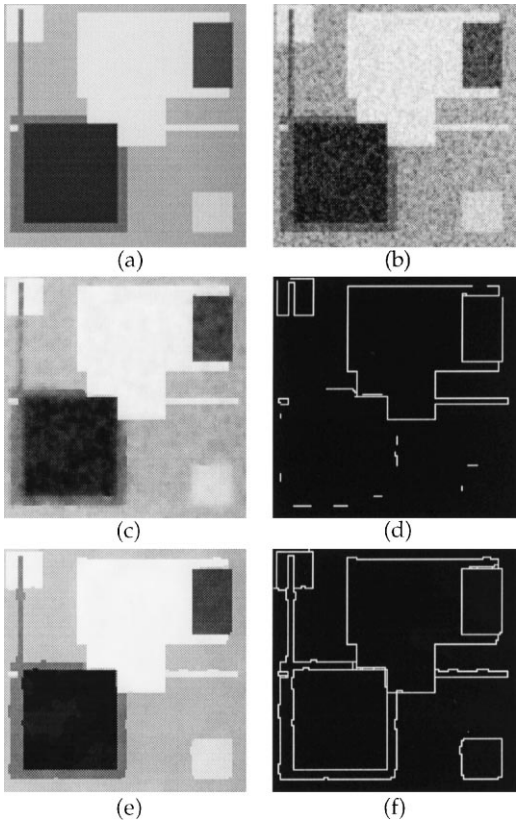

Fig. 3. 128 × 128 synthetic image: (a) piecewise constant original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 25$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0016$, $\alpha = 4$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0016$, $\alpha_i = 4$; final parameters: $\lambda_f = 0.014$, $\alpha_f = 0.39$); (f) related edge map.

described in Section 4 we tested it on a serial computer on synthetic and real images belonging to the class of piecewise smooth images. As synthetic images we considered a piecewise constant 'Mondrian' image (Fig. 1(a)) and a piecewise planar image (Fig. 2(a)), both of size 64 × 64 over 256 grey levels, artificially degraded by adding Gaussian uncorrelated noise with zero mean and three different levels of standard deviation, i.e. $\sigma = 10, 20, 30$, respectively. For comparison with the results obtained on a previous work [38], we also considered another piecewise constant 'Mondrian' image (Fig. 3(a)), of size 128 × 128, degraded by Gaussian noise with standard deviation $\sigma = 25$. As real images we used two 128 × 128 images of printed characters, which can be also roughly considered piecewise constant, and a 170 × 290 portion of a 512 × 512 image of 'Lena', which can be roughly considered piecewise smooth. The first one was degraded by randomly selecting 50% of the original image and adding Gaussian noise with $\sigma = 12$ (Fig. 4(a)); the second one (Fig. 5(a)) was degraded by adding Gaussian noise with standard deviation $\sigma = 25$. Image 'Lena' (Fig. 6(a)) was degraded by adding Gaussian noise with $\sigma = 20$.

Although the MAP–ML procedure has been established for the general case of an image model with self-interacting

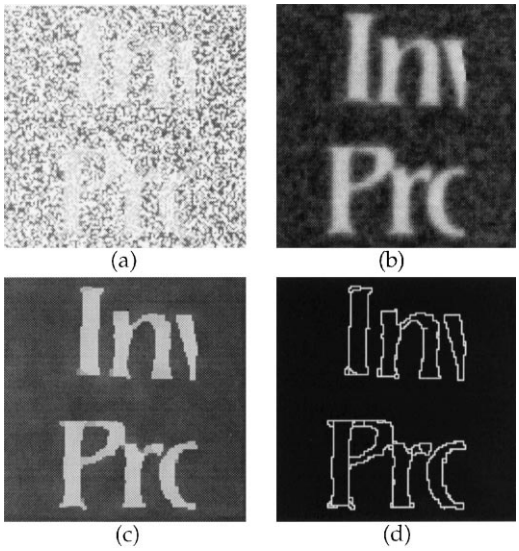

Fig. 4. 128 × 128 real image: (a) image degraded by randomly removing the 50% of pixels and adding uncorrelated, Gaussian noise ($\sigma = 12$); (b) MAP reconstruction with fixed parameters ($\lambda = 0.007$, $\alpha = 17.36$); (c) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.007$, $\alpha_i = 17.36$; final parameters: $\lambda_f = 0.046$, $\alpha_f = 2.08$); (d) related edge map.

explicit line processes, in this first application we restricted ourselves to the simpler case of the weak membrane model, characterized by a prior energy of the form given in Eq. (16). Thus the vector of the Gibbs parameters **w** reduces to parameters λ and α alone.

In all the experiments, we adopted $\lambda_i = 2/2\sigma^2$ and $\alpha_i = 5000/2\sigma^2$ as starting values for the parameters, which give an initial threshold $\sqrt{\alpha/\lambda} = 50$. The initial value of temperature was $T_0 = 1500$, and the law for decreasing it was chosen to be exponential, according to:

$$T_{k+1} = 0.9T_k$$

At each temperature we performed $L = 10$ updates of the image field, according to steps 2a and 2b of the algorithm, while the length of the Markov chains for updating the line elements in step 2b was chosen to be $2N(N - 1)$, corresponding to a single visitation of each line element in turn. Indeed, a single complete update is sufficient, because, owing to the independence of the line elements, the Gibbs sampler is memoryless. The minimization of the posterior energy with respect to *f*, in step 2a, was executed by means of a conjugate gradient algorithm.

The single-step update of each hyperparameter (step 4) was executed by decreasing its current value of a variable percentage of the gradient of the negative conditional log-prior, where the percentage value η is computed as the ratio between the gradient itself and the clamped expectation.

5.1. Synthetic images

In a first set of experiments we tested the performance and the convergence rate of the method on the two synthetic

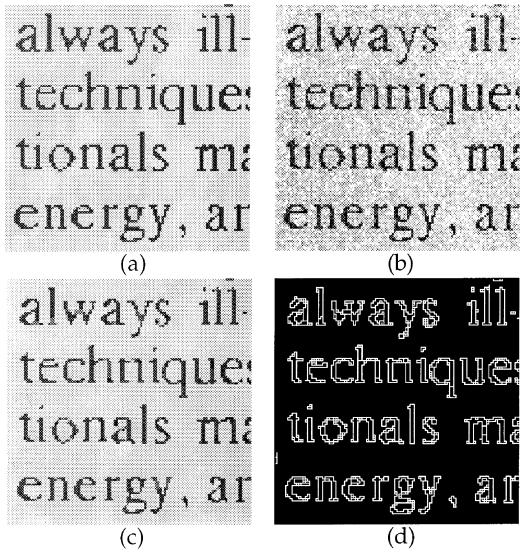

Fig. 5. 128 × 128 real image: (a) original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 25$); (c) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0016$, $\alpha_i = 4$; final parameters: $\lambda_f = 0.0078$, $\alpha_f = 0.69$); (d) related edge map.

images of Fig. 1(a) and Fig. 2(a), degraded by uncorrelated Gaussian noise with different standard deviations. The results obtained for the piecewise constant 'Mondrian' and for the piecewise planar image, when degraded by a noise with $\sigma = 10, 20, 30$, are shown in Figs. 1, 7 and 8 and in Figs 2, 9 and 10, respectively. For comparison purposes, for each degraded image we also computed the reconstruction corresponding to the initial values of the parameters. These reconstructions were obtained by running the mixed-annealing alone, at the same conditions of initial temperature, annealing schedule and Markov chain length.

From a qualitative analysis of the results obtained, the performance of our method can be considered quite satisfactory. In particular, while the reconstructions obtained with the initial values of the parameters are, at the same time, excessively rough inside the homogeneous areas and excessively smooth across them, due to a too low smoothness parameter and a too high threshold, the final reconstructions globally show a nice recovery of the smooth areas, and an almost perfect reconstruction of the related edges. Indeed, the edge curves are all connected, while the few defects that are observable in the cases of higher noise level ($\sigma = 20, 30$) are essentially related to a lack of straightline continuity and thinness. This is not surprising, if we think that the weak membrane model adopted herein for the images does not enforce any constraints on the line geometry. In fact, although the weak membrane model is generally sufficient to describe piecewise constant images, there are practical situations where the particular noise realization, or the very low value of the smallest jump in the true image, do not allow for any couple (λ,α) to give a perfect edge map. This is particularly evident in the regions of the image where the values of the jumps go to zero or for planar

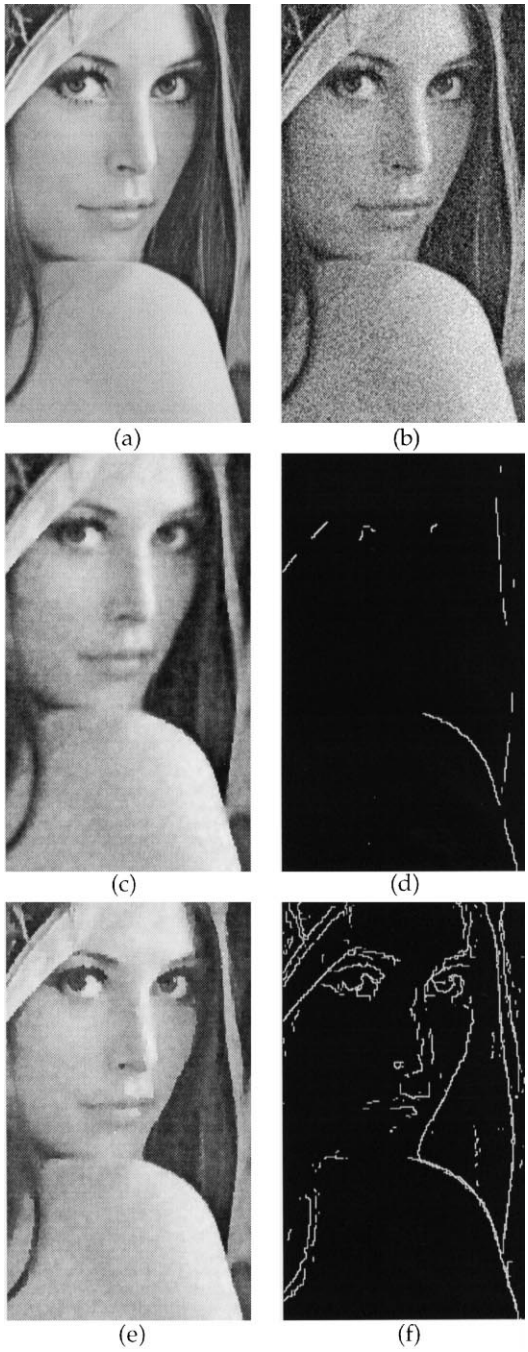

Fig. 6. 170 × 290 real image: (a) original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 20$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0025$, $\alpha = 6.25$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0025$, $\alpha_i = 6.25$; final parameters: $\lambda = 0.0019$, $\alpha_f = 0.97$); (f) related edge map.

images (see Fig. 2(a)). In those regions the reconstruction is not correct. In these cases, the aid of extra constraints on the regularity of the line configurations, such as straightline connectivity and double line penalization, becomes necessary. Further improvements can be obtained also by using more complex models including higher order derivatives and/or implicitly addressed graded discontinuities [43].

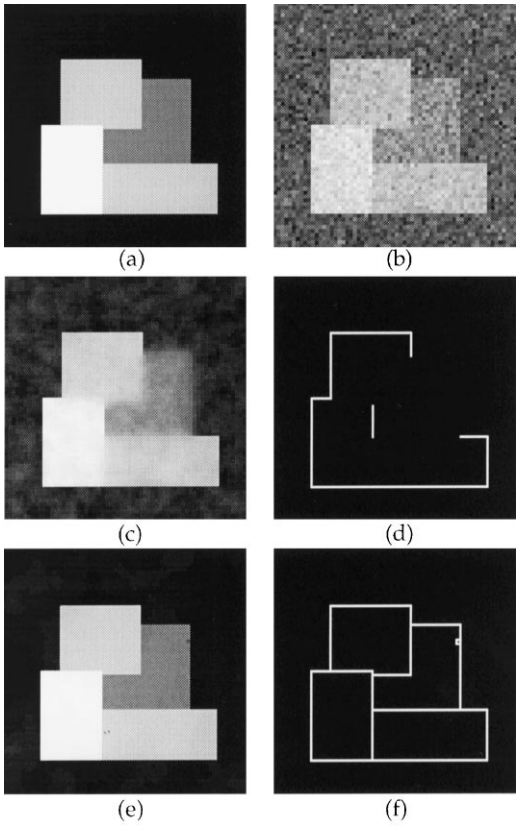

Fig. 7. 64 × 64 synthetic image: (a) piecewise constant original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 20$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0025$, $\alpha = 6.25$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0025$, $\alpha_i = 6.25$; final parameters: $\lambda_f = 0.029$, $\alpha_f = 0.7$); (f) related edge map.

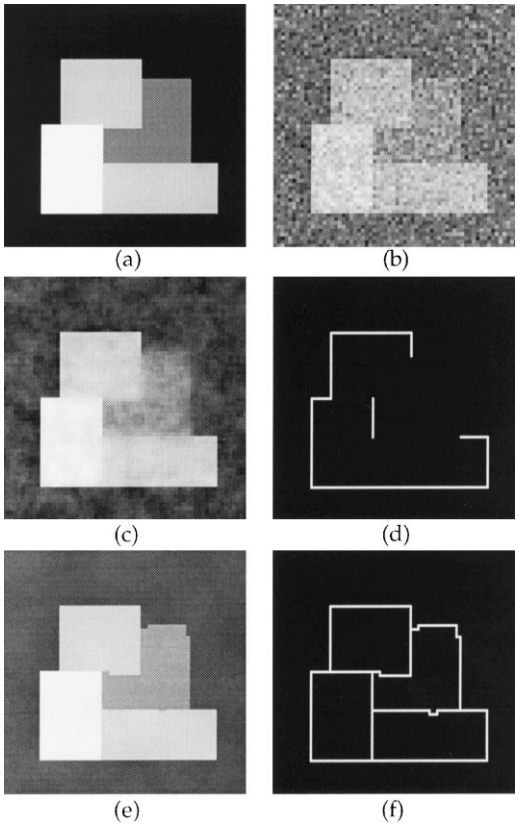

Fig. 8. 64 × 64 synthetic image: (a) piecewise constant original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 30$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0011$, $\alpha = 2.78$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0011$, $\alpha_i = 2.78$; final parameters: $\lambda_f = 0.025$, $\alpha_f = 0.65$); (f) related edge map.

Another effect, still due to a poor image modelization, is the slight flattening of the junction between the two planar patches with different slopes in the image of Fig. 2(a), which is easily perceivable in the perspective view provided. This image feature is not a true intensity edge, in that it does not correspond to a discontinuity in the intensity values. Instead, it is mathematically defined as a discontinuity of the intensity first derivative. To preserve it a local smoothness constraint should be recursively enforced on the line process itself.

The final values of the parameters obtained in these six examined cases are reported in Table 1 for the 'Mondrian' image and in Table 2 for the planar image; a typical plot of the values of the parameters versus the number of iterations is shown in Fig. 11. In all cases, convergence to the final values and stabilization of the reconstruction were reached in less than 30 iterations of the whole procedure, corresponding to a final value of the temperature around 70. The study of the Tables allows for a quantitative analysis of the convergence properties of the method. According to the MAP–ML principle, the final reconstructions should be a sample of the MRF corresponding to the given prior and the final parameters. In this sense, we should expect the parameters to converge to the same values for a given original image and almost independently of the degradation level and the initial values. Thus, for the λ parameters, starting with very different initial values, ranging from 0.0011 to 0.01 in the three noise level cases, we obtained final values ranging from 0.015 to 0.04, for both images. Similarly, for the α parameters, starting with initial values ranging from

Table 1
Initial and final values of the hyerparameters for the experiments shown in Figs. 1, 7 and 8

noise	λ_i	λ_f	α_i	α_f
$\sigma = 10$	0.01	0.035	25	1.075
$\sigma = 20$	0.0025	0.029	6.25	0.7
$\sigma = 30$	0.0011	0.025	2.78	0.65

Table 2
Initial and final values of the hyperparameters for the experiments shown in Figs. 2, 9, and 10

noise	λ_i	λ_f	α_i	α_f
$\sigma = 10$	0.01	0.04	25	2.4
$\sigma = 20$	0.0025	0.028	6.25	1.37
$\sigma = 30$	0.0011	0.015	2.78	0.63

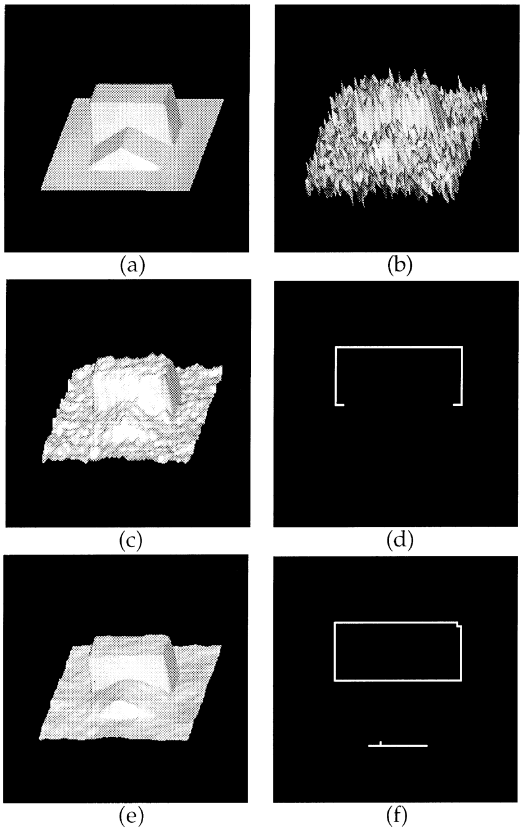

Fig. 9. 64 × 64 synthetic image: (a) piecewise planar original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 20$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0025$, $\alpha = 6.25$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0025$, $\alpha_i = 6.25$; final parameters: $\lambda_f = 0.028$, $\alpha_f = 1.37$); (f) related edge map.

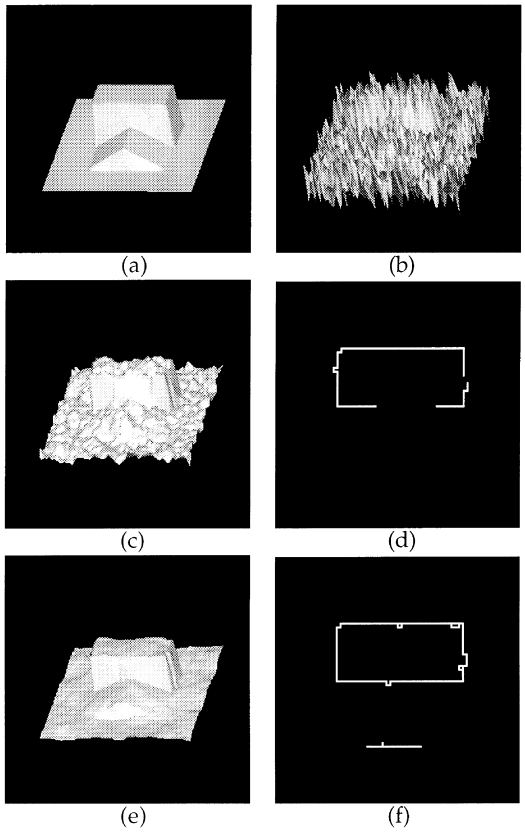

Fig. 10. 64 × 64 synthetic image: (a) piecewise planar original image; (b) image degraded by adding uncorrelated, Gaussian noise ($\sigma = 30$); (c) MAP reconstruction with fixed parameters ($\lambda = 0.0011$, $\alpha = 2.78$); (d) related edge map; (e) MAP–ML reconstruction (initial parameters: $\lambda_i = 0.0011$, $\alpha_i = 2.78$; final parameters: $\lambda_f = 0.015$, $\alpha_f = 0.63$); (f) related edge map.

2.78 to 25, we obtained final values ranging from 0.63 to 2.4. However, in all our simulations we found that the obtained values decrease slightly when the values of the noise standard deviation increase. We are not able to decide if that is due to the random variability of the estimates or if it is intrinsically connected to the method. Nevertheless, it is to be noted that the average variability of the reconstructions computed inside the homogeneous zones (which is governed by the λ parameter), as well as the total number of edges (which is governed by the α parameters), remain almost constant for a given image in the three noise cases.

As a different example of synthetic image we considered the piecewise constant 'Mondrian' image shown in Fig. 3(a), degraded by uncorrelated Gaussian noise with $\sigma = 25$. We again reached the convergence in less than 30 iterations with final values $\lambda = 0.014$ and $\alpha = 0.39$. The very small value of the α parameter obtained in this case is related to the very small jump between the background and the square in the bottom-right corner of the image. From an inspection of the reconstructed edge map, we can see that the boundaries are all complete and close, although not well behaved. Indeed, it is possible to observe, besides many deviations from the straightline, some unwanted double parallel lines. This is a clear consequence of the low α value. In another paper [38], we observed that, for this image, the recovery of a well behaved edge map is only possible if the weak membrane model is augmented by a straight-line line continuity constraint. In that case, the best values for the λ and α parameters that we found by trial-and-error were, of course, different from those obtained here, due to the presence of the line propagation factors ε.

5.2. Real images

The results obtained for the two real images of printed characters are shown in Fig. 4 and in Fig. 5, respectively. For display purposes, in Fig. 4(a) the missing pixels are replaced by white pixels. In both cases convergence to the values $\lambda = 0.046$, $\alpha = 2.08$ for the first image, and $\lambda = 0.0078$, $\alpha = 0.69$ for the second one, was again reached in less than 30 iterations. The results obtained gave rise to interesting observations. Let us consider first Fig. 4. In this case, the degration, besides the addition of white Gaussian noise with standard deviation $\sigma = 12$, consists of randomly removing 50% of the pixels. Thus the restoration problem becomes here a problem of noise removal plus interpolation of the missing data. The good quality of the reconstruction clearly indicates that the use of even very

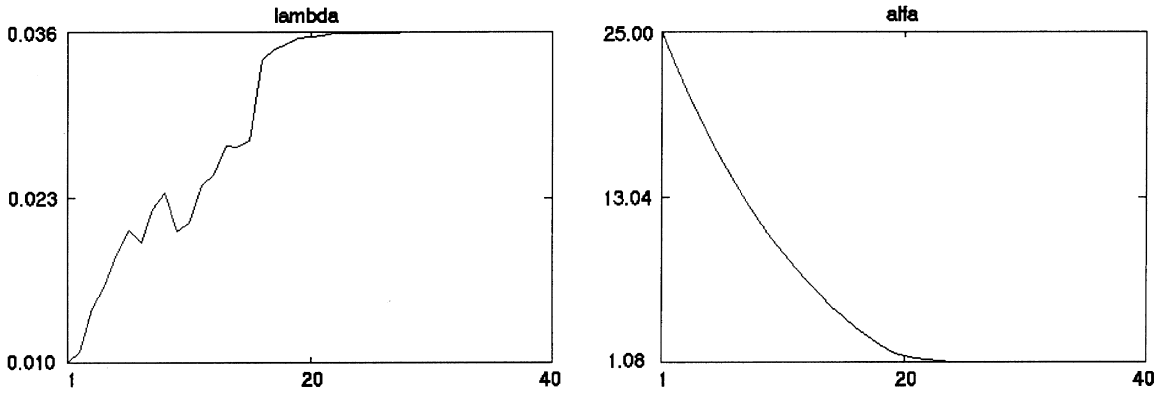

Fig. 11. Graphical behaviour of the two parameters versus the number of iterations, for the experiment in Fig. 1. Left: the λ parameter; right: the α parameter.

simple MRF models for the image allows for an excellent solution of the interpolation problem, even when the missing data are a significant percentage of the pixels. Thus the true problem here is again the noise removal. We can observe that, both from the point of view of the image model (piecewise constant), and with respect to the amount of degradation, this experiment is similar to that shown in Fig. 1. Indeed, we obtained for the two experiments very similar values of the smoothness degree (0.035 and 0.046). The α values are instead different in the two cases, higher in the second one, where the jump between the printed characters and the background is higher than the minimum jump in the 'Mondrian' image. The artefacts in the edge map of Fig. 4(d) are clearly related to the presence of very fine details in the various characters. Finally, as shown in Fig. 4(b), the reconstruction obtained with the initial parameter values almost shows no edge detected. This is in accordance with the high value of the threshold used. With respect to the second image of printed characters (Fig. 5), we can observe that the very low value of the final smoothness parameter ($\lambda = 0.0078$), as compared with the amount of noise ($\sigma = 25$), is due to the very fine scale of the characters. Despite of the obvious presence of many spurious edges, the quality of the reconstruction shown in Fig. 5(c) is excellent. We again emphasize that better results for the edge maps, with consequent further improvement of the intensity maps, could be obtained using image models which incorporate constraints on the line configuration geometry.

In a last experiment, we considered the more realistic image 'Lena', degraded by the addition of Gaussian noise with standard deviation $\sigma = 20$ (see Fig. 6(b)). Although this image does not exactly fit the weak membrane model, which is more suitable for piecewise constant images rather than for piecewise smooth images, we obtained the satisfactory results shown in Fig. 6(e)–(f). The final values of the parameters, after 25 iterations, are $\lambda = 0.0019$ and $\alpha = 0.97$. These values are well balanced, since they allow for a sufficient removal of the noise and a good detection of the most salient edges. In particular, the very low value of λ prevents the reconstruction from being too stylized, in accordance with the fact that the true image does not contain large constant patches, and is slowly varying. The same considerations already made about possible actions to be taken in order to further improve the edge map hold also in this case. Finally, it is to be noted the poor performance of the initial set of parameters (Fig. 6(c)–(d)).

All the computations were executed on the IBM systems RISC6000. For the 64×64 images the CPU computation time was about 30 seconds for 30 iterations; for the 128×128 images the times were around 120 seconds. We found that the number of iterations for all the experiments were approximately the same needed for a typical mixed-annealing procedure with known parameters. Thus the computational cost of the whole MAP–ML method can be considered equivalent (apart from the cheap step 4) to that of a simpler mixed-annealing algorithm for MAP estimation, with the advantage that the optimal values of the parameters do not need to be known in advance.

6. Conclusions

We have proposed an iterative procedure for the unsupervised MAP restoration of images, when piecewise smooth MRF models with explicit and constrained line processes are considered. The method consists of an external mixed-annealing algorithm for the maximization of the posterior distribution with respect to the image field, periodically interrupted to compute a new set of parameters. The new parameters are computed by ML estimation based on the conditional prior distribution of the lines given the intensities, evaluated on the current estimate of the whole image field. We have shown that in the case of the weak membrane model, this estimator corresponds to an ML estimator based on the joint posterior of intensities and lines, when a saddle point approximation for the partition function is taken. Our approach allows the computation of the expectations involved in the ML estimation to be reduced to sampling with respect to the binary line elements alone, or, in the case of a weak membrane model, to analytical calculations. In both cases, the reduction of the computational complexity, with respect to maximum pseudo-likelihood methods, is

significant, and makes the method very fast and cuts the cost down to almost a single MAP restoration. This computational saving was confirmed by our experiments, and the quality of the reconstructions obtained for the sub-case of non-interacting discontinuities is encouraging.

References

[1] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Patt. Anal. Mach. Intell. 6 (1984) 721–740.

[2] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill-posed problems in computational vision, J. Am. Stat. Ass. 82 (1987) 76–89.

[3] S. Kirkpatrick, C.D. Gellatt, M.P. Vecchi, Optimisation by simulated annealing, Science 220 (1983) 671–680.

[4] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, Cambridge, USA, 1987.

[5] D. Geiger, F. Girosi, Parallel and deterministic algorithms for MRFs: surface reconstruction, IEEE Trans. Patt. Anal. Mach. Intell. 13 (1991) 401–412.

[6] G.L. Bilbro, W.E. Snyder, Applying mean field annealing to image noise removal, J. Neural Network Comput. Fall, (1990) 5–17.

[7] J. Besag, On the statistical analysis of dirty pictures, J. R. Statist. Soc. ser. B 48 (1986) 259–302.

[8] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum Likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc. ser. B 39 (1977) 1–38.

[9] T. Hebert, R. Leahy, A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors, IEEE Trans. Med. Imag. 8 (1989) 194–202.

[10] L. Bedini, E. Salerno, A. Tonazzini, Edge-preserving tomographic reconstruction from Gaussian data using a Gibbs prior and a generalized Expectation-Maximization algorithm, Int. J. Imag. Systems Technol. 5 (1994) 231–238.

[11] G. Winkler, Image analysis, random fields and dynamic Monte Carlo methods: a mathematical introduction, Springer-Verlag, Berlin, 1995.

[12] B. Gidas, Parameter estimation for Gibbs distributions from fully observed data, in: R. Chellappa and A. Jain (Eds.), Markov Random Fields Theory and Application, Academic Press, San Diego, 1993, pp. 471–498.

[13] L. Younes, Estimation and annealing for Gibbsian fields, Ann. Inst. Henri Poincare 24 (2) (1988) 269–294.

[14] S.Z. Li, Markov Random Field Modeling in Computer Vision, Springer, Tokyo, 1995.

[15] G.R. Cross, A.K. Jain, Markov random field texture models, IEEE Trans. Patt. Anal. Mach. Intell. 5 (1983) 25–39.

[16] H. Derin, S. Guler, Realizations and parameter estimation for line processes, Proc. Int. Conf Acoust. Speech Signal. Process. 90 IV (1990) 2213–2216.

[17] L. Bedini, X. Qiao, A. Tonazzini, A generalized Boltzmann machine for learning Gibbs priors in edge-preserving image restoration, IEI Technical Report B4–55, 1995.

[18] R. Azencott, Synchronous Boltzmann machines and Gibbs fields: learning algorithms, in: F. Fogelman and J. Hérault (Eds.), Neurocomputing, NATO ASI F 68, 1990, pp. 51–64.

[19] R. Azencott, Boltzmann machines: high-order interactions and synchronous learning, in: P. Barone and A. Frigessi (Eds.), Stochastic models in image analysis, Lecture Notes in Statistics, Springer-Verlag, 1992, pp. 14–45.

[20] G.E. Hinton, T.J. Sejnowski, D.H. Hackley, Boltzmann machine: constraint satisfaction networks that learn, Carnegie-Mellon University Technical Report CNU-CS-84-119, 1984.

[21] E. Aarts, J. Korst, Simulated Annealing and Boltzmann Machines, Wiley, 1989.

[22] L. Bedini, S. Pandolfi, A. Tonazzini, Training a Boltzmann machine for edge-preserving image restoration, SPIE's 1993 International Symposium on Optical Instrumentation and Applied Sciences, Proc, Proc. Neural Stochastic Meth. Image Signal Process. II 2032 (1993) 192–202.

[23] M. Bertero, T. Poggio, V. Torre, Ill-posed problems in early vision, Proc. IEEE 76 (8) (1988) 869–889.

[24] D. Geman, G. Reynolds, Constrained restoration and the recovery of discontinuities, IEEE Trans. Patt. Anal. Mach. Intell. 14 (1992) 367–383.

[25] G. Wahba, Practical approximate Solutions to Linear Operator Equations when the Data Are Noisy, SIAM J. Numer. Anal. 14, 1977.

[26] P. Hall, D.M. Titterington, Common structure of techniques for choosing smoothing parameters in regression problems, J. R. Stat. Soc. B 49 (1987) 184–198.

[27] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–223.

[28] S. Geman, D.E. McClure, Statistical methods for tomographic image reconstruction, Proc. of the 46th session of the ISI, Bull. ISI 52 (1987) 1–18.

[29] S. Lakshmanan, H. Derin, Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing, IEEE Trans. Patt. Anal. Mach. Intell. 11 (1989) 799–813.

[30] J. Zhang, The mean field theory in EM procedures for Markov random fields, IEEE Trans. on Signal. Process. 40 (10) (1992) 2570–2583.

[31] J. Zhang, The mean field theory in EM procedures for blind Markov random field image restoration, IEEE Trans. Image Process. 2 (1) (1993) 27–40.

[32] C.J. Geyer, E.A. Thompson, Constrained Monte Carlo maximum likelihood for dependent data, J. R. Statist. Soc. ser. B 54 (1992) 657–699.

[33] D.M. Higdon, V.E. Johnson, T.G. Turkington, J.E. Bowsher, D.R. Gilland, R.J. Jaszczak, Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data, Technical Report 96–21, Institute of Statistics and Decision Sciences, Duke University, 1995.

[34] X. Escombes, R. Morris, J. Zerubia, M. Berthod, Maximum likelihood estimation of Markov random field parameters using Markov chain Monte Carlo algorithms, in: M. Pelillo and E.R Hancock (Eds.), Lecture Notes in Computer Science 1223, 1997, pp. 133–148.

[35] A. Tonazzini, L. Bedini, S. Minutoli, Joint MAP image restoration and ML parameter estimation using MRF models with explicit lines, IASTED Int. Conf. on Signal and Image Processing (SIP'97), New Orleans, 1997.

[36] H. C. Andres, B. R. Hunt, Digital Image Restoration, Prentice-Hall, Englewood Cliffs, NJ, USA, 1977.

[37] L. Bedini, I. Gerace, E. Salerno, A. Tonazzini, Models and algorithms for edge-preserving image reconstruction, in: P.W. Hawkes (Ed.), Adv. Imag. Electron Phys., vol. 97, 1996, pp. 86–189.

[38] L. Bedini, I. Gerace, A. Tonazzini, A deterministic algorithm for reconstructing images with interacting discontinuities, CVGIP: Graphic. Models Image Process. 56 (2) (1994) 109–123.

[39] J.L. Marroquin, Probabilistic solution of inverse problems, MIT-AI Technical Report 860, 1985.

[40] L. Bedini, A. Tonazzini, Image restoration preserving discontinuities: the Bayesian approach and neural networks, Image Vision Comput. 10 (1992) 108–118.

[41] J.J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Natl Acad. Sci. USA 81 (1984) 3088–3092.

[42] D. Chandler, Introduction to modern statistical mechanics, Oxford University Press, 1987.

[43] L. Bedini, I. Gerace, A. Tonazzini, A GNC algorithm for constrained image reconstruction with continuous-valued line processes, Pattern Recog. Lett. 15 (9) (1994) 907–918.