Entropy-Controlled Quadratic Markov Measure Field Models for Efficient Image Segmentation

Mariano Rivera, Omar Ocegueda, and Jose L. Marroquin

***Abstract*—We present a new Markov random field (MRF) based model for parametric image segmentation. Instead of directly computing a label map, our method computes the probability that the observed data at each pixel is generated by a particular intensity model. Prior information about segmentation smoothness and low entropy of the probability distribution maps is codified in the form of a MRF with quadratic potentials so that the optimal estimator is obtained by solving a quadratic cost function with linear constraints. Although, for segmentation purposes, the mode of the probability distribution at each pixel is naturally used as an optimal estimator, our method permits the use of other estimators, such as the mean or the median, which may be more appropriate for certain applications. Numerical experiments and comparisons with other published schemes are performed, using both synthetic images and real data of brain MRI for which expert hand-made segmentations are available. Finally, we show that the proposed methodology may be easily extended to other problems, such as stereo disparity estimation.**

***Index Terms*—Bayesian methods, energy minimization, image segmentation, magnetic resonance image (MRI) segmentation, Markov random fields (MRFs).**

I. Introduction

IMAGE segmentation is an important task in image analysis and image editing tasks. Its importance for low-level image processing stems from several facts: boundaries between segmented regions may be highly correlated with perceptually significant edges or contours, thus, the relevance of segmentation for edge detection, edge-preserving filtering, and piecewise image reconstruction and restoration. Besides, there are many problems for which the core of the solution procedure is a segmentation algorithm; for instance: medical image analysis (including the localization of tumors and other pathologies; measurement of tissue volumes; computer-aided surgery; anatomical studies, etc.) [1], [2]; foreground extraction (image matting) [3], [4]; motion computation [5]–[9]; interactive image segmentation (trimap) [3], [10]–[13]; pattern recognition systems, etc. Therefore, any improvement to segmentation methods in their computational complexity, reduction in memory requirements, or error reduction will have an important impact in many image processing and computer vision applications.

Manuscript received November 1, 2006; revised August 26, 2007. This work was supported in part by CONACYT (Mexico) under Grants 46270 and 61367. O. Ocegueda was supported in part by a scholarship from CONACYT. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mario A. T. (G. E.) Figueiredo.

The authors are with the Department of Computer Science, Centro de Investigacion en Matematicas (CIMAT) A.C., Guanajuato, Gto. 36000 Mexico (e-mail: mrivera@cimat.mx; ocegueda@cimat.mx; jlm@cimat.mx).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2007.909384

In its general form, image segmentation may be formulated as follows: an image I may be regarded as a mapping from a pixel lattice L to a state space E. A segmentation S of I is a partition of L into a set of nonoverlapping (not necessarily connected) regions $\{R_i, i = 1, \ldots, M\}$ such that $L = \bigcup_i R_i$ and I is uniform in some sense over every region R_i. In particular, one may consider the case in which the values of the image I in each region R_i may be represented as the sum of some fixed deterministic functions plus noise. Considering these functions as parametric models whose parameters are constant for each region, one obtains

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

where x denotes a pixel in L; $\Phi(x;\theta_k)$ is a function (parametric model) $\Phi : L \mapsto E$ that depends on the parameter vector θ_k; $\{n(x) : x \in L\}$ is a set of independent random variables and b_k is the indicator function of region R_k, i.e., $b_k(x) = 1$ iff $x \in R_k$ and

$$\sum_{k=1}^{M} b_k(x) = 1 \tag{2}$$

for all $x \in L$. In what follows, $b(x)$ will denote the vector $(b_1(x), \ldots, b_M(x))$ and b will denote the set $\{b(x) : x \in L\}$.

The appearance of the image I will depend on the nature of the functions Φ; for example, if each function $\Phi(x;\theta_k)$ is smooth, I will be a piecewise smooth image, with discontinuities located at the boundaries between adjacent regions R_i, R_j. A particular instance of this case is related to the segmentation of brain magnetic resonance images (MRI) in terms of tissue type; in this case, $\Phi(x;\theta_k)$ represents the intensity associated with tissue type k and R_k represents the portion of the image classified as tissue k (see Section III-2). Other important problems that may also be formulated in these terms include: edge-preserving smoothing [14]–[17], color-based segmentation [18], [19], stereo disparity estimation [20]–[23], motion-based segmentation [24], etc.

This problem has been approached from different perspectives: if the parameters $\{\theta_k\}$ are known, the segmentation problem has been solved using, for instance: the k-means algorithm and its variants [25], [26], region merging [19], region-growing [17], [27] and active contour [28] approaches, eigendecomposition-based partitions [29]–[31], variational methods [5], [32], and probabilistic (Bayesian) formulations [14], [33], [34]. Of these, the Bayesian formulations are one of the most powerful and general, since they allow for the inclusion of spatial coherence constraints that regularize the solution [via Markov random field (MRF) models] and makes it more robust with respect to noise, although they may be

computationally expensive. In the general case, when the parameters $\{\theta_k\}$ are not known, some of these methods may be extended using two-step algorithms, in which one is given an initial estimate $\{\theta_k^{(0)}\}$, and then one iterates the following steps:

1) estimate S given $\bar{\theta}$;
2) estimate $\bar{\theta}$ given S;

where $\bar{\theta}$ denotes the collection of all parameter vectors θ_k, until convergence [17], [35]–[38].

One particularly important class of these two-step methods is derived from the expectation–maximization (EM) algorithm, which was originally proposed for computing maximum likelihood estimators from incomplete data [35]. This approach has also been used for computing estimators with respect to posterior distributions for segmentation tasks (classification) in image analysis and computer vision problems where the class model parameters are unknown [14], [33]. However, prior probability distributions based on MRF models introduce high correlation among the variables (labels in the segmentation task) that increments the computational complexity of the EM algorithm [1], [2], [39]–[42]. For this reason, instead of Monte-Carlo Markov chain methods, approximations such as mean field theory [43] or Gauss–Markov measure fields [44] have been used for computing the marginal probabilities in the expectation (E) step with relative success.

The problem with these approaches is their high computational complexity, and their high sensitivity with respect to noise and to the choice of $\{\theta_k^{(0)}\}$. The reason for this is that these two-step approaches can be guaranteed to converge only to a local maximum of the posterior distribution [35]; since, in most cases, this distribution has multiple maxima, if one starts the iterations from a "bad" point $\{\theta_k^{(0)}\}$, the local maximum to which the method converges may not be the global maximum, i.e., one may get suboptimal solutions [16].

One may get more robust methods if one formulates the problem in such a way that S and $\bar{\theta}$ are progressively refined at the same time, e.g., by the iterative minimization of a differentiable function that depends on $\bar{\theta}$ and on the probability of assigning models to each pixel (i.e., on a "soft" version of the segmentation S). These direct methods, such as the one in [16], do exhibit a better performance than that of two-step approaches; their computational complexity, however, is still relatively high, since the solution involves the minimization of a highly nonlinear function, and the hyperparameters of the corresponding algorithms are in general not easy to tune.

The purpose of this work is to present a direct method, which is rigorously based on a Bayesian framework, which is computationally efficient—since the solution is found by minimization of a quadratic function with linear constraints—and whose hyperparameters are easy to tune. The derivation of the method and its implementation are presented in Section II; in Section III, the experimental performance of this approach is compared with other published schemes, using both synthetic images and real MRI volumes where expert segmentations are available. Finally, in Section IV, some properties and extensions of the method are discussed and some conclusions are drawn.

II. Theoretical Derivation and Implementation

This section presents the mathematical derivation of the proposed probabilistic segmentation method. The method is initially presented by assuming that the noise, n in (1), has a Gaussian distribution and the intensity models' parameters, $\bar{\theta}$, are given; then, the generalization to other noise distributions and the problem of parameter estimation is presented. Throughout this section we will use the following notation: $r(x)$ will denote the region to which pixel x belongs, i.e., $r(x) = k$ iff $x \in R_k$; $\delta(\cdot)$ will denote the Kronecker delta function; $\bar{\theta}$ will denote the collection of all the parameter vectors $\{\theta_k, k = 1, \ldots, M\}$ that correspond to each parametric model $\Phi(x; \theta_k)$; N_x will denote the first order neighborhood of pixel x, i.e., $N(x) = \{y \in L : \|y - x\| = 1\}$ and $\sharp N_x$ will denote its cardinality.

A. Entropy-Controlled Quadratic Markov Measure Field Model

Assuming at first that the parameters $\{\theta_k\}$ are known, the goal is to estimate the field b defined in (1). To this end, one may use a Bayesian approach and model b as an MRF, so that prior constraints may be introduced as a prior Gibbs distribution

$$P_b(b) = \frac{1}{Z_b} \exp[-U_b(b)] \tag{3}$$

where Z_b is a constant and U_b is an "energy" function, which we would like to be differentiable with respect to b. To achieve this, one may relax the constraint (2), and replace it by

$$b_k(x) \geq 0, \quad k = 1, \ldots, M, x \in L \tag{4}$$

$$\sum_{k=1}^{M} b_k(x) = 1, x \in L \tag{5}$$

$$b_k(x) \approx \delta(k - r(x)), k = 1, \ldots, M, x \in L. \tag{6}$$

Also, one would like b to be spatially smooth almost everywhere, to control the granularity of the regions, i.e., to avoid the proliferation of very small regions assigned to a particular model, which may occur due to noise in the data. All these constraints may be expressed by the prior energy U_b in (3)

$$U_b(b) = \sum_{\langle x,y \rangle} \beta_{xy} \|b(x) - b(y)\|^2 - \mu \sum_{x \in L} \sum_{k=1}^{M} b_k^2(x). \tag{7}$$

The first term embodies the smoothness constraint: the sum is taken over all nearest-neighbor pairs of sites $\langle x, y \rangle$, and the interaction field β controls the granularity of the resulting regions; in what follows we assume for simplicity a constant interaction $\beta_{xy} = \lambda$, where λ is a positive hyperparameter (see, however, Section IV). The second term, together with the constraints (4) and (5), enforces constraint (6), as explained in the following; one may say that b is a Markov random measure field (MRMF)—since constraints (4) and (5) imply that each $b(x)$ is a discrete probability measure, where $b_k(x)$ represents the probability that $r(x) = k$, i.e., we are assuming that given b, each label $r(x)$ is obtained as an independent sample from $b(x)$, so that $P(r(x) = k \,|\, b) = b_k(x)$. The entropy of each one of these discrete distributions is equal to minus the expected value of $\ln b_k(x)$, taken with respect to the same distribution, i.e., $-\sum_k b_k(x) \ln b_k(x)$. To obtain a quadratic form, one may take instead the expected value of $b_k(x)$ itself, to get $-\sum_k b_k^2(x)$. This quantity (plus an additive constant) is known in the machine learning community as the Gini index [45], and it is known to closely resemble the behavior of the entropy [46]: it is minimal when the distribution becomes a delta function, and it is maximal for a uniform distribution. Thus, for positive values of

μ, (7) constrains each $b(x)$ to have low entropy, i.e., to approach a delta function.

If one assumes that the random variables $\{n(x), x \in L\}$ in (1) are independent, zero-mean Gaussian random variables with variance σ^2, the likelihood of the observed image I is obtained from (1) as

$$P(I \,|\, b, \bar{\theta}) = \frac{1}{Z} \exp\left[-\frac{1}{2\sigma^2} \sum_{x \in L} \left(I(x) - \sum_{k=1}^{M} b_k(x) \Phi(x; \theta_k) \right)^2 \right] \quad (8)$$

where Z is a constant. Since the prior on b, whose Gibbsian energy is given by (7) depends only on the spatial granularity of the segmentation and not on the model parameters, we have that $P(b \,|\, \bar{\theta}) = P_b(b)$, with P_b given by (3). Also, noting that $P(I \,|\, \bar{\theta})$ is just a normalizing constant—since the observations are given—one gets from (7) and (8), using Bayes rule, the posterior distribution

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

where Z is a normalizing constant and

$$U(b, \bar{\theta}) = \frac{1}{2\sigma^2} \sum_{x \in L} \left[\left(I(x) - \sum_{k=1}^{M} b_k(x) \Phi(x; \theta_k) \right)^2 - \mu \sum_{k=1}^{M} b_k^2(x) \right] + \lambda \sum_{\langle x,y \rangle} ||b(x) - b(y)||^2. \quad (9)$$

Note that, because of (5)

$$I(x) - \sum_{k=1}^{M} b_k(x) \Phi(x; \theta_k) = \sum_{k=1}^{M} b_k(x) (I(x) - \Phi(x; \theta_k)).$$

In the low-entropy limit, for each x, only one of the $b_k(x)$ becomes almost equal to one, and all the others become almost equal to zero, so that (6) holds and

$$\left[\sum_{k=1}^{M} b_k(x) (I(x) - \Phi(x; \theta_k)) \right]^2 \approx \sum_{k=1}^{M} b_k^2(x) (I(x) - \Phi(x; \theta_k))^2 \quad (10)$$

so that one may write the posterior energy in the simple form

$$U(b, \theta) \approx \sum_{x \in L} \sum_{k=1}^{M} b_k^2(x) \times \left[\frac{1}{2\sigma^2} (I(x) - \Phi(x; \theta_k))^2 - \mu \right] + \lambda \sum_{\langle x,y \rangle} ||b(x) - b(y)||^2. \quad (11)$$

The maximum *a posteriori* (MAP) estimator for b may, therefore, be found by minimization of the quadratic form (11) subject to the constraints (4) and (5). This is the basis for the main contribution of this work.

It is possible to remove the Gaussian assumption for n and obtain a more general expression. Let

$$v_k(x) = P(I(x) \,|\, r(x) = k, \theta_k). \quad (12)$$

Note that in the model we are using, given that $r(x) = k, I(x)$ is obtained by $I(x) = \Phi(x; \theta_k) + n(x)$, so that $I(x)$ is not affected by the probability with which $r(x) = k$ was selected, namely, $b_k(x)$, and also, if $r(x) = k, I(x)$ is independent of θ_j, for $j \neq k$. Therefore, one may write

$$P(I(x) \,|\, r(x) = k, \theta_k) = P(I(x) \,|\, r(x) = k, b, \bar{\theta}).$$

Similarly, because the probability with which $r(x) = k$ is selected does not depend on $\bar{\theta}$, one has that $P(r(x) = k \,|\, b, \bar{\theta}) = P(r(x) = k \,|\, b) = b_k(x)$. We may, therefore, write

$$\begin{aligned} P(I(x) \,|\, b, \bar{\theta}) &= \sum_{k=1}^{M} P(I(x) \,|\, r(x) = k, b, \bar{\theta}) P(r(x) = k \,|\, b, \bar{\theta}) \\ &= v(x) \cdot b(x) = \sum_{k=1}^{M} v_k(x) b_k(x). \end{aligned} \quad (13)$$

Using Bayes rule as before, one obtains the posterior distribution $P(b \,|\, I, \bar{\theta}) = 1/Z_b \exp[-U(b, \bar{\theta})]$, where the energy $U(b, \bar{\theta})$ is

$$U(b, \bar{\theta}) = -\sum_{x \in L} \left[\log(v(x) \cdot b(x)) - \mu \sum_{k=1}^{M} b_k^2(x) \right] + \lambda \sum_{\langle x,y \rangle} ||b(x) - b(y)||^2$$

but in the low-entropy limit, since only one of the $b_k(x)$ becomes almost equal to one, and the rest become almost equal to zero, one may write

$$\log(v(x, \bar{\theta}) \cdot b(x)) \approx \sum_{k=1}^{M} b_k^2(x) \log v_k(x, \theta_k)$$

so that one finally gets the general expression

$$U(b, \bar{\theta}) \approx \sum_{x \in L} \sum_{k=1}^{M} b_k^2(x) [-\log v_k(x) - \mu] + \lambda \sum_{\langle x,y \rangle} ||b(x) - b(y)||^2 \quad (14)$$

which, in the Gaussian case, noting that the normalizing constant $\log(\sqrt{2\pi})\sigma$, may be absorbed in μ, reduces to (11). Note that even in the general case, (14) is still quadratic in b. For this reason, we call models (11) and (14) entropy-controlled quadratic markov measure field (ECQMMF) models.

The equality constraints (5) may be easily handled using Lagrange multipliers: the Lagrangian associated with the objective function U is

$$\mathcal{L}(b,\bar{\theta},\gamma) = U(b,\bar{\theta}) - 2\sum_{x\in L}\gamma_x\left[\sum_{k=1}^{M} b_k(x) - 1\right] \tag{15}$$

where $\{\gamma_x\}$ are the Lagrange multipliers. Setting the gradient of (15) with respect to b and γ equal to zero, solving for $b_k(x)$ and substituting in the equality constraint (5) to eliminate the γ multipliers, one obtains the update equation

$$b_k(x) = \frac{\alpha_k(x)}{\beta_k(x)} + \frac{1 - \sum_{i=1}^{M}\frac{\alpha_i(x)}{\beta_i(x)}}{\sum_{i=1}^{M}\frac{\beta_k(x)}{\beta_i(x)}} \tag{16}$$

where

$$\alpha_k(x) = \lambda \sum_{y\in N_x} b_k(y) \tag{17}$$

and

$$\beta_k(x) = -\log v_k(x) - \mu + \lambda \sharp N_x. \tag{18}$$

Note that $b(x)$ computed with (16) does not necessarily satisfy the non-negativity constraint (4). If this constraint is violated, it is necessary to project $b(x)$ back to the feasible region, which is the M-dimensional simplex defined by the constraints (4) and (5). A simple way to perform this projection is to make the negative $b_k(x)$ equal to zero (so that (4) is satisfied) and renormalize the vector $b(x)$ so that (5) is satisfied. Note that other projection schemes are possible, e.g., finding the point in the simplex that is closest to $b(x)$, but we have found that the proposed projection is faster and works properly. In order to guarantee the convergence of this method, one should check if the energy (12) does not increase at the updated $b(x)$; if it increases, it is necessary to find a linear combination of the old and updated $b(x)$ where the constraints are satisfied and the energy does not increase, and replace the updated $b(x)$ with this point. In our experiments, however, we have found that this additional checking is unnecessary, since the method always converges with the simple projection and re-normalization procedure that enforces the constraints.

B. Model Parameter Estimation

If the parameters $\bar{\theta}$ are not known, one may apply Bayes rule again to obtain $P(b,\bar{\theta}\,|\,I)$. With a uniform prior on $\bar{\theta}$, the posterior energy is still given by (11) and (14), and the optimal estimators for b and $\bar{\theta}$ may be obtained by minimization of either one of these functions with respect to both groups of variables, subject to the constraints (4) and (5). In order to get estimators that are robust with respect to the initial values, it is desirable that in this minimization the values of both b and $\bar{\theta}$ are incrementally updated; one usually starts with a coarse estimate for $\bar{\theta}$ and $b_k(x) = 1/M$ for all k and x, i.e., a completely "fuzzy" segmentation that sharpens gradually, as the estimates for $\bar{\theta}$ become more reliable. In contrast, a typical two-step procedure (e.g., [17] based on the $\alpha-\beta$ swap graph-cut algorithm [7]) may produce sharp segmentations at the beginning of the procedure which are based on wrong estimates for $\bar{\theta}$, which may send the algorithm off-track to a local minimum from which it cannot recover.

In the Gaussian case, and assuming that the functions $\Phi(x;\theta_k)$ are linear in θ_k, i.e., are of the form

$$\Phi(x;\theta_k) = \sum_{j=1}^{m}\theta_{kj}B_j(x) \tag{19}$$

where $\{B_j, j = 1,\ldots,m\}$ are some basis functions (which, in general, may be nonlinear in x), that are chosen depending on the particular application, this minimization problem is particularly simple, since (11) is quadratic in b given $\bar{\theta}$ and also quadratic in $\bar{\theta}$ given b. In this case, the incremental updates of b and $\bar{\theta}$ may be effected very efficiently by a generalized Gauss–Seidel scheme: b is updated using (16) as before. The $\bar{\theta}$ update is effected by solving the linear system that results from setting the gradient of (11) with respect to $\bar{\theta}$ equal to zero. This system is

$$\sum_{i=1}^{m}\theta_{ki}\sum_{x\in L} b_k^2(x)B_i(x)B_j(x) = \sum_{x\in L} b_k^2(x)B_j(x)I(x) \tag{20}$$

for $k = 1,\ldots,M$ and $j = 1,\ldots,m$.

In the special case where $n = 1$ and $B_1(x) = 1$ (i.e., the case of constant models), the $\bar{\theta}$ update is simply

$$\theta_k = \frac{\sum_{x\in L} b_k^2(x)I(x)}{\sum_{x\in L} b_k^2(x)} \tag{21}$$

for $k = 1,\ldots,M$ (note that in this case θ_k is a scalar). In the general case, however, one cannot give an explicit formula and the $Mn \times Mn$ system (20) must be solved numerically.

This scheme for the minimization of (11) bears some resemblance with a generalized EM procedure in which the M step is only partially implemented [36], improving the likelihood, but not necessarily maximizing it. The b update, however, is different from the one that would be obtained from a classical EM formulation; moreover, if one uses a different optimization scheme this resemblance will be lost.

C. Hyperparameters Selection

An important issue for the successful application of this method is the determination of appropriate values for the hyperparameters of the system. In the case of (11), these hyperparameters are: the noise variance σ, the entropy control parameter μ, and the regularization parameter λ. The noise variance σ can usually be obtained from the data, for instance, as the mode of the empirical distribution of the local variance, which may be estimated using sliding windows (e.g., of 3×3 pixels). For the determination of μ, one notes that the numerical stability of the iterative procedure described above can only be guaranteed if the quadratic function U is positive definite. A sufficient condition for this to happen is that the coefficients of $b_k^2(x)$ in the first term of (11) are non-negative for all k and x, which happens if

$$\mu \le \min_{k,x}\frac{1}{2\sigma^2}(I(x) - \Phi(x;\theta_k))^2.$$

We have found, however, that in practice this condition may be relaxed, and it is enough if it holds on average, i.e., if

$$\mu \leq \langle \min_k \frac{1}{2\sigma^2}(I(x) - \Phi(x;\theta_k))^2 \rangle_x \tag{22}$$

where $\langle \cdot \rangle_x$ denotes average over all x. Since under the Gaussian noise assumption

$$\langle \min_k (I(x) - \Phi(x;\theta_k))^2 \rangle_x \approx E\left[\left(I(x) - \Phi(x;\theta_{r(x)})\right)^2\right] = \sigma^2 \tag{23}$$

the constraint (22) is simply $\mu \leq 0.5$. Since one wants μ to be as large as possible in order to have an adequate entropy control, an appropriate choice is $\mu = 0.5$.

Regarding parameter λ, we have found that a value of $\lambda = 4$ gives very good results, regardless of the noise level. Therefore, for all the experiments reported in the next section, we use: σ estimated from the data; $\mu = 0.5$ and $\lambda = 4$.

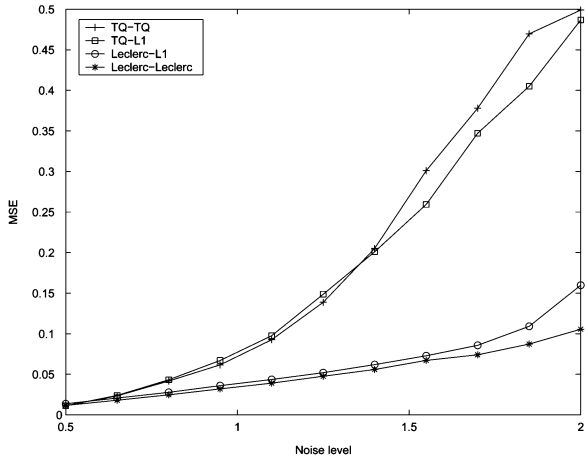

Fig. 1. MSE as a function of the noise level. First acronym indicates the potential used in the data term. Second acronym indicates the potential used in the regularization term.

III. Experimental Results

In the first set of experiments, we compare the performance of the ECQMMF method with three of the most competitive published schemes: a two-step algorithm based on a multiway ($\alpha - \beta$ swap) graph-cut method [7], [17], the MPM-MAP algorithm of [37], and the hidden Markov measure field (HMMF) direct method of [16].

In the first case, following [17], the hard segmentation was computed by alternating minimizations with respect to the label map, b, with the $\alpha - \beta$ swap algorithm [7] and with respect to the parameters, $\overline{\theta}$, by performing a half-quadratic minimization [47]–[50]. In our notation, that corresponds to the minimization of the cost function

$$\begin{aligned} U_h(b,\theta) &= \sum_x \sum_k \rho_1\left(I(x) - \theta_k^T b(x), \gamma_1\right) \\ &\quad + \lambda \sum_{\langle x,y \rangle} \rho_2\left(\theta_k^T b(x) - \theta_k^T b(y), \gamma_2\right) \end{aligned} \tag{24}$$

where $\rho_i(\cdot)$ (with $i = 1, 2$) are robust potentials, such as [7], [47], and [48]: Ising: $\rho(z) = 1 - \delta(z)$; L1: $\rho(z) = |z|$; truncated quadratic (TQ): $\rho(z) = \min\{z^2, \gamma\}$; or Leclerc: $\rho(z) = 1 - \exp(-\gamma z^2)$, where γ is a parameter. In Fig. 1, we show the computed MSE for different combinations ρ_1, ρ_2. According to our experiments, using Leclerc's potential for both ρ_1 and ρ_2 is better for nonparametric graph cut-based segmentation than using any of the potentials proposed in [7].

The synthetic test images were generated with the model

$$I(x) = \sum_{k=1}^{5} b_k(x)\theta_k + n(x) \tag{25}$$

i.e., with piecewise constant intensity models $\Phi(x;\theta_k) = \theta_k$ (a scalar) and with $n(x) \sim N(0, \sigma^2)$. The task is to estimate both the segmentation (i.e., the field b) and the parameters $\{\theta_k, k = 1, \ldots, 5\}$. The actual parameter values (assumed unknown) were $\{\theta_k = k, k = 1, \ldots, 5\}$, and the actual images appear in Fig. 2. We tested the performance of the four methods for noise levels $\sigma = 0.5 + 0.15j, j = 0, \ldots, 15$. In all cases, the initial estimate for the models was obtained by dividing the dynamic range of the observed image into four equal intervals and setting $\{\theta_k\}$ equal to the extremes and dividing points. The hyperparameters for the MPM-MAP, HMMF, and GC methods were determined by sampling the parameter space near the parameter values recommended in the corresponding publications; a total of 200 sets of values for each noise level were explored in each case, and the best set [in terms of mean squared error (MSE)] was selected for each noise level; for the ECQMMF method the same value for the hyperparameters, with σ estimated from the data as explained above, was used in all cases. Examples of the results of typical runs are shown in Fig. 2 in [51]. Average performances, over ten realizations for each noise level, are shown in Fig. 2. As one can see, the performance of direct methods (HMMF and ECQMMF) is better than that of two-step approaches based on "hard" segmentations, such as MPM-MAP and GC. This confirms the results in [16]. The performance of ECQMMF, however, is better than that of HMMF. The computational load is also significantly smaller, as one can see in Fig. 2.

The second set of experiments is meant to illustrate the performance of the ECQMMF method in a practical, more complex application, namely, the segmentation of brain MRI in terms of tissue type [white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF)]. This is an important problem in biomedicine, which has received a lot of attention [37], [52]–[55]. What makes it difficult is that the intensity of each tissue type is, in general, not constant across the volume, due to irregularities in the magnetic fields, spatially varying magnetic properties of the biological tissues, etc. The methods used to perform this segmentation involve, in general, several steps, computational processes and model specifications that influence the results [56]–[59]; some of these factors are as follows.

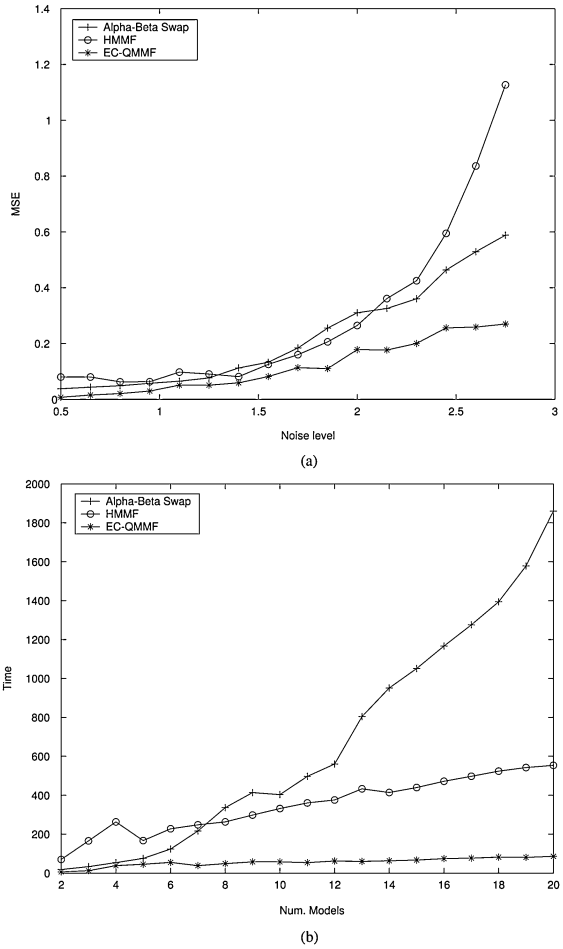

Fig. 2. EC-QMMF performance comparison. (a) Average mean squared error (MSE) for ten independent realizations of the noise fields for the images of Fig. 2, as a function of the noise level. (b) Average execution time (in hundredths of seconds) over ten images distorted with the first noise level ($\sigma = 0.5$), as a function of the number of models. All methods were run until convergence.

1) The use of spatially varying intensity models for each tissue type: in many cases a single smooth multiplicative bias field is assumed to affect equally the mean intensity of all tissues [1], [60]; other choices are finite element membrane models [37], spline functions [16], etc.
2) The use of spatially varying prior probabilities for each tissue type, taken from statistical studies [37], [61]. This requires the registration of the specimen to be segmented with a reference volume (anatomical atlas), and the registration method used clearly affects the result.
3) Preprocessing of the data: sometimes edge-preserving filtering (e.g., anisotropic diffusion [62]) is used to eliminate noise prior to the segmentation; also, some form of intensity normalization is often applied.
4) The use of separate intensity models for each slice vs the use of coupled 3-D models.
5) The use of a 2-D segmentation algorithm for each slice vs a 3-D segmentation scheme.

Since we are interested in the evaluation of precisely the image segmentation algorithm, we make the simplest choices for the other factors; specifically, we use the 2-D ECQMMF segmentation scheme described above and assume no spatially varying prior probabilities and no preprocessing. To model the spatially varying intensity of each tissue on each slice, we use a 2-D spline-based model with a Gibbsian prior on the corresponding coefficients. In particular, the intensity of tissue type k is modeled by (19), where the basis functions $B_j(x)$ are tensor product quadratic splines

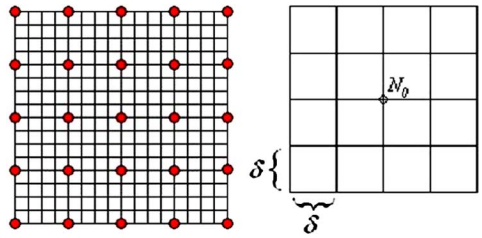

Fig. 3. Control points in the tensor product B-splines model.

$$B_j(x) = T\left(\frac{x_1 - x_{j1}}{\delta}\right) T\left(\frac{x_2 - x_{j2}}{\delta}\right)$$

$$T(x) = \begin{cases} \frac{(1.5-2x^2)}{2}, & |x| \in [0, 0.5) \\ \frac{(x^2-3|x|+2.25)}{2}, & |x| \in [0.5, 1.5) \\ 0, & |x| \geq 1.5 \end{cases}$$

where $x = (x_1, x_2)$ are the pixel coordinates and $\{x_j = (x_{j1}, x_{j2}), j = 1, \ldots, J\}$ are the coordinates of the nodes of a coarse grid, where δ is the node spacing (see Fig. 4). The coefficients $\{\theta_{kj}\}$ are assumed to have a prior Gauss–Markov distribution

$$P_\theta(\bar{\theta}) = \frac{1}{Z} \exp\left[-\eta \sum_{\langle i,j \rangle} \sum_{k} (\theta_{ki} - \theta_{kj})^2\right]$$

where Z is a normalizing constant and the sum is taken over nearest-neighbor pairs of nodes $\langle i, j \rangle$ of the coarse grid. This model produces smoothly varying intensities, with the degree of smoothness controlled by the parameter η.

Assuming additive, Gaussian observation noise, one gets the following ECQMMF posterior energy for the segmentation problem:

$$U(b, \bar{\theta}) = \sum_{x \in L} \sum_{k=1}^{M} b_k^2(x) \times \left[\frac{1}{2\sigma^2}(I(x) - \Phi(x; \theta_k))^2 - \mu\right] + \lambda \sum_{\langle x,y \rangle} \|b(x) - b(y)\|^2 + \eta \sum_{\langle i,j \rangle} \|\theta_i - \theta_j\|^2. \quad (26)$$

The performance of this simplified segmenter is evaluated using a set of 20 real brain MRI volumes which can be obtained from [63], and for which expert hand made segmentations are available. In this data set, each voxel corresponds to an actual volume of $1 \times 1 \times 3$ mm, where the higher resolution corresponds to coronal slices. Due to the low resolution (3 mm) in the other dimension, the partial volume effect is significant; in particular, when considering a coronal slice, there is a substantial number of voxels whose intensity corresponds to a mixture

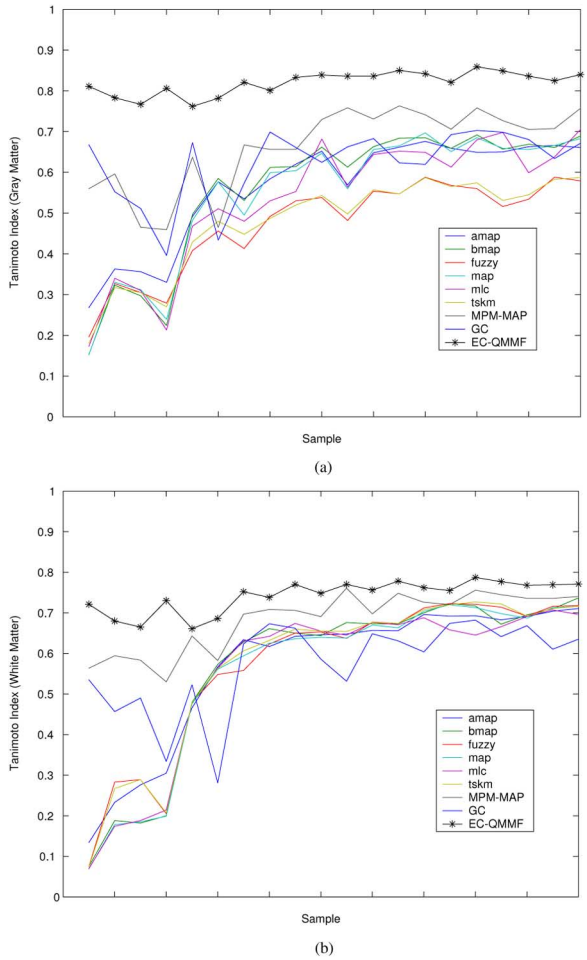

Fig. 4. Tanimoto coefficient for the segmentation results on the 20 $1 \times 1 \times 3$-mm-thick MRI volumes data set of [63]. Sample volumes in the same order as in Table II. (a) White matter. (b) Gray Matter (see text).

TABLE I
TANIMOTO COEFFICIENTS FOR DIFFERENT METHODS ON THE $1 \times 1 \times 3$-MM RESOLUTION DATA SET OF [63]

	GM		WM	
	Average	Std. dev.	Average Average	Std dev.
Adaptive MAP	0.564	0.130	0.567	0.180
Biased MAP	0.558	0.169	0.562	0.214
Fuzzy c-means	0.473	0.115	0.567	0.196
MAP	0.55	0.164	0.554	0.213
Max. Lik.	0.535	0.160	0.551	0.207
TSKM[1]	0.477	0.118	0.571	0.198
MPM-MAP	0.662	0.102	0.683	0.072
Graph-cut	0.623	0.087	0.574	0.112
Manual[2]	0.876	-	0.832	-
EC-QMMF	**0.819**	**0.028**	**0.742**	**0.039**

[1] Tree Structured K-Means. [2] Average of 2 experts.

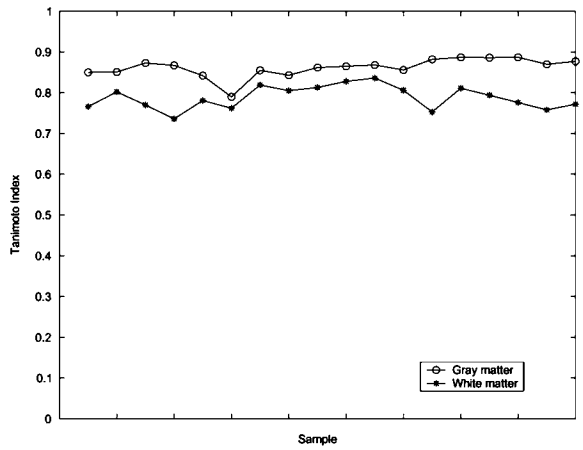

Fig. 5. Tanimoto coefficient for the segmentation results on the 1.5 mm thick in the data set of [63]. Sample volumes in the same order as in Table II (see text).

of GM and CSF; although there are several strategies for dealing with this partial volume effect [54], [64], in this case, we simply consider four tissue classes: CSF (0), CSF+GM (1), GM (2), and WM(3). Observing the expert segmentations, one can see that there are only three classes, since classes (1) and (2) are assigned to GM. Therefore, we performed the same grouping after the four-class segmentation.

The results are summarized in Fig. 4 and Table I. To evaluate the performance of our method with higher resolution and better quality data, we also tested it with 18 additional volumes with $1 \times 1 \times 1.5$ mm resolution, which are also available in [63]; the results, presented in Figs. 5 and 7 and in the second column of Table II, indicate a good, consistent performance in this case as well. In all cases, we used the same parameter values as before: $\lambda = 4$, $\mu = 0.5$ and σ estimated from each slice as explained above. For the parameters related to the spline model we used $\eta = 5$ and $\delta = 256/31$ (32×32 control points). The comparison with the expert (ground truth) segmentations is done using the *Tanimoto coefficient* (Jaccard similarity measure), which is defined as

$$\mathrm{TC}(k) = \frac{V_{p \cap g}(k)}{V_{p \cup g}(k)}$$

where $V_{p \cap g}(k)$ denotes the number of voxels classified as class k by both the proposed method and the expert (taken in this case as ground truth) and $V_{p \cup g}(k)$ denotes the number of voxels classified as class k by either the proposed method or the expert. In all cases, the brain parenchyma was presegmented by hand.

A meaningful comparison with other published methods for this task is not easy to do, since, as we pointed out above, published approaches include different processes and model specifications, besides the image segmentation algorithm itself. As a reference, however, we also include in Fig. 4 and Table I the performances (in the 3-mm slices data set) of other published schemes taken from [63] and also with the method of [37] (labeled MPM-MAP). In Fig. 6, we show as an example, the segmentation of a "difficult" volume (i.e., [63, vol. ibsr2-4]), where most published methods show poor accuracy, mostly due to the high spatial inhomogeneities that are present in this case. The segmentation obtained by the method in [37] is also shown as a reference. As one can see, the performance of ECQMMF, even with the simplified algorithm used, is highly competitive. This good performance should be attributed mainly to the superiority of the segmentation algorithm itself, which was also shown

TABLE II
TANIMOTO COEFFICIENTS OBTAINED BY ECQMMF FOR EACH INDIVIDUAL VOLUME ON THE DATA SET OF [63]

$1 \times 1 \times 3$mm voxel size			$1 \times 1 \times 1.5$mm voxel size		
Vol.	GM	WM	Vol.	GM	WM
5_8	0.811	0.721	ibsr01	0.850	0.766
4_8	0.783	0.68	ibsr02	0.851	0.802
2_4	0.767	0.665	ibsr03	0.873	0.77
6_10	0.806	0.73	ibsr04	0.867	0.736
15_3	0.762	0.661	ibsr05	0.842	0.781
16_3	0.782	0.686	ibsr06	0.790	0.762
17_3	0.821	0.752	ibsr07	0.855	0.819
8_4	0.801	0.738	ibsr08	0.843	0.805
7_8	0.833	0.77	ibsr09	0.862	0.813
110_3	0.839	0.748	ibsr10	0.865	0.828
111_2	0.836	0.77	ibsr11	0.868	0.836
112_2	0.836	0.756	ibsr12	0.856	0.806
100_23	0.85	0.778	ibsr13	0.882	0.753
202_3	0.842	0.762	ibsr14	0.887	0.811
191_3	0.821	0.755	ibsr15	0.886	0.794
12_3	0.859	0.787	ibsr16	0.887	0.776
13_3	0.849	0.777	ibsr17	0.870	0.758
1_24	0.836	0.768	ibsr18	0.877	0.772
205_3	0.825	0.769	-	-	-
11_3	0.84	0.771	-	-	-
Averg.	**0.819**	**0.742**	**Averg.**	**0.862**	**0.788**

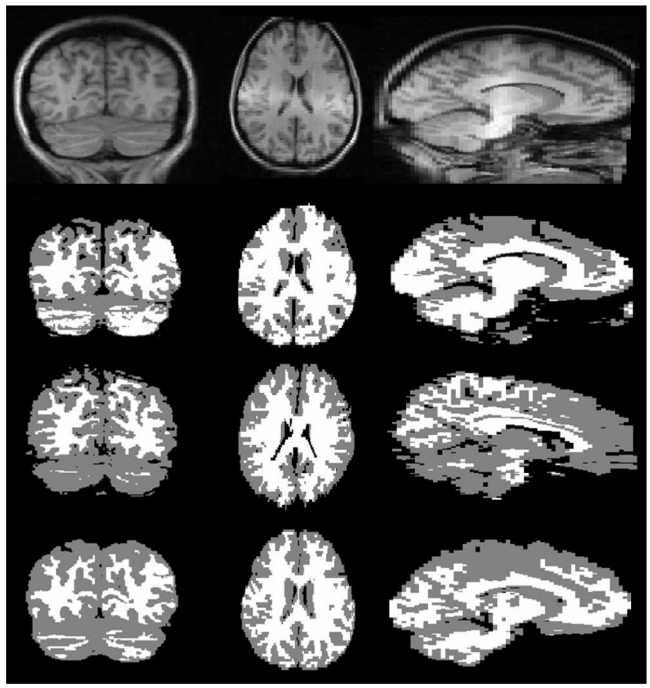

Fig. 6. Segmented results for the volume 2–4 from [63] ($1 \times 1 \times 3$ mm). From top to bottom: data, expert segmentation, MPM-MAP segmentation and EC-QMMF segmentation. The columns correspond to coronal, axial and sagittal views, respectively. Note that the pixels in columns 2 and 3 are rectangular to reflect the voxel size.

in the experiments with synthetic images described above. Another reason may be related to the fact that the noise variance is estimated for each individual slice separately, which effectively adapts the amount of smoothing to the local granularity of the data.

Our method may be extended to 3-D in a straightforward way; it is enough to add to the energy (14) the term

$$\lambda_{\text{inter}} \sum_{\langle x,y \rangle_I} \|b(x) - b(y)\|^2 \tag{27}$$

where the sum is taken over all nearest-neighbor pairs of voxels that belong to different slices and λ_{inter} is a positive parameter that controls the degree of interslice coupling. We performed the segmentations of the 18 1.5-mm-thick volumes from [63] using the 3-D version, hand-adjusting the value of λ_{inter} to 0.1, which gave the best results. The computation time increased from an average of 3.2 h per volume in the 2-D version to 4.1 h per volume (on a 3-GHz machine), and we obtained only a marginal increase (less than 0.5 %) in the average Tanimoto coefficients, so that it may not be worthwhile to use this additional complexity. We think that the reason why one obtains such a small improvement with the 3-D version is twofold: on one hand the performance of the 2-D segmenter is already quite good, and it introduces practically no artifacts, as one can verify from the axial and sagittal views of the segmented volumes (see Fig. 7 for an example segmentation). On the other hand, the hand-made expert segmentations are quite likely also done slice by slice, so that a good 2-D method should approximate them quite well.

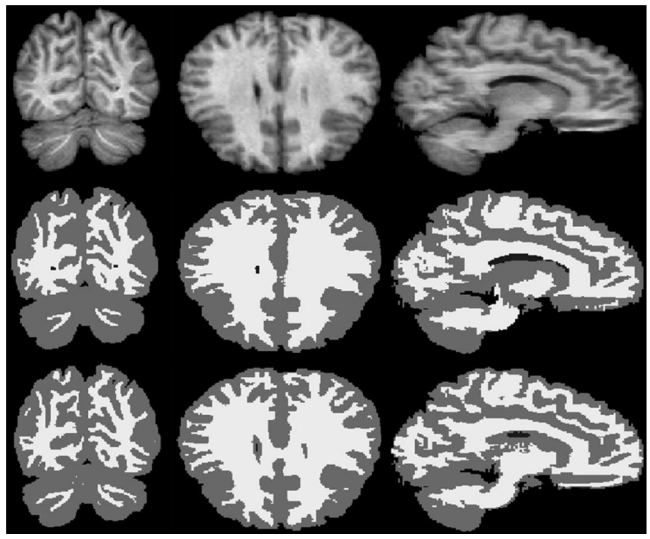

Fig. 7. Segmented results for the volume ibsr02 from [63] ($1 \times 1 \times 1.5$ mm). From top to bottom: data, expert segmentation and EC-QMMF segmentation. The columns correspond to coronal, axial and sagittal views, respectively. Note that the pixels in columns 2 and 3 are rectangular to reflect the voxel size.

IV. DISCUSSION AND CONCLUSIONS

We have presented a general model for parametric segmentation based on a hidden QMMF model with controlled entropy. This method is closely related to the ones in [37] and [16], which also use a Markov measure field model with prior quadratic potentials. The main contribution of this work derives from a new quadratic term that is included in the prior energy,

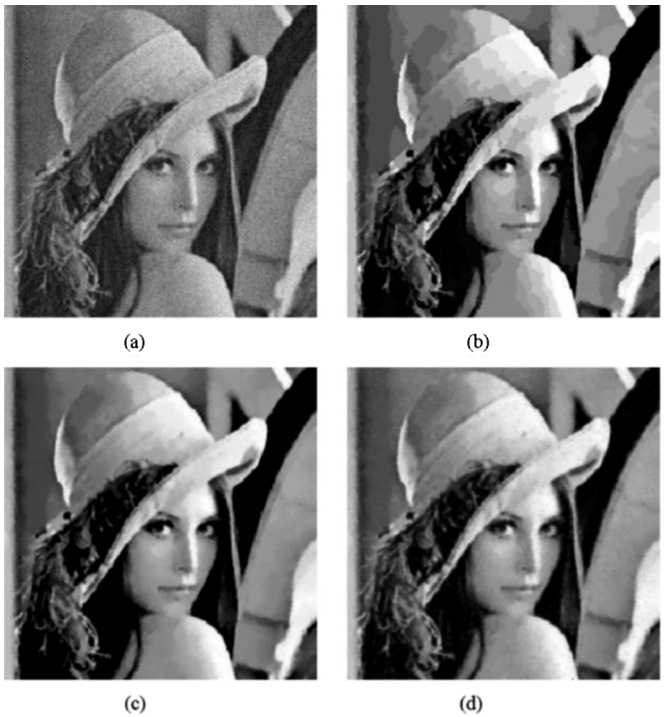

Fig. 8. Edge preserving image restoration using different estimators provided by EC-QMMF (see text).

which controls the entropy of the discrete distributions associated with each pixel. This term plays a crucial role, since it permits the derivation of a data term which, unlike the one in [16], is quadratic on the b variables given $\bar{\theta}$, and, in the important case of Gaussian noise, is also quadratic in $\bar{\theta}$ given b, unlike [37] and [44]. This term also improves the convergence of the method when $\bar{\theta}$ and b are simultaneously estimated, so that it leads to optimal estimators that are both accurate and efficient to compute, with hyperparameters that are easy to tune. This scheme is quite flexible: it allows for piecewise constant or piecewise smooth solutions, such as the ones presented in the previous section. Besides, unlike other methods that produce only "hard" segmentations, the fact that in our case one has a probability vector associated to each pixel, opens the possibility of using optimal estimators different from the mode, such as the mean or median, which in certain cases may be preferable. As an example, consider the edge-preserving image restoration problem illustrated in Fig. 8. Using $M = 15$ piecewise constant intensity models $\Phi(x;\theta_k) = \theta_k$, where θ_k now denotes a scalar, one may segment the noisy input image of Fig. 8(a) and obtain, taking the mode of each probability vector, the piecewise constant restoration of Fig. 8(b); as one can see, the noise has been eliminated, but the discretization causes many artifacts. If one takes the mean or the median (obtained via linear interpolation of the cumulative distribution) of each $b(x)$, however, one obtains the piecewise smooth restorations of Fig. 8(c) and (d), respectively, that have a much better appearance.

In the examples presented so far, we have used a constant regularization parameter for the complete image. It is also possible to use the general, spatially varying β field that was introduced in (7) to introduce a coupling between different perceptual modalities via a spatially varying field that modulates the amount of smoothness imposed over the field b. As an illustration, consider the problem of stereo disparity estimation, which is one of the most extensively studied problems in computational vision (see, for instance, references in [65] and, recently, [21]–[23], and [66]–[68]). Consider, without loss of generality, that the optical axes of the two cameras that take the stereo pair are parallel, so that the epipolar lines are horizontal (otherwise, the images may be prewrapped, so that this condition is fulfilled). The observation model takes the form

$$I_R(x) = I_L(x + \bar{d}(x)) + n(x) \tag{28}$$

where I_L, I_R are the right and left images, $\bar{d}(x) = (d(x), 0)$ is the disparity vector associated to pixel x, and n is a zero-mean, Gaussian, white noise field with variance σ^2. If one assumes M constant, integer-valued disparity models $\{\theta_k, k = 1, \ldots, M\}$ that span the total disparity range of the stereo pair, one may compute the likelihood as

$$v_k(x) = \frac{1}{Z} \exp\left[-\frac{1}{2\sigma^2}(I_R(x) - I_L(x + \bar{d}_k(x))^2\right]$$

where $\bar{d}_k(x) = (\theta_k, 0)$ and Z is a normalizing constant.

One may now modify the model (14) by including in the prior a term that expresses the constraint that disparity discontinuities often coincide with intensity edges in the reference image. In particular, if one defines an edge indicator variable e_{xy} to be equal to 1 if there is an intensity edge between pixels x and y of I_R, and equal to zero otherwise, one may use the general prior energy (7) with $\beta_{xy} = \lambda[\epsilon e_{xy} + (1-\epsilon)(1-e_{xy})]$, where ϵ is a parameter (we use $\epsilon = 0.1$), so that the presence of an intensity edge partially decouples the measure field b at sites x and y and, thus, favors the appearance of a disparity discontinuity. As a result, the term $\lambda \sum_{\langle x,y\rangle} \|b(x) - b(y)\|^2$ in (14) is replaced by

$$\lambda \sum_{\langle x,y\rangle} \|b(x) - b(y)\|^2[\epsilon e_{xy} + (1-\epsilon)(1-e_{xy})].$$

The field e may be precomputed in a variety of ways; here, we simply set $e_{xy} = 1$ if there is a zero crossing of the Laplacian of I_R between pixels x and y and $|I_R(x) - I_R(y)|$ is greater than 0.2 times the dynamic range of I_R.

A piecewise constant disparity field may now be estimated by running the procedure described in Section III. Note that one is using fixed, integer-valued models in this case, so that the model update step is omitted. An example of the obtained results (using the standard "Tsukuba" stereo pair [65], [69], [70]) is presented in Fig. 9. Fig. 9(c) shows the disparity field estimated with the constant interaction field $\beta_{xy} = \lambda$. As one can see, there is an apparent over-smoothing of some of the boundaries between constant disparity regions, due to the fact that in areas with low texture (i.e., almost constant intensity) disparity is not well defined, since the likelihood term may give almost the same value for a whole range of disparities. Therefore, the smoothness term of the posterior energy dominates and produces a partial "spilling" over these areas.

The disparity obtained using the spatially varying interaction field β is shown in Fig. 9(f). As one can see, the inclusion of this field effectively improves the localization of the disparity

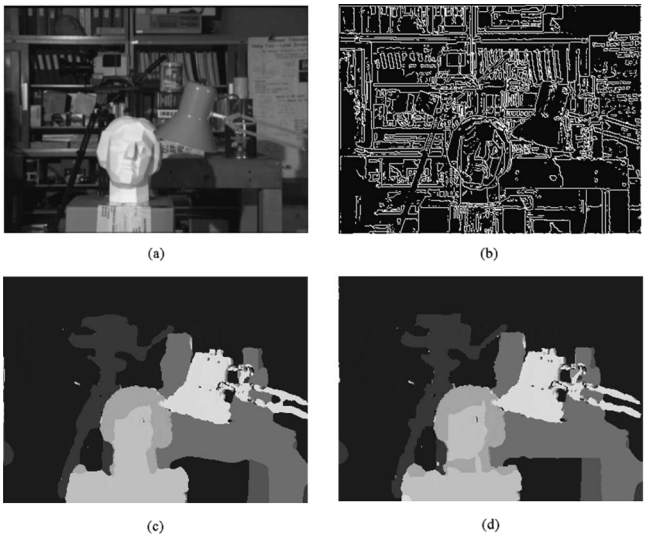

Fig. 9. (a) Frame from the Tsukuba sequence. (b) Simple edge detection (see text). (c), (d) Computed piecewise constant disparity without and with edge information, respectively (see text).

discontinuities; note that this is possible only because of the entropy control term, which makes $b(x)$ and $b(y)$ peaked at different disparities at opposite sides of a discontinuity. These results should be taken only as an illustration of the potential of this approach for solving the stereo disparity computation problem. A competitive approach should incorporate features like automatic occlusion detection, use of color information, more refined models that obtain disparities with subpixel accuracy, etc. What we think we have shown, however, is that the ECQMMF model presented here may be used as a flexible and computationally efficient component to develop sophisticated systems that produce high quality solutions for this and other specific problems.

This paper extends our previous results in [51].

REFERENCES

[1] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, no. 4, pp. 429–442, Aug. 1996.

[2] J. Rajapakse, J. Giedd, and J. Rapoport, "Statistical approach to segmentation of single-channel cerebral MR images," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 176–186, Apr. 1997.

[3] C. Rother, V. Kolmogprov, and A. Blake, "Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[4] J. Wang and M. Cohen, "An interactive optimization approach for unified image segmentation and matting," *ICCV*, vol. 3, pp. 936–943, 2005.

[5] Y. Weiss and E. Adelson, "A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models," *CVPR*, pp. 321–326, 1996.

[6] H. S. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 814–830, Aug. 1996.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[8] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother, "Probabilistic fusion of stereo with color and contrast for bi-layer segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1480–1492, Sep. 2006.

[9] R. C. P. Smith and T. Drummond, "Layered motion segmentation and depth ordering by tracking edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 479–494, Apr. 2004.

[10] M. A. Ruzon and C. Tomasi, "Alpha estimation in natural images," *CVPR (1)*, pp. 18–25, 2000.

[11] Y. Boykov and M. P. Jolly, "Interactive graph cut for optimal boundary & region segmentation of objects in N-D images," *ICIP (1)*, pp. 105–112, 2001.

[12] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive GMMRF model," *ECCV*, vol. 1, pp. 414–427, 2004.

[13] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "Random Walks for interactive organ segmentation in two and three dimensions: Implementation and validation," *MICCAI (2)*, vol. LNCS 3750, pp. 773–780, 2005.

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[15] S. Li, K. Chan, and H. Wang, "Bayesian image restoration and segmentation by constrained optimization," *CVPR*, pp. 1–6, 1996.

[16] J. L. Marroquin, E. Arce, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, Nov. 2003.

[17] M. Rivera and J. C. Gee, "Two-level MRF models for image restoration and segmentation," *BMVC (2)*, pp. 809–818, 2004.

[18] D. Panjwani and G. Healey, "Markov random field models for unsupervised segmentation of textured color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 10, pp. 939–954, Oct. 1995.

[19] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.

[20] V. Kolmogorov, R. Zabih, and S. J. Gortler, "Generalized multi-camera. Scene reconstruction using graph cuts," *EMMCVPR*, vol. LNCS 2683, pp. 501–516, 2003.

[21] J. Sun, Y. Li, S. B. Kang, and H.-Y. Shum, "Symmetric stereo matching for occlusion handling," *CVPR*, pp. 371–378, 2005.

[22] M. Gong and Y.-H. Yang, "Fast unambiguous stereo matching using reliability-based dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 998–1003, Jun. 2005.

[23] Y. Tsin, S. B. Kang, and R. Szeliski, "Stereo matching with linear superposition of layers," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 28, no. 2, pp. 290–301, Feb. 2006.

[24] S. Birchfield and C. Tomasi, "Multiway cut for stereo and motion with slanted surfaces," *ICCV*, pp. 489–495, 1999.

[25] A. K. Jain and R. C. Dubes, *Algorithm for Clustering Data.*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[26] D. Pham and J. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 737–752, Sep. 1999.

[27] S. C. Zhu and A. L. Yuille, "Region.competition: Unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, Sep. 1996.

[28] A. Blake and M. Isard, *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion.*. New York: Springer, 2000.

[29] P. Perona and W. Freeman, "A factorization approach to grouping," *ECCV (1)*, vol. LNCS 1406, pp. 655–670, 1998.

[30] Y. Weiss, "Segmentation using eigenvectors: A unifying view," *ICCV (2)*, pp. 975–982, 1999.

[31] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[32] C. Samson, L. Blanc-Feraud, J. Zerubia, and G. Aubert, "A variational model for image classification and restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 5, pp. 460–472, May 2000.

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

[34] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Tokyo, Japan: Springer-Verlag, 2001.

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

[36] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Boston, MA: Kluwer, 1998, pp. 355–368.

[37] 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. Med. Imag.*, vol. 21, no. 8, pp. 934–945, Aug. 2002.
[38] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognit.*, vol. 36, pp. 131–144, 2003.
[39] J. Wang and E. Adelson, "Representing images with layers," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 625–628, Sep. 1994.
[40] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Color- and texture-based image segmentation using EM and its application to content-based image retrieval," *ICCV*, pp. 675–682, 1998.
[41] D. Pham and J. Prince, "A generalized em algorithm for robust segmentation of magnetic resonance images," *CISS*, pp. 558–563, 1999.
[42] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using Expectation-Maximization and?its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, Aug. 2002.
[43] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2570–2583, Oct. 1992.
[44] J. L. Marroquin, F. Velazco, M. Rivera, and M. Nakamura, "Gauss–Markov measure field models for low-level vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 337–348, Apr. 2001.
[45] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
[46] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.
[47] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.
[48] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications in early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–92, 1996.
[49] M. Rivera and J. Marroquin, "Adaptive rest condition potentials: First and second order edge-preserving regularization," *J. Comput. Vis. Image Understand.*, vol. 88, pp. 76–93, 2002.
[50] M. Rivera and J. Marroquin, "Half-quadratic cost functions with granularity control," *Image Vis. Comput.*, vol. 21, pp. 345–357, 2003.
[51] M. Rivera, O. Ocegueda, and J. L. Marroquin, "Entropy controlled Gauss–Markov random measure fields for early vision," *VLSM*, vol. LNCS 3752, pp. 137–148, 2005.
[52] Y. Lin, J. Tian, and H. He, "Image segmentation via fuzzy object extraction and edge detection and its medical application," *J. X-Ray Sci. Technol.*, vol. 10, no. 1–2, pp. 95–106, 2001.
[53] 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. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[54] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic resonance image tissue classification using a partial volume model," *Neuroimage*, vol. 13, no. 5, pp. 856–876, 2001.
[55] N. M. John, M. R. Kabuka, and M. O. Ibrahim, "Multivariate statistical model for 3D image segmentation with application to medical images," *J. Digit. Imag.*, vol. 16, no. 4, pp. 365–377, 2003.
[56] A. Dale, B. Fischl, and M. Sereno, "Cortical surface-based analysis I: Segmentation and surface reconstruction, neuroimage," *Neuroimage*, vol. 9, no. 2, pp. 179–194, 1999.
[57] B. Fischl, M. Sereno, and A. Dale, "Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system," *Neuroimage*, vol. 9, no. 2, pp. 195–207, 1999.
[58] J. Ashburner and K. J. Friston, "Unified segmentation," *Neuroimage*, vol. 26, pp. 839–851, 2005.
[59] R. Heckemann, J. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *Neuroimage*, vol. 33, no. 1, pp. 115–126.
[60] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Trans. Med. Imag.*, vol. 10, no. 10, pp. 897–908, Oct. 1999.
[61] A. Evans, D. Collins, S. Mills, E. Brown, R. Kelly, and T. Peters, "3D statistical neuroanatomical models from 305 MRI volumes," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, 1993, pp. 1813–1817.
[62] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, no. 12, pp. 629–639, Dec. 1990.
[63] *Massachusetts General Hospital. The Internet Brain Segmentation Repository (IBSR)*, [Online]. Available: http://www.cma.mgh.harvard.edu/ibsr/
[64] M. A. Gonzalez-Ballester, A. Zisserman, and M. Brady, "Estimation of the partial volume effect in MRI," *Med. Imag. Anal.*, vol. 6, no. 4, pp. 389–405.
[65] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1/2/3, pp. 7–42, 2002.
[66] J. Y. Jang, K. M. Lee, and S.-U. Lee, "Stereo matching using iterated graph cuts and mean shift filtering," *LNCS*, vol. 3851, pp. 31–40, 2006.
[67] O. Veksler, "Stereo correspondence by dynamic programming on a tree," *CVPR (2)*, pp. 384–390, 2005.
[68] A. Criminisi, J. Shorten, A. Blake, C. Rother, and P. Torr, "Efficient dense stereo with occlusions for. new view-synthesis by four-state dynamic programming," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 89–110, 2007.
[69] Y. Nakamura, T. Matsuura, K. Satoh, and Y. Ohta, "Occlusion detectable stereo-occlusion patterns in camera matrix," *CVPR*, pp. 371–378, 1996.
[70] D. Scharstein and R. Szeliski, Middlebury Stereo Vision Page [Online]. Available: http://www.middlebury.edu/stereo

Mariano Rivera received the B.E. degree in electronics from the Durango Institute of Technology, Mexico, in 1989, the M.Sc. degree in electronics from the Chihuahua Institute of Technology, Mexico, in 1993, and the D.Sc. degree in optics from the Center for Research in Optics (CIO), Leon, Mexico, in 1997.

Since 1997, he has been with the Computer Science Department, Center for Research in Mathematics (CIMAT), Guanajuato, Mexico. His current interests include computer vision, image processing, machine learning, and optical metrology. His research is summarized in more than 40 papers in scientific journals and conference proceedings.

Dr. Rivera is Fellow of the National Researcher System (SNI) of the Mexican Government.

Omar Ocegueda received the B.S. degree in computer science from the University of Guanajuato, Mexico, in 2004, and the M.Sc. degree in computer science and industrial mathematics from the Center for Research in Mathematics (CIMAT), Guanajuato, Mexico, in 2006.

His main research interests include computer vision, image processing, machine learning, and human-computer interaction.

Jose L. Marroquin received the B.S. degree in chemical engineering from the National University of Mexico in 1968 and the M.Sc. and Ph.D. degrees in systems science from the Massachusetts Institute of Technology, Cambridge, in 1985.

He was with PEMEX, the Mexican petroleum company, as a Project Leader in geophysical data processing. Currently, he is the Head of the Computer Science Department, Center for Research in Mathematics, Guanajuato, Mexico, and is conducting research related to the computer processing of visual information.

Dr. Marroquin is a Fellow of the National Research System of the Mexican Government and of the Mexican Academy of Science.