Hidden Markov Measure Field Models for Image Segmentation

Jose L. Marroquin, Edgar Arce Santana, and Salvador Botello

Abstract—Parametric image segmentation consists of finding a label field that defines a partition of an image into a set of nonoverlapping regions and the parameters of the models that describe the variation of some property within each region. A new Bayesian formulation for the solution of this problem is presented, based on the key idea of using a doubly stochastic prior model for the label field, which allows one to find exact optimal estimators for both this field and the model parameters by the minimization of a differentiable function. An efficient minimization algorithm and comparisons with existing methods on synthetic images are presented, as well as examples of realistic applications to the segmentation of Magnetic Resonance volumes and to motion segmentation.

Index Terms—Markov random fields, segmentation, motion.

1 Introduction

After the seminal work by Besag [1] and Geman and Geman [2], probabilistic methods and, in particular, Markov Random Field (MRF) models, have been used with great success for the solution of a number of important problems in image analysis; there is a vast amount of published works on the subject that include applications in image restoration, segmentation, edge-preserving filtering, reconstruction in inverse problems, etc. (see [3], [4], and references contained therein). There are two main reasons for this success: discrete MRFs provide a systematic way—firmly rooted in Bayesian estimation theory—for including prior constraints about the shape and average size of homogeneous regions in an image; since these macroscopic properties result from local interactions, a wide variety of behaviors may be obtained, simply by varying a few parameters in the definition of local potentials in the MRF model. The second reason is that, even when exact optimal estimators cannot be precisely computed, it is possible to design reasonable approximate algorithms that work well in many cases, although sometimes with high-computational costs.

A particular problem, that has been approached with this kind of model, is image segmentation: It consists of partitioning an image into a set of nonoverlapping regions $\{R_1, \ldots, R_M\}$, so that the variation of some property (such as intensity, depth, velocity, color, etc.) within each region R_k is either constant, or follows a simple model Φ_k. What makes this problem specially difficult is the fact that one has to estimate both the parameters that characterize each model Φ_k , and the corresponding regions of validity R_k at the same time. To solve it, prior MRF models have been used in conjunction with iterative procedures—in particular, the Expectation Maximization (EM) algorithm [5], which is reasonably effective, but entail a high-computational cost. The goal of this paper is to present a new class of probabilistic models that permits the characterization of the solution for complex segmentation problems in terms of the minimization of a differentiable energy function, for which efficient algorithms can be devised. We will show that these models, which are also rigorously based on Bayesian estimation theory, represent a significant improvement over classical MRF's, both in terms of the accuracy of the solutions and of computational complexity, and are also versatile and generally applicable. The plan of our presentation is the following: In Section 2, we review the classical MRF formulation of parametric segmentation problems, introduce our new model, and present efficient estimation algorithms. In Section 3, we compare experimentally the performance of the new scheme with that of classical ones, and discuss the problem of control parameter selection. In Section 4, we present two examples of applications to illustrate the versatility of our approach: to the segmentation of brain Magnetic Resonance (MR) volumes and to motion segmentation. Finally, some conclusions are drawn in Section 5.

• *The authors are with the Center for Research in Mathematics (CIMAT), Apartado Postal 402, Guanajuato, Gto. 36000, Mexico. E-mail: {jlm, roman, botello}@cimat.mx.*

Manuscript received 25 Apr. 2002; revised 23 Jan. 2003; accepted 6 May 2003.
Recommended for acceptance by M.A.T. Figueiredo, E.R. Hancock, M. Pelillo, and J. Zerubia.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118620.

2 Hidden Markov Field Models for Image Segmentation

2.1 Classical MRF Models

To describe the probabilistic models that have been used in most cases to formulate segmentation problems, we introduce the following notation: Let L represent the pixel (or voxel, in 3D problems) lattice, where images I are observed. The model assumes that there are M regions, $\{R_1, \ldots, R_M\}$, such that $L = \bigcup_{k=1}^{M} R_k; R_i \cap R_j = \emptyset, i \neq j$, so that the observation at pixel $r \in L$ is given by:

$$I(r) = \sum_{k=1}^{M} \Phi(r, \theta_k) b_k(r) + n(r), \tag{1}$$

where $n(r)$ is a white noise field with known distribution P_n (e.g., $\{n(r), r \in L\}$ are 0-mean, independent, identically

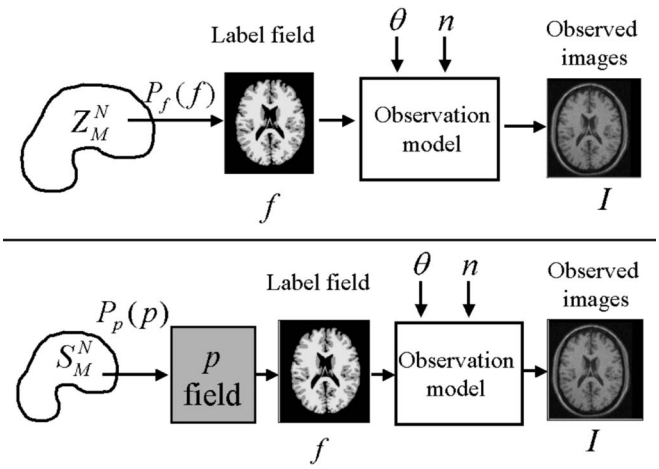


Fig. 1. Probabilistic models for image segmentation. Above: Classical MRF Model. Below: Hidden Measure Field Model (see text).

distributed Gaussian random variables with standard deviation σ), $\Phi(\cdot,\cdot)$ is a parametric model, θ_k is the parameter vector that corresponds to region R_k, and $b_k(r)$ is the corresponding indicator function: $b_k(r) = 1 \Leftrightarrow r \in R_k$; note that $b(r)$ satisfies the constraints:

$$\sum_{k=1}^{M} b_k(r) = 1,\ b_k(r) \in \{0,1\},\ \text{for all } r \in L. \tag{2}$$

Associated with b, there is a label field f, with $f(r) \in Z_M = \{1,\ldots,M\}$, that indicates to which region r belongs, i.e., $b_k(r) = \delta(f(r) - k)$, where $\delta(x)$ equals 1 if $x = 0$, and equals 0 otherwise. In this model, the field f is assumed to be a sample from a MRF, i.e., a sample from the state space Z_M^N (where N is the cardinality of L), obtained with a Gibbsian distribution:

$$P_f(f) = \frac{1}{Z_f}\exp[-\sum_C V_C(f)], \tag{3}$$

where Z_f is a normalizing constant and the sum in the exponent ranges over the cliques of a given neighborhood system on L, and $\{V_C(f)\}$ are "potential functions," each one of which depends only on the value of f at the sites that belong to the clique C (see [2], [3] for details). These potential functions, together with the neighborhood system selected, control the appearance of the sample field f and, hence, the properties of the estimated segmentation. A potential that is often used is the generalized Ising model, which considers cliques of size 2 (e.g., pairs of sites that are one unit apart), and potentials of the form:

$$\begin{aligned} V_C(f_i, f_j) &= -\beta, \text{ if } f_i = f_j \\ &= \beta, \text{ otherwise,} \end{aligned}$$

where β is a parameter that controls the granularity of the field. Since the field f is not directly observable, it is often called a hidden MRF model. This is schematically represented in the upper panel of Fig. 1.

2.1.1 Estimation Algorithms

The segmentation problem consists of finding an optimal estimator for both the field f and the parameter vector $\theta = (\theta_1,\ldots,\theta_M)$, given the observations I. To obtain it, using Bayesian estimation theory, one follows the steps [6]:

1. Find the likelihood of the observations $P(I|f,\theta)$.
2. Using the prior distribution $P_f(f)$ (and $P_\theta(\theta)$, if available), find the posterior distribution $P(f,\theta|I)$, using Bayes rule.
3. Define an appropriate cost function $C(\hat{f},\hat{\theta},f,\theta)$, that associates a cost to estimators $\hat{f},\hat{\theta}$, given that the true values are f,θ (note that the function C may be arbitrarily chosen).
4. Find the optimal estimators f^*,θ^* by minimization of:

$$Q(\hat{f},\hat{\theta}) = E[C(\hat{f},\hat{\theta},f,\theta)|I]. \tag{4}$$

We now analyze them in detail.

The likelihood of the observations is obtained from the observation model (1) and the noise distribution P_n (assumed known):

$$P(I|f,\theta) = \prod_{r\in L} v_{f(r)}(r), \tag{5}$$

where the kth component of each M-vector $v(r)$ is defined by:

$$v_k(r) = P(I(r)|f(r) = k,\theta) = P_n(I(r) - \Phi(r,\theta_k)). \tag{6}$$

For example, for Gaussian noise, we have:

$$v_k(r) = \sqrt{\frac{\gamma}{\pi}}\exp[-\gamma|I(r) - \Phi(r,\theta_k)|^2], \tag{7}$$

where γ is a parameter that depends on the noise variance.

Using (7), (3), and Bayes rule, one finds the posterior distribution as:

$$P(f,\theta|I) = \frac{1}{Z_p}\exp[-U(f,\theta)],$$

where Z_p is a normalizing constant, and

$$U(f,\theta) = -\sum_{r\in L}\log v_{f(r)}(r) + \sum_C V_C(f) - \log P_\theta(\theta), \tag{8}$$

where a noninformative (constant) prior P_θ may be used, if there are no prior constraints on θ.

The minimization of Q (4) is usually performed by two-step procedures, which may be generically called Segmentation/Model Estimation (SM) algorithms (the EM algorithm is a particular example), in which the best segmentation, given the current estimate for the model parameters θ is found in step S, and the best estimate for θ, given the current estimate for the segmentation, is found in step M. The problem with this approach is that it is not possible to find the exact optimal segmentation in the S step and, hence, one must resort to approximations; the most precise are based on stochastic, Markov Chain Monte Carlo (MCMC) algorithms [2], [7], and are computationally very expensive; fast approximations (e.g., the ICM algorithm [8]) are highly sensitive to noise. Approximations based on Mean Field theory [9], [10] are faster than MCMC, but still relatively expensive, and also sensitive to noise (see Section 3). A recent algorithm for finding a segmentation based on the Maximizer of the Posterior marginals (MPM) estimator [11], based on a Gaussian approximation for the posterior marginals, is fast and resistant to noise; however,

since the MPM estimator corresponds to a hard segmentation, the corresponding MPM-MAP procedure is very sensitive to the initial estimate for θ.

2.2 Hidden Markov Measure Field Models

The difficulties mentioned above may be solved if one uses a different probabilistic model for the generation of the label field f. Instead of the one-step procedure described in the upper panel of Fig. 1, we propose to use the two-step probabilistic model, with an additional hidden field p, which appears in the lower panel: In a first step, a Markov random vector field p is generated with distribution $P(p) = \frac{1}{K}\exp[-\sum_C W_C(p)]$, where K is a normalizing constant, C are the cliques of a given neighborhood system, and W_C are given potential functions, and where each vector $p(r)$ takes values on the M-vector simplex S_M:

$$S_M = \{u \in \Re^M : \sum_{k=1}^{M} u_k = 1,\ u_k \geq 0, k = 1, \ldots, M\}. \quad (9)$$

Hence, $p(r)$ may be interpreted as a discrete probability measure on Z_M (the label state space). In a second step, the label field f is generated in such a way that each $f(r)$ is an independent sample from the distribution $p(r)$, so that

$$P(f|p) = \prod_{r \in L} p_{f(r)}(r), \quad (10)$$

where $p_{f(r)}(r)$ is the component of vector $p(r)$ corresponding to class $f(r)$. Note that the prior for f is:

$$P_f(f) = \int_{S_M^N} P(f|p) dP_p(p)$$

which is not Gibbsian (see [12] for a proof), so that this class of models is different from the classical ones. As in the classical case, however, potential functions (for the p field in this case) may be used to enforce the appropriate prior constraints on the label field. The spatial coherence of regions $\{R_1, \ldots, R_M\}$, for instance, may be enforced by requiring that each vector $p(r)$ is similar to its spatial neighbors. A simple quadratic potential that expresses this condition is:

$$W_{rs}(p(r), p(s)) = \lambda |p(r) - p(s)|^2 = \lambda \sum_{k=1}^{M} (p_k(r) - p_k(s))^2, \quad (11)$$

where λ is a positive parameter, and $< r, s >$ are neighboring sites in L. Other potentials may be defined to enforce more complex constraints, but here we concentrate on this simple one.

The posterior distribution $P(p, \theta|I)$ is obtained from Bayes rule as:

$$P(p, \theta|I) = \frac{1}{Z} P(I|p, \theta) P_p(p) P_\theta(\theta), \quad (12)$$

where Z is a normalizing constant. The conditional distribution $P(I|p, \theta)$ is obtained as:

$$P(I|p, \theta) = \prod_{r \in L} P(I(r)|p, \theta).$$

The conditional distribution $P(I(r)|p, \theta)$ may be obtained by first computing the joint conditional distribution

$$P(I(r), f(r)|p, \theta) = P(I(r)|f(r), p, \theta) P(f(r)|p, \theta),$$

and then marginalizing over $f(r)$:

$$P(I(r)|p, \theta) = \sum_{k=1}^{M} P(I(r)|f(r) = k, p, \theta) P(f(r) = k|p, \theta).$$

Using the fact that $P(I(r)|f(r), p, \theta) = P(I(r)|f(r), \theta)$ and that $P(f(r) = k|p, \theta) = p_k(r)$, one obtains:

$$P(I(r)|p, \theta) = \sum_{k=1}^{M} v_k(r) p_k(r) = v(r) \cdot p(r), \quad (13)$$

where $v_k(r)$ is given by (7).

From (13) and (12), one finally gets:

$$P(p, \theta|I) = \frac{1}{Z} \exp[-U(p, \theta)] \quad (14)$$

with

$$U(p, \theta) = -\sum_{r \in L} \log(v(r) \cdot p(r)) + \sum_C W_C(p) - \log P_\theta(\theta), \quad (15)$$

where we consider cliques of size 2 and potentials given by (11). To obtain the optimal estimator f^* for the label field, we use the following procedure:

1. Find the MAP estimators p^*, θ^* for p, θ:

$$p^*, \theta^* = \arg \max_{p \in S_M^N, \theta} P(p, \theta|I). \quad (16)$$

2. Find f^* as the maximizer of $P(f|p = p^*, \theta = \theta^*, I)$.

The first step is equivalent to the minimization of $U(p, \theta)$, given by (15), subject to the constraints:

$$p(r) \in S_M, \text{ for all } r \in L \quad (17)$$

with S_M defined by (9), while the second step consists simply of finding the mode for each discrete measure $p^*(r)$ in a decoupled way:

$$f^*(r) = \arg \max_k p_k^*(r). \quad (18)$$

The computational burden, thus, lies on the first step; since (15) is differentiable, however, this minimization may be carried out very efficiently, as we now show.

2.3 Energy Minimization Algorithm

The minimization of (15) may be effected using any general purpose constrained optimization technique; we have found, however, that due to the simplicity of the constraints (17), and the structure provided by the Markovianity of the p field, a multiscale gradient projection Newtonian descent (GPND) [13], [14] gives best results. This method is based on the idea of moving, at each iteration, in a direction d such that $\nabla U \cdot d < 0$ (so that it is a descent direction), and that the new point lies in the feasible region. This is achieved by choosing d as the projection of the negative gradient onto the tangent subspace defined by the set of active constraints

(see [14, pp. 331-339]). The convergence may be accelerated if one considers each element $p_k(r)$ (or $\theta_j(r)$) as the position of a particle of unit mass, subject to a force equal to $-\partial U/\partial p_k(r)$ (respectively, $-\partial U/\partial \theta_j$). The equations of motion for these particles may be obtained from Newton's second law:

$$\ddot{\theta} = -\nabla_\theta U - 2\alpha\dot{\theta}$$
$$\ddot{p} = -\nabla_p U - 2\alpha\dot{p},$$

where α is the friction coefficient. The discretization of these equations gives an iterative gradient descent algorithm with inertia; to satisfy the constraints (17), each new particle position $p_k(r)$ must be projected back into S_M, to get the complete iteration as:

$$\theta^{(t+h)} = \frac{2}{\alpha h+1}\theta^{(t)} + \frac{\alpha h-1}{\alpha h+1}\theta^{(t-h)} - \frac{h^2}{\alpha h+1}\nabla_p U(p^{(t)}, \theta^{(t)})$$
$$\tilde{p} = \frac{2}{\alpha h+1}p^{(t)} + \frac{\alpha h-1}{\alpha h+1}p^{(t-h)} - \frac{h^2}{\alpha h+1}\nabla_p U(p^{(t)}, \theta^{(t)})$$
$$p^{(t+h)}(r) = \Pi_{S_M}(\tilde{p}(r)), \text{ for all } r \in L, \tag{19}$$

where the operator $\Pi_{S_M}(\tilde{p}(r))$ finds the closest point $x \in S_M$ to the vector $\tilde{p}(r) \in \Re^M$. This is done by the following algorithm (see [12] for details):

1. set $x = \tilde{p}(r)$ and $A = Z_M$;
2. while $x \notin S_M$ do:

   a. for $k = 1, \ldots M$, set

$$x_k = \tilde{p}_k(r) - \frac{\sum_{i\in A}\tilde{p}_i(r) - 1}{|A|}, \text{ if } k \in A$$
$$= 0, \text{ if } k \notin A.$$

   b. set $A = \{k : x_k \geq 0\}$;

Note that this algorithm will converge at most in M iterations.

The minimization of (15) may be further accelerated using a multiscale approach, which is the one we use in the experiments reported below (see [12] for details).

3 Experimental Performance

In this section, we use synthetic images to compare the experimental performance of the new approach presented here (labeled HMMF) with that of classical MRF models; we compare with MPM estimators, first, because they are known to perform better than MAP estimators, particularly for high noise levels [6], and, second, because they are based on the estimation of the posterior marginals, which are the basis for EM procedures, whose performance is also interesting to compare. For the classical case, we use the generalized Ising model and three methods for computing the posterior marginals for the f field: a stochastic MCMC algorithm (the Gibbs Sampler [2]); the Mean Field (MF) approximation [9], and the Gaussian approximation reported in [11] (labeled GMMF).

In the first set of experiments, the task is to perform intensity-based segmentation from noisy data when the regions corresponding to each class have known constant intensity (i.e., $\Phi(r, \theta_k) = \theta_k$, assumed known), and the purpose is to compare the robustness of each method with respect to noise. We assumed five classes, with the class distribution shown in Fig. 2a, and with $\theta_k = k$, $k = 1, \ldots, 5$. The observed images are obtained by adding white Gaussian noise with increasing variance. As a performance measure, we choose the average number of segmentation errors. The results are summarized in Fig. 2c. As one can see, for low noise levels, all methods give similar results; as the noise level increases, the performance of the MF and MCMC approximations break down (for $\sigma = 1.5$ and $\sigma = 2$, respectively), while GMMF and HMMF degrade more gracefully, with HMMF giving the best results. In all cases, the control parameters for each method were hand adjusted to get the best possible performance. We have also compared the performances of HMMF and GMMF using the same eight-class multiband segmentation problem presented in [11]; the results are consistent with those of Fig. 2 (see [12] for details).

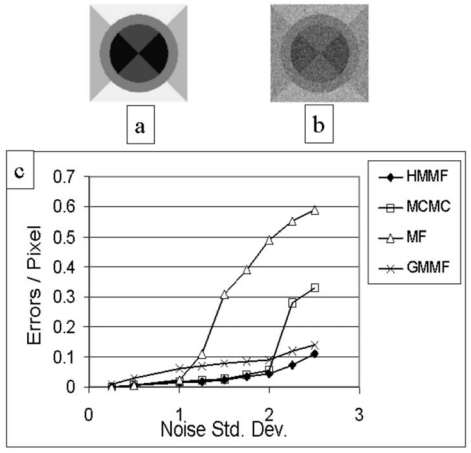


Fig. 2. For the experiments described in the text: (a) Class distribution. (b) 128×128 observed image for noise std. dev. = 15. (c) Comparative performance for four Bayesian estimators, for different noise levels.

In a second set of experiments, we test the relative robustness of the EM algorithm and our proposed procedure, with respect to initialization. To do this, we take again a synthetic, piecewise constant image with three classes, but this time we assume that the intensities $(\theta_1, \theta_2, \theta_3)$ are not known. We then generate 20 random starting points, with uniform distribution on the dynamic range of the observed image, and note whether the corresponding algorithm converged to a neighborhood (a ball of radius 0.1) of the true values of θ, in which case, the run was labeled as a "success." We tested the EM algorithm, using MF and MCMC to compute the posterior marginals and the direct HMMF method presented here. The results are summarized in Fig. 3, which also includes the corresponding average processing times. As one can see, the EM algorithm, even with an accurate approximation for the marginals (obtained with MCMC) is quite sensitive to initialization, giving a maximum success rate of 60 percent. This rate falls down to 0 for high noise levels if the MF approximation is used. HMMF, on the other hand, is much more robust (giving an 100 percent success rate in both cases) and, since it does not need to iterate, alternating between E and M steps, achieves this at a fraction of the computational time (all times refer to a PC-based workstation running at 1.8GHz).

	Success rate $\sigma = 1$	Success rate $\sigma = 1.5$	Avg. Time (sec)
EM/MCMC	60 %	60 %	800
EM/Mean Field	60 %	0 %	200
HMMF	100 %	100 %	5

Fig. 3. Top: Observed image with noise std. dev. of 1.5 (left) and sample successful (center) and unsuccessful (right) segmentations. Bottom: Success rates and average processing times for different estimation procedures for the experiment explained in the text.

HMMF has an additional advantage: If the exact number of models is not known in advance, one may initialize the procedure with a relatively large number of models, and the superfluous models will be automatically eliminated, in the sense that, if the parameter vectors for two models j, k become almost equal, the p distributions will exhibit only one dominant mode in the corresponding support region, corresponding to either one of these models.

A final word must be said about the setting of the control parameters for these methods. In all cases (i.e., for EM/MCMC, EM/MF, and HMMF), there are two control parameters: one that corresponds to the noise variance and the regularization parameter that controls the granularity of the reconstructed regions. Ideally, these parameters should be estimated—or at least fine-tuned—from the data and, in principle, some of the procedures that have been proposed to do this in the classical case [15], [9], [16] may be extended to the case of HMMF as well. The problem is that these procedures have, in general, a very high-computational cost, which we are trying to avoid in this case. The development of efficient hyperparameter estimation methods for HMMF is thus an important open problem, which we are currently investigating; however, HMMF is not too sensitive to the precise setting of these parameters: We have found, experimentally, that there is a large region around the optimal setting where the error surface (i.e., average number of segmentation errors) is very flat, and the overlap between these flat regions for different values of the SNR is also quite large (see [12] for details). This means that it is possible to callibrate the method for a particular problem class, selecting "good" values for the control parameters for a test image that belongs to the class, and use these values for the whole class of problems, getting acceptable results. This is the approach we follow for the applications described in the next section.

4 Applications

4.1 Segmentation of Brain Magnetic Resonance Images

Magnetic resonance (MR) images of the brain provide a means for imaging tissue at very high resolutions, and the assignment of each voxel to a specific class (i.e., White Matter (WM), Gray Matter (GM), or Cerebro Spinal Fluid (CSF)) is important for visualization (as in surgical planning), for solving inverse problems (e.g., in electric tomography), and for relative volume quantification, which is important for the diagnosis and prognosis of certain illnesses, etc. The main difficulties found in the automatic segmentation of brain MR volumes are due to two reasons: one is the presence of noise in the data, which cause voxel-by-voxel classification methods to produce granular or fragmented regions that violate anatomical constraints, and the other is that image intensities are, in general, non-constant for each tissue class, due to irregularities in the magnetic fields, varying magnetic properties of biological tissues, operating conditions of the MR equipment, etc. For these reasons, a precise segmentation method should include an appropriate model for spatial interactions—to control the spurious granularity due to noise —and also the simultaneous estimation of smoothly varying intensity models for each class. This makes this problem an ideal candidate for the probabilistic segmentation methods that we have described in the previous sections.

Bayesian estimation, with prior MRF models for the label field, combined with SM methods (such as EM) for the estimation of smooth intensity models, have in fact been used by a number of researchers [17], [18], [19], [20], with the problems and limitations discussed in Section 2. In most of these works, the smooth intensity models Φ are assumed to be of the form:

$$\Phi(r, \theta, k) = \mu_k \beta(r, \theta),$$

where $(\mu_1, \ldots, \mu_M)$ are the mean intensities for each tissue class and $\beta(r, \theta)$ is a multiplicative bias field that is supposed to affect all tissue classes in the same way (see also [21] for a related approach). If one wants a model that depends linearly on the parameters, however, it is necessary to perform a logarithmic transformation on the image intensities, which alters the noise distribution in a complex way, so that the Gaussian assumption is no longer valid, and also alters the image histogram, making the separation more difficult. For these reasons, and also because this simple model does not take into account the spatial variation of magnetic properties of specific tissues, we prefer to use more flexible, global spline-based models defined as linear combinations of basis functions $\{N_j, j = 1, \ldots, J\}$ each one of which has local support (see [22]):

$$\Phi(r, \theta, k) = \sum_{j=1}^{J} \theta_{kj} N_j(r) \tag{20}$$

each one of the functions $\{N_j\}$ corresponds to a quadratic tensor product B-spline basis function B^2 [23], translated to a node of a regular subgrid of the voxel lattice, which we call the *spline subgrid*: $N_j(r) = B^2((r - n_j) \cdot d)$, where n_j denotes the coordinate vector (in voxels) of the jth node of the spline subgrid and $d = (1/\Delta_x, 1/\Delta_y, 1/\Delta_z)^T$ is a scaling vector, with $\Delta_x, \Delta_y, \Delta_z$ denoting the distance between neighboring nodes on the spline subgrid for each direction. Since interslice intensity variations in MRI are usually larger than intraslice ones, we use a value of 32 voxels for Δ_x and Δ_y, and of one voxel for Δ_z.

To further control the rigidity of the models, we impose a "membrane" Gibbsian prior on θ, of the form:

$$P_\theta(\theta) = \frac{1}{Z_\theta} \exp\left[-\sum_{k=1}^{M} \sum_{<u,v>} \eta_{uv} (\theta_{ku} - \theta_{kv})^2 \right], \tag{21}$$

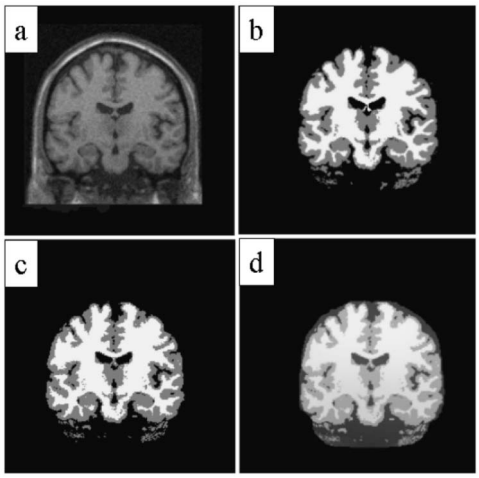


Fig. 4. (a) Sample slice of a simulated brain MR volume with 9 percent noise and 40 percent inhomogeneity. (b) Anatomical model (ground truth). (c) HMMF segmentation. (d) Reconstructured intensity $\Phi(r, \theta^*_{f_{(r)}})$.

where the second sum is taken over nearest neighbor pairs of nodes $< u, v >$ in the spline subgrid. For η_{uv}, we used 0.1 in the $x - y$ direction and 0.01 in the z direction.

To validate this application of the HMMF procedure, we use the Brainweb MRI simulator [24], which allows one to generate high quality simulated MRI volumes from known (ground truth) anatomical models, for different levels of noise and spatial inhomogeneities. Fig. 4 shows a sample slice of the simulated MRI, the anatomical model, the HMMF segmentation, and the reconstructed intensity $\Phi(r, \theta^*_{f^*(r)})$. Fig. 5 shows a comparison between HMMF results and other published results on the same data, namely, the procedure presented in [25], which uses an EM/MF approach. The performance index ξ used for the comparison is:

$$\xi_k = \frac{2V_{GPk}}{V_{Pk} + V_{Gk}},$$

where V_{GPk} denotes the total number of voxels that were correctly assigned to class k by a given procedure, V_{Pk} is the total (correct + incorrect) number of voxels assigned to class k by this procedure, and V_{Gk} denotes the total number of voxels belonging to class k in the anatomical model (ground truth). Note that ξ_k is always between 0 and 1, with 1 corresponding to a perfect segmentation. As one can see, the performance of HMMF is practically insensitive to the presence of spatial inhomogeneities, indicating they are adequately modeled. The values for the control parameters were: $\lambda = 0.01$ in the $x - y$ direction; $\lambda = 0.001$ in the z direction, and $\gamma = 1$. It is important to note that the same values were used for the complete set of experiments (i.e., for different values of the noise intensity and of the spatial inhomogeneities). The initial values for p and θ were: $p_k(r) = 1/M$, for all k, r, and θ selected in such a way that each model $\Phi(r, \theta_k)$ corresponded to a constant intensity, with these intensities corresponding to the minimum, middle, and maximum intensities of the MR volume. The brain parenchyma was segmented using the procedure described in [22] prior to the segmentation step.

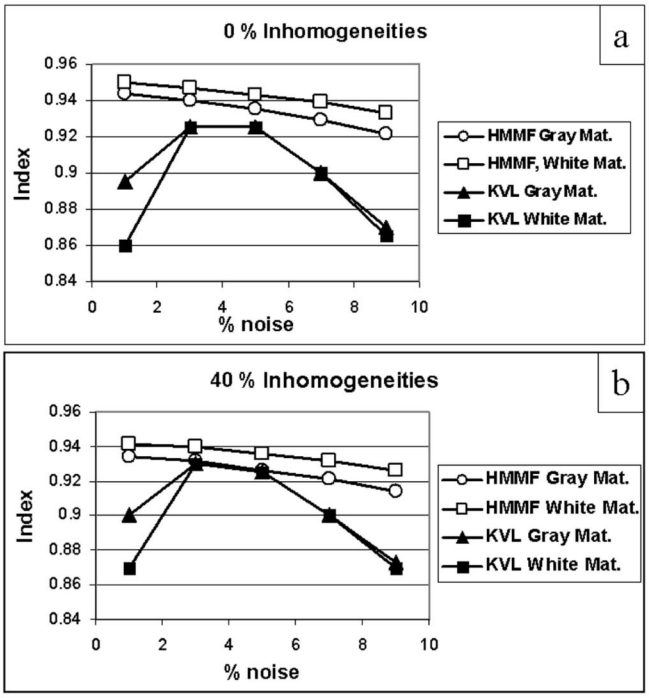


Fig. 5. Performance index for different noise levels for the HMMF segmentation and the one reported in [25] (lebeled KVL) for a simulated MR volume, for (a) 0 percent and (b) 40 percent spatial inhomogeneities.

4.2 Motion Segmentation

We now present an application example where the parameter vector θ enters into the energy function in a highly nonlinear way, and show that the HMMF method still gives very good results. This example is the segmentation of objects moving according to different velocity models from an image sequence. This is an important problem in computational vision: Useful descriptions of complex scenes are usually composed of several moving objects and simple parametric descriptions of their motions. What makes this problem difficult is that one has to find both the model parameters and the corresponding objects (i.e., the region where the model is applicable) at the same time. Although other approaches are possible (e.g., [26]), the most successful follow the Bayesian paradigm discussed in Section 2.1 [27], [28], [29], [30], [16]. In this case, the models Φ are vector-valued (since they represent velocities in 2D), and the observation model is:

$$I_1(r) = I_2(r + \Phi(r, \theta_{f(r)})) + n(r), \tag{22}$$

where I_1, I_2 represent two successive frames from the sequence, and $f(r)$ indicates the active model in pixel r, as before. Note that, since θ enters as an argument of the intensity I_2, one would need to solve a highly nonlinear optimization problem in the M step in the EM procedure, if this approach is used. To avoid this problem and lower the computational cost, often a globally smooth optic flow is precomputed and then segmented [30], in which case, the energy function in the M step becomes quadratic. However, this has a disadvantage; since a regularization method must be used to compute the optic flow, the final segmentation results are likely to lose small details and localization of the boundaries between regions. With the HMMF approach, one may work directly with the image intensities since the

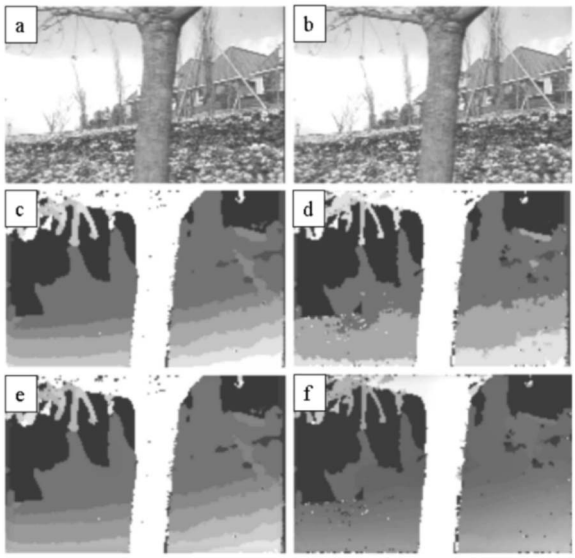


Fig. 6. (a) and (b) Two frames of a real motion sequence. (c) and (d) Piecewise constant and piecewise smooth segmentations. (e) and (f) Magnitude of the horizontal component of the velocity for the piecewise constant and piecewise smooth cases.

added nonlinearity represents only a marginal increase in the computational complexity of the procedure. For the motion models Φ, we use the spline models (20) with a Gibbsian prior ("membrane splines"), as described above, for each component of the velocity, since these models extrapolate as constants outside their support region and, hence, are less prone to produce spurious interactions with other regions and are numerically more stable than affine models. Besides, they provide a way (via the η parameter in (21)) to control the rigidity of the model and, hence, the character of the reconstructed optic flow, which may go from piecewise constant (for high values of η) to piecewise smooth (for low values).

The initial values for θ for the minimization of (15), are not very critical in the piecewise constant (pure translation) case (see Section 3). We have found that it is not necessary in this case to precompute the optic flow; we took 15×15 pixel windows, randomly placed in image I_1, and found, for each window W_k, the parameter vector θ_k as:

$$\theta_k = \arg\min_{\theta} \sum_{r \in W_k} [I_1(r) - I_2(r + \Phi(r, \theta))]^2,$$

which may be very efficiently done using the Gauss-Newton algorithm [31]. Once the minimum of (15) is found for the piecewise constant case, the piecewise smooth segmentation is initialized with constant membrane spline models which are set equal to the optimal translations. The performance of this scheme, using two frames of a real motion sequence, is illustrated in Fig. 6. For the piecewise constant case, we used eight models, which were automatically reduced to five for the piecewise smooth case. Since the vertical component of the motion was very small in this case, we show only the horizontal component of the reconstructed flow in panels Figs. 6e and 6f. The values for the control parameters were $\gamma = 0.0001$ and $\lambda = 1$ in both cases, and the distance between nodes of the membrane spline grid was 32 pixels.

Other examples of applications, e.g., to edge-preserving denoising, may be found in [12].

5 Discussion

We have presented a new energy-minimization method for image segmentation, in the case when the parameters of the models that describe the spatial variation of a given attribute within each segment are not known and when it is necessary to include prior constraints for the spatial coherence of the supports for each model. This method is rigorously based on Bayesian estimation theory, and its key idea is to introduce a hidden Markov random measure field, so that the (also hidden) label field is generated by a two-step stochastic procedure. The resulting posterior energy, given by (15), may be directly minimized with respect to p and θ, subject to the constraints $p(r) \in S_M$, instead of using costly two-step iterative procedures, such as EM, and without having to use approximations, such as MF. For the minimization of this function, any nonlinear constrained optimization method may be used. We have tried, for instance, a Quasi-Newton scheme, using a barrier function to handle the constraints [31]; the results we have gotten so far, however, are practically indistinguishable from the ones obtained with the simple gradient projection scheme reported here, and the computational cost is significantly higher.

We have presented examples that illustrate the performance of this method in a variety of situations: in intensity-based segmentation, using simple constant models, and also parametric models of high order (membrane splines), and in motion segmentation, where the model parameters enter in a highly nonlinear fashion. For simplicity, we have used Gaussian noise models in all cases, but any other model (e.g. Rayleigh noise in MRI segmentation) may be easily incorporated, simply modifying the definition of the likelihood vector (7). In all cases, one gets a consistently robust behavior, both with respect to noise, and with respect to initialization, with a reasonable computational cost. The enhanced performance of this method may be due, in part, to the nonlinear data term $-\sum_r \log(v(r) \cdot p(r))$, which is used instead of the classical $-\sum_r \log v(r)$ (which is quadratic only for Gaussian noise): This term, in combination with the quadratic regularization term $\sum_{<r,s>} |p(r) - p(s)|^2$ permits the energy function (15) to strike a good balance between two opposing tendencies. On one hand, the data term pushes each distribution $p(r)$ toward low entropy configurations, in which one component $p_k(r)$ dominates, because the minimum of $-\log(v(r) \cdot p(r))$, subject to $p(r) \in S_M$, is attained by $p_j(r) = \delta(j - k)$, $j = 1, \ldots, M$ where $k = \arg\max_j v_j(r)$. On the other hand, the regularization term acts as a diffusion and, hence, tends to produce high entropy (uniform) configurations; this balance permits the solution to evolve from an initial uniform state to a final low entropy configuration at an appropriate rate, so that the model parameters θ can escape from local minima at the beginning when the segmentation induced by p is "soft" and be optimally adjusted at the end when each $p(r)$ is sharply peaked.

As in the classical case, the potentials W_C may be adjusted to include constraints that are relevant to particular applications. We have found that the simple quadratic potentials given by (11) are sufficient to enforce general spatial coherence prior assumptions, but variations

of these potentials and the inclusion of additional terms should improve the results in specific cases.

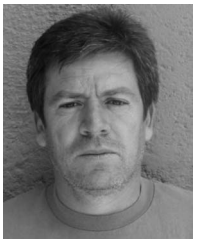
ACKNOWLEDGEMENTS

J.L. Marroquin and S. Botello were supported in part by Conacyt, Mexico, under grant 34575-A. E. Arce Santana was supported in part by a grant from UASLP, Mexico.

REFERENCES

[1] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *J. Royal Statistical Soc. B,* vol. 36, no. 2, pp. 192-236, 1974.
[2] D. Geman and S. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, 1984.
[3] S.Z. Li, *Markov Random Field Modeling in Image Analysis.* Tokyo: Springer, 2001.
[4] *Markov Random Fields: Theory and Applications,* R. Chellappa and A. Jain, eds. Academic Press, 1993.
[5] A.P. Dempster, N.M. Larid, and D.B. Bubin, "Maximum Likelihood Form Incomplete Data via EM Algorithm," *J. Royal Statistical Soc., Series B,* vol. 39, pp. 1-38, 1977.
[6] J.L. Marroquin, S. Mitter, and T. Poggio, "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *J. Am. Statistical Assoc.,* vol. 82, no. 397, pp. 76-89, 1987.
[7] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of State Calculations by Fast Computational Machine," *J. Chemical Physics,* vol. 21, pp. 1087-1092, 1953
[8] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Royal Statistical Soc. B,* vol. 48, no. 3, pp. 259-302, 1974.
[9] J. Zhang, "The Mean Field Theory in EM Procedures for Markov Random Fields," *IEEE Trans. Image Processing,* vol. 40, no. 10, pp. 2570-2583, 1992.
[10] D. Geiger and F. Girosi, "Parallel and Deterministic Algorithms from MRFs: Surface Reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 5, pp. 401-412, 1991.
[11] J. L. Marroquin, F. Velasco, M. Rivera, and M. Nakamura, "Gauss-Markov Measure Field Models for Low-Level Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 4, pp. 337-348, Apr. 2001.
[12] J.L. Marroquin, E. Arce, and S. Botello, "Hidden Markov Measure Field Models for Image Segmentation," Centro de Investigacion en Matematicas, Technical Report 1-02-05 (CC), Guanajuato, Gto. , Mexico, 2002, available at: http://www.cimat.mx/reportes/enlinea/I-02-05.pdf.
[13] J.L. Marroquin, "Deterministic Interactive Particle Models for Image Processing and Computer Graphics," *Graphical Models and Image Processing,* vol. 55, no. 5, pp. 408-417, 1993.
[14] D.G. Luenberger, *Linear and Nonlinear Programming.* Reading, Mass.: Addison Wesley, 1989.
[15] A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "Hyperparameter Estimation for Satellite Image Restoration Using a MCMC Maximum-Likelihood Method," *Pattern Recognition,* vol. 35, no. 2, pp. 341-352, Feb. 2002.
[16] N. Vasconcelos and A. Lippman, "Empirical Bayes Motion Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 217-221, Feb. 2001.
[17] J.C. Rajapakse and F. Krugge, "Segmentation of Magnetic Resonance Images with Intensity Inhomogeneities," *Image and Vision Computing,* vol. 16, pp. 165-180, 1998.
[18] K. Held, E.R. Kopsa, B.J. Krause, W.M. Wells, R. Kikinis, and H.W. Muller-Gartner, "Markov Random Field Segmentation of Brain MR Images," *IEEE Trans. Medical Imaging,* vol. 16, no. 6, pp. 878-886, 1997.
[19] W.M. Wells III, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz, "Adaptive Segmentation of MRI Data," *IEEE Trans. Medical Imaging,* vol. 15, pp. 429-442, Aug. 1996.
[20] Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Trans. Medical Imaging,* vol. 20, no. 1, pp. 45-57, Jan. 2001.
[21] M. Styner, C. Brechbuler, G. Szekely, and G. Gerig, "Parametric Estimate of Intensity Inhomogeneities Applied to MRI," *IEEE Trans. Medical Imaging,* vol. 19, no. 3, pp. 153-165, 2000.
[22] J.L. Marroquin, B.C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI," *IEEE Trans. Medical Imaging,* vol. 21, no. 8, pp. 934-945, 2002.
[23] I.J. Schoenberg, *Cardinal Spline Interpolation.* Philadelphia, P.A.: Soc. for Industrial and Applied Math., 1973.
[24] C.A. Cocosco, V. Kollokian, R.K. Kwan, and A.C. Evans, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," *NeuroImage,* vol. 5, no. 4, parts 2/4, S425, 1997.
[25] K.V. Leemput, F. Maes, D. Vandermuelen, and P. Suetens, "Automated Model-Based Tissue Classification of MR Images of the Brain," *IEEE Trans. Medical Imaging,* vol. 18, no. 10, pp. 897-908, 1999.
[26] J. Shi and J. Malik, "Motion Segmentation and Tracking Using Normalized Cuts," *Proc. Int'l Conf. Computer Vision,* 1998.
[27] J. Zhang and G.G. Hanauer, "The Application of Mean Field Theory to Image Motion Estimation," *IEEE Trans. Image Processing,* vol. 4, no. 1, pp. 19-33, Jan. 1995.
[28] F. Heitz and P. Bouthemy, "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 12, pp. 1217-1232, Dec. 1993.
[29] J. Konrad and E. Dubois, "Bayesian Estimation of Motion Vector Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 9, pp. 910-927, Sept. 1992.
[30] Y. Weiss and E.H. Adelson, "A Unified Mixture Framework for Motion Segmentation: Incorporing Spacial Coherence and Estimating the Number of Models," *Proc. IEEE Computer Vision and Pattern Recognition,* pp 321-326, 1996.
[31] J. Nocedal and S.J. Wright, *Numerical Optimization.* New York: Springer, 1999.

Jose L. Marroquin received the BS degree in chemical engineering in 1968 from the National University of Mexico and the MSc and PhD degrees in systems science in 1985 from the Massachusetts Institute of Technology. He has worked for PEMEX, the Mexican petroleum company, as project leader in geophysical data processing. Currently, he is the head of the Computer Science Department at the Center for Research in Mathematics, Guanajuato, Mexico, and is conducting research related to the computer processing visual information. Dr. Marroquin is a fellow of the National Research System of the Mexican Government and of the Mexican Academy of Science.



Edgar Arce Santana received the BSc degree in computer science engineer from the Technical Institute of San Luis Potosi, Mexico, in 1987 and the MSc degree in 2000 from the Center of Research in Mathematics (CIMAT) in Guanajuato, Mexico. He is currently pursuing his PhD degree at CIMAT. His research interests are in the areas of computer vision and digital image processing. He has worked on problems relating to parametric image segmentation, disparity estimation, and motion segmentation.

Salvador Botello received the BSc degree in civil engineering from the University of Guanajuato in 1985, the MSc degree in structure engineering from ITESM, Mexico, in 1987, and the PhD degree from the Polytechnic University of Cataluña, Barcelona, Spain, in 1993. He is currently a research-professor in the Department of Computer Science at the Center for Research in Mathematics (CIMAT) in Guanajuato, Mexico. His research interests include image processing (medical image analysis, fringe pattern analysis), computational vision (stereo vision), optimization, and finite element applications in solid mechanics.