Entropy Controlled Gauss-Markov Random Measure Field Models for Early Vision

Mariano Rivera, Omar Ocegueda and Jose L. Marroquin

Centro de Investigacion en Matematicas A.C.
Guanajuato, Gto., MEXICO, 36000
{mrivera,omar,jlm}@cimat.mx

Abstract. We present a computationally efficient segmentation–restoration method, based on a probabilistic formulation, for the joint estimation of the label map (segmentation) and the parameters of the feature generator models (restoration). Our algorithm computes an estimation of the posterior marginal probability distributions of the label field based on a Gauss Markov Random Measure Field model. Our proposal introduces an explicit entropy control for the estimated posterior marginals, therefore it improves the parameter estimation step. If the model parameters are given, our algorithm computes the posterior marginals as the global minimizers of a quadratic, linearly constrained energy function; therefore, one can compute very efficiently the optimal (Maximizer of the Posterior Marginals or MPM) estimator for multi–class segmentation problems. Moreover, a good estimation of the posterior marginals allows one to compute estimators different from the MPM for restoration problems, denoising and optical flow computation. Experiments demonstrate better performance over other state of the art segmentation approaches.

1 Introduction

Image segmentation from different attributes (such as gray level, local orientation or frequency, texture, motion, color, etc.) is typically formulated as a clustering problem. Although generic clustering algorithms as K-Means or ISODATA have been used with relative success [8], the consideration of spatial interactions among pixel labels provides additional, useful constraints on the problem [2][3] [6][7] [9][10] [11][12] [13][14] [17][18] [19][20] [21][22] [23][24].

Therefore, most successful algorithms for image segmentation taked into account both the observed pixel values (which are generally noisy) and pixel context information.

We assume that an image of features, g, in the regular lattice L, is an assembly of K disjoint regions, $R = \{R_1, R_2, ..., R_K\}$. Moreover, such features are generated with a generic parametric model ϕ, with $\theta = \{\theta_1, \theta_2, ..., \theta_K\}$ as the corresponding parameters for each region, i.e.:

$$g_r = \sum_{k=1}^{K} \phi_{kr} b_{kr} + \eta_r \tag{1}$$

where $r = [x, y]^T$ is the position of a pixel in the regular lattice L; $\phi_{kr} \stackrel{def}{=} \phi(\theta_k, r)$ is the parametric model of the actual value of the feature at pixel r; θ_k is the parameter set corresponding to the k^{th} region; b_{kr} is an indicator variable equal to one if the pixel r was generated with the model k and equal to zero otherwise and η is an additive, independent, identically distributed noise process. In the general case, the one we consider here, the parameter set is unknown and needs to be estimated.

Bayesian regularization framework has been successfully used for finding the solution to these problems [6][10] [11][12] [13][14] [18] [19][20] [21][22] [24][25]. In this framework, the solution is computed as a statistical estimator from a posterior probability distribution. In particular, one needs to estimate the label map, b, and the model parameters, θ, from the posterior distribution $P_{b,\theta|g}$. If it is assumed independence between b and θ and a uniform prior distribution for θ, then this posterior distribution is given by:

$$P_{b,\theta|g} = P_{g|b,\theta} P_b / P_g \tag{2}$$

where the likelihood of the whole label field is obtained from a mixture model :

$$P_{g|b,\theta} = \prod_k \prod_r (v_{kr})^{b_{kr}};$$

with v_{kr} as the probability (individual likelihood) that the observed value at pixel r was generated with model k (that uses the set of parameters θ_k). For instance, if η is Gaussian with zero mean and variance σ^2:

$$v_{kr} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(g_r - \phi_{kr})^2}{2\sigma^2}\right], \tag{3}$$

for a real valued feature image g. In the framework of Bayesian Estimation Theory, b is modelled as a Markov Random Field (MRF) with prior distribution, P_b, in the form of a Gibbs distribution:

$$P_b = \frac{1}{z} \exp\left[-\beta \sum\nolimits_C V_C(b)\right]; \tag{4}$$

where z is a normalization constant and V_C is a potential such that it assigns larger probabilities to smooth label fields than to granular ones and β is a positive parameter. The most popular potential is the Ising model:

$$V_{krs}(b) = \begin{cases} -1 \; if \; b_{kr} = b_{kr} \;\; \forall k \\ \;\; 1 \; if \; b_{kr} \neq b_{ks}. \end{cases} \tag{5}$$

Finally, P_g is a normalization constant independent of the unknowns b and θ.

In most cases, approximate solutions for this complex estimation problem are found by 2-step procedures,[2][4][12][14][20] in which the best segmentation, given the parameters is found in the first step, and the optimal estimator for the parameters, given the segmentation is found in the second step, iterating these 2 steps until convergence. Usually, one chooses the models, ϕ in such a way that the maximum a posteriori (MAP) estimator for the parameters θ given b is relatively easy to compute. For instance, flat, planar or spline models have successfully been used [12][14][18]. However, the MAP estimator for the label field requires the solution of a combinatorial optimization problem. Graph–Cuts based algorithms[2][3] [7][19][23] can be used to compute the exact MAP estimator in the case of binary segmentation or an approximation for problems with more than two classes. The problem here is that using a “hard” segmentation in the first step of the procedure, makes the 2-step algorithm prone to get trapped in local minima, producing suboptimal results. A better strategy is to compute, instead of a set of binary indicator variables, their expected value (i.e., the posterior marginal distributions), in which case the 2-step procedure is equivalent to the Expectation-Maximization (EM) algorithm [4][15][20]. Upon convergence, a hard segmentation may be computed, if desired, using, for instance, the MPM estimator [11]:

Definition 1 (MPM Estimator). *The MPM estimator of the label field is given by:*

$$b_{kr}^{MPM} = \begin{cases} 1 \text{ if } \pi_{kr} \geq \pi_{lr}, \text{ for } k \neq l \\ 0 \text{ otherwise}, \end{cases} \tag{6}$$

where π_r is the marginal probability distribution of the pixel r.

An estimation, p, of the true marginals, π, can be computed with Markov Chain Monte Carlo (MCMC) based methods [6][10]. In such a case, samples $\{b^{(1)}, b^{(2)}, \ldots, b^{(N)}\}$, of the posterior distribution, $P_{b,\theta|g}$, are used to compute the empirical marginals: $p_{kr} = \frac{1}{N}\sum_{j=1}^{N} b_{kr}^{(j)}$, that satisfy

$$E[p_{kr}] = \pi_{kr} \tag{7}$$

where $E[\cdot]$ denotes the expectation operator. The problem with these methods lies in their high computational cost.

A more efficient approach considers the empirical marginals as a vector-valued random field (i.e., a random measure field) that needs to be modeled. There are 2 main models that have been proposed: one based in a Mean Field (MF) approximation [20][24], and the other in a Gauss-Markov Measure Field (GMMF) model [13]. Both of them, however, have certain drawbacks: the MF approach leads to algorithms that are relatively slow and sensitive to noise, while the GMMF approach presented in [13] produces estimators for the marginal distributions that differ significantly from the true ones, in the sense that these distributions (one for each pixel) have significantly higher entropy than the ones found asymptotically by MCMC approaches. If the model parameters are known, this difference is not too important, since usually the modes of these distributions (and hence, the MPM estimator) are correct; if the model parameters are unknown, however, this high entropy produces an unstable behavior of the EM algorithm, producing bad results. The goal of this paper is to present a better GMMF model for the empirical marginals that produces estimates that are in agreement with the true ones, and that can be efficiently computed.

2 Entropy Controlled GMMF (EC-GMMF) Models

The use of GMMF models for estimating the posterior marginal distributions is based in the following:

Theorem 1 (Gauss-Markov Measure Fields (GMMF)). *Let the binary label field b be a Markov random field (MRF) with posterior distribution (2) and v the likelihood field, then the empirical marginal field, p, is itself a MRF with posterior distribution:*

$$P_{p|v} = P_{v|p} P_p / P_v \tag{8}$$

with the following properties:

1. *It is Markovian with the same neighborhood system as b.*
2. *It is Gaussian, i.e. $P_{p|v} \propto \exp\left[-U(p;v)\right]$, where the energy $U(p;v)$ is a Quadratic, Positive Definite (QPD) function of p.*

The proof is presented in [13]. This theorem establishes important properties of the marginal probabilities, π, of the label field, b, with posterior distribution (2), but it does not determine the exact form of the QPD energy U. Given (7), and Theorem 1, π can be estimated as the MAP estimator of (8). In order to find a particular form for U, an additional consistency constraint is imposed in [13]:

Consistency Constraint 1 (GMMF) *If no prior information is provided [i.e. P_π is the uniform distribution] then the maximizer of (8) is $\pi^* = \bar{v}$, where $\bar{v}_{kr} = v_{kr} / \sum_j v_{jr}$.*

Based on these properties, the function U that is proposed in [13] is:

$$U(p;v) = \sum_{k=1}^{K} \sum_{r \in L} \left[(p_{kr} - \bar{v}_{kr})^2 + \frac{\lambda}{2} \sum_{s \in \mathcal{N}_r} \|p_{kr} - p_{ks}\|_2^2 \right], \tag{9}$$

where $\mathcal{N}_r = \{s \in L \colon |r - s| = 1\}$ is the set of nearest neighbor pixels to r. In spite of the fact that the minimization of (9) can be done with computationally efficient algorithms, the use of this equation has two disadvantages: first, it depends on the model parameters (via the likelihoods v) in a highly non-linear way (3), which makes difficult its incorporation in EM procedures, and second, the set of distributions ($\{p_r\}$ field) that minimizes (9) are relatively flat (i.e., they have high entropy), which makes them unsuitable for EM procedures.

In order to propose a new QPD function U that overcomes these difficulties, we need to relax the consistency constraint 1; the new constraint is:

Consistency Constraint 2 (EC-GMMF) *If no prior information is provided, the mode of the optimal estimators for the posterior marginal distributions π^* coincides with the maxima of the corresponding likelihoods v, i.e., the MPM estimator for the label field, computed using π^* coincides with the maximum likelihood estimator.*

With this relaxed constraint, we may introduce $\log v$ instead of v in the data term, to get a quadratic dependence on the model parameters θ (assuming a Gaussian noise model). Entropy control is introduced by adding a penalization term of high entropy distributions; to keep quadratic the energy function, we use the Gini coefficient [5] (instead of the Shannon's entropy [22]): $-\sum_k \sum_r p_{kr}^2$; so that we finally get:

$$U_{EC}(p, \theta) = \sum_k \sum_r \left[p_{kr}^2 \left(-\log v_{kr} - \mu \right) + \frac{\lambda}{2} \sum_{s \in \mathcal{N}_r} (p_{kr} - p_{ks})^2 \right] \tag{10}$$

subject to

$$\sum\nolimits_k p_{kr} = 1,\ \ \forall r \quad \text{and} \quad p_{kr} \geq 0,\ \ \forall k, r; \tag{11}$$

where the parameter μ controls the entropy of the marginals. Note that for $\mu < 2\lambda$, we can assure that (10) is QPD. It is found that the performance of the estimation algorithms does not depend critically on the precise value for μ; we have used $\mu = 0.1\lambda$ in all the experiments reported here.

2.1 Computation of the Optimal Estimators

The minimization of (10) may be carried out by a 2-step (EM) procedure. The MAP estimators for the marginals, p^{MAP}, and the models, θ^{MAP}, are computed by iterating alternated minimizations of (10) w.r.t. p and θ, respectively. These minimizations have the following interesting properties:

Theorem 2 (EC-GMMF Convergence). *Assuming that U_{EC} is QPD with respect to p and V_{kr} is a uni-modal distribution, we have:*

(i) If θ is given, the problem of minimizing U_{EC} w.r.t. p, subject to the constraints (11) has a unique local minimum, which coincides with the constrained global minimum.
(ii) If p is given, the problem of minimizing U_{EC} w.r.t. θ has a unique local minimum, which coincides with the global minimum.
(iii) The iterated alternate minimizations of U_{EC} w.r.t. p and θ converges, at least, to a local minimum.

The proof of (i) and (ii) follows from the facts that U_{EC} is a QPD function of p for fixed θ, and of θ for fixed p, and the constraints (11) are linear (see [16]). (iii) Follows from (i) and (ii) and from the fact that $U_{EC} \geq 0$.

Last theorem establishes that any descent algorithm will converge to the global minimum in the E and M steps of the EM procedure. In particular, if the model parameters are given, one may find very efficiently the optimal (MPM) segmentation even for multi-class segmentation problems, which represents a significant advantage over algorithms like those based on graph cuts, which guarantee global optimality only for 2-class problems.

In the algorithm we propose here, the equality constraint in (11) may be incorporated using the Lagrangian method: the Lagrangian, that incorporates the equality constraints, is given by:

$$\mathcal{L}(p, \theta) = U_{EC}(p, \theta) - \sum_r \gamma_r \big(1 - \sum\nolimits_k p_{kr}\big) \tag{12}$$

where γ are the Lagrange multipliers of the equality constraints. Now, we define $n_{kr} \stackrel{def}{=} \lambda \sum_{s \in \mathcal{N}_r} p_{ks}$ and

$$m_{kr} \stackrel{def}{=} (-\log v_{kr} - \mu) + \lambda \sharp \mathcal{N}_r. \tag{13}$$

where $\sharp S$ is the cardinality of the set S. Equating to zero the gradient of (12) w.r.t. p, solving for p_{kr} and substituting in the equality constraint (11), one finally obtains the Gauss-Seidel update equation:

Algorithm 1 Gauss-Seidel Implementation of Parametric Segmentation

1: Set the initial model parameters θ_0 and initially set $p_0 = v$;
2: Given the tolerance $\epsilon > 0$;
3: **for** $i = 1, 2, \ldots$ **do**
4:   **for all** the pixels r **do**
5:     **for all** the models k **do**
6:       Compute p_{ikr} with (14);
7:       Project $p_{ikr} = \max\{0, p_{ikr}\}$;
8:     **end for** all the models
    {The renormalization of the p_{ir} can be performed here}
9:   **end for** all the pixels
10:   Update the models θ_i with (15);
11:   **if** $\|p_i - p_{i-1}\| < \epsilon$ **then**
12:     STOP with solution $p^*_{kr} = p_{ikr}$ and $\theta^* = \theta_i$;
13:   **end if**
14: **end for**

$$p_{kr} = \frac{n_{kr}}{m_{kr}} + \frac{1 - \sum_{l=1}^{N} \frac{n_{lr}}{m_{lr}}}{\sum_{l=1}^{N} \frac{m_{kr}}{m_{lr}}}. \tag{14}$$

Note, however, that p_{kr} computed with (14) does not necessarily satisfy the non-negativity constraint in (11). If such a constraint is violated, one makes the negative p_{kr} equal to zero, and renormalizes the vector p_r. In our experiments we found that this simple method works properly and is faster than more sophisticated methods(such as gradient projection).

In practice, one gets better performance, in terms of computational efficiency, if the θ variables are updated after every Gauss-Seidel iteration, instead of waiting until convergence of the E step, i.e. by using a Generalized EM algorithm (GEM) [15]. One gets then a direct procedure, in which the p and θ variables are simultaneously optimized.

Given that θ is not constrained, it may be computed, after every update of the p field, by

$$\theta^{MAP} = \arg\min_{\theta} \sum_k \sum_r \left[-p_{kr}^2 \log v_{kr}(\theta) \right]. \tag{15}$$

The complete procedure is summarized in algorithm 1. Line 10 in such an algorithm is generic and depends on the specific feature model; in the next section we present a particular case.

3 Experiments

The purpose of the experiments in subsection 3.1 is to evaluate the relative performance of the EC-GMMF model with respect to other state of the art segmentation methods. Three aspects are evaluated: Noise robustness, computational efficiency and entropy control. In subsection 3.2 we show an application of our method to other Computer Vision problems: Image Denoising and Optical Flow estimation. We also show that using the same p^* obtained, we can directly compute different estimators that serve as solutions to Piecewise Constant and Piecewise Smooth regularization.

3.1 Numerical Experiments

The normalized test images (two models synthetic image and Lenna's portrait) were corrupted with additive Gaussian noise with mean zero and variance 0.30 and 0.05 respectively. The first experiment, illustrated by Fig. 1, demonstrates the robustness of

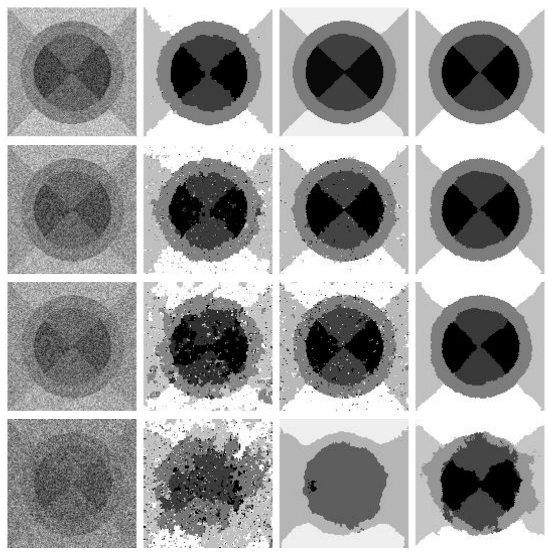

Fig. 1. Segmentation method performance for different level noise, see text for details.

the EC-GMMF method to noise. The task is to segment and estimate the models' parameters for a five models synthetic image with levels 1,2,3,4 and 5. First column in Fig. 1 shows the data corrupted with Gaussian noise with: $\sigma = 0.7, 1.0, 1.2, 2.0$, respectively. The second column shows the segmentations computed with a multi–way graph cut based algorithm. This algorithm has the drawback of being based on a MAP criterion, and it is known that for low SNR data the MPM estimator exhibits better performance. The third column shows the results computed with other state of the art parametric segmentation method, namely HMMF, which in [14] is shown to have better performance than the Mean Field and MCMC-based EM procedures. Our results are consistent with the ones reported by the authors [14]: HMMF models are more robust to noise than graph-cuts based algorithms; however, we found that HMMF algorithm is very sensitive to the precise selection of the parameters' values: the noise variance, the regularization parameter, the initial p–field values and the minimization algorithm parameters (i.e. the friction coefficient and the step size for the gradient projection Newtonian descent algorithm [14]). In part, such a difficulty lies in the fact that the energy function, in the HMMF model, is not convex, which makes the descent algorithm prone to be trapped by local minima. Last column shows segmentations computed with the proposed EC-GMMF method. The experiment shows the superior performance of the proposed method: EC-GMMF produces acceptable segmentations even for low SNR data and in a fraction of the computational time of the compared methods. The number of iterations for all the algorithms were 500 in all cases. The initial estimates for the models ($\phi_{rk} = \theta_k$, with $\theta_k \in \Re$) where uniformly distributed in the dynamic range of the noisy data; we initialize $p_{0kr} = v_{kr}$ and $p_{0kr} = 1/5$ for the EC-GMMF and the HMMF algorithms, respectively. In this case $-log v_{kr} = (g_r - \theta_k)^2$. So that step 10 in algorithm 1 is computed, for each θ model at the i^{th} iteration, with: $\theta_{ik} = \frac{\sum_r g_r p_{ikr}^2}{\sum_r p_{ikr}^2}$.

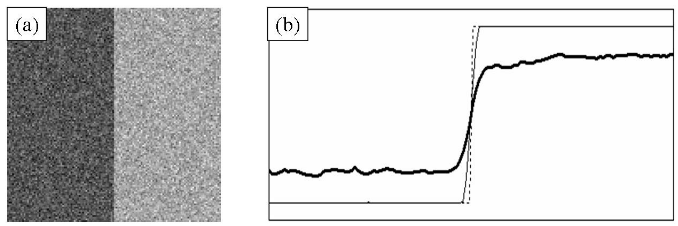

Fig. 2. Comparison of the robustness to noise of different methods. (a) Test image. (b) Computed marginals corresponding to the central row (see text).

Figure 2-(b) shows a comparison of the computed marginals for the central row in figure 2-(a). The model parameters are in this case assumed known, i.e. $\phi_{1r} = 0$ and $\phi_{2r} = 1$, for all the pixels r. The thin solid line corresponds to the p_2 marginals computed with the Gibbs Sampler algorithm (a MCMC method) after 2,000 iterations, with a computational time of 15.92 sec. The heavy line shows the marginals computed with the original GMMF formulation (with $\lambda = 10$ in $0.18 secs.$) and the dotted line, the marginals computed with the EC-GMMF method in 1 second (with $\lambda = 10$ and $\mu = 3$ in $0.33 secs.$). The proposed EC-GMMF approximates very closely the marginals computed with the MCMC method but at a fraction of the computational time. If no entropy control is applied in the EC-GMMF formulation (i.e., for $\mu = 0$), then the computed marginals are similar to the ones computed with the original GMMF method.

For the multi–class problem such entropy reduction can be observed in the maximum p-value maps (i.e $\max_k p_{kr}$). Figure 3 shows the max p–values for the case of 10 models computed with: (a) Gibbs Sampler (comp. time, 318.12 secs.), b) original GMMF (comp. time 0.75 secs.) and (c) EC-GMMF (comp. time, 2.03 secs.), respectively. The apparently low entropy of the Gibbs sampler results may be explained by an incomplete convergence, even after about 5 min.

Next experiment compares the computational efficiency of the EC-GMMF as the number of models is incremented. Parametric segmentations of the Lena's portrait were performed with the HMMF and the EC-GMMF algorithms. Figure 4 shows the corresponding computational times; the Gibbs sampler method (not shown in the plot) required 5 min for 10 models.

Gibbs Sampler GMMF EC-GMMF

Fig. 3. Maximum value of the marginals: Dark pixels denotes low (high entropy) values.

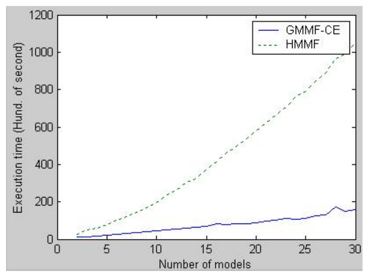

Fig. 4. Comparison of the computational times for the EC-GMMF and HMMF algorithms for different number of models.

3.2 Optimal Estimators for Piecewise Smooth Reconstruction

The fact that one can obtain very precise approximations for the marginals, allows one to compute estimators other than the MPM, which give very good results for piecewise smooth restoration problems. These estimators may be obtained even with fixed models that sample uniformly the search space, producing highly efficient methods. The idea is to compute, instead of the posterior mode (MPM estimator), the mean or median of the estimated posterior marginal distribution at each pixel. The resulting estimators will have sharp discontinuities when these are present in the image, and produce smooth transitions between adjacent models in other places. The posterior mean is computed using: $\bar{f}_r = \sum_k \theta_k p_{kr}$.

This is illustrated in Fig. 5, where the first row shows the results for the piecewise smooth restoration of a noise-corrupted Lena image. Second row in Fig. 5 shows details of the corresponding images in first row. A similar procedure may be used for the computation of piecewise smooth optical flow from a pair of images [1][12][14][22]. In this case, the models are 2-vectors that correspond to discrete velocities that sample the space of allowed displacements: $\phi_{ij} = \theta_{ij} = [u_i, v_j]^T$ where we make a slight notation change by substituting the index model k by the more intuitive pair ij. For the example of Fig. 6 we use $u_i = v_i = d_m(2i/\Delta - 1)$ for $i = 0, 1, \ldots, \Delta$), where d_m is the largest expected displacement and Δ is number of models in $[-d_m, d_m]$; we use $d_m = 2.5$ and $\Delta = 6$ with bi–cubic interpolation for the fractional displacement that corresponds to consider displacements of $\{\pm 2.50, \pm 1.66, \pm 0.83, 0\}$ in the x and y directions. Now, let f_1 and f_2 be two consecutive frames from a sequence. Then, assuming Gaussian noise, we have: $-\log v_{ijr} = [f_1(r) - f_r(r - \theta_{ij})]^2$. One may now

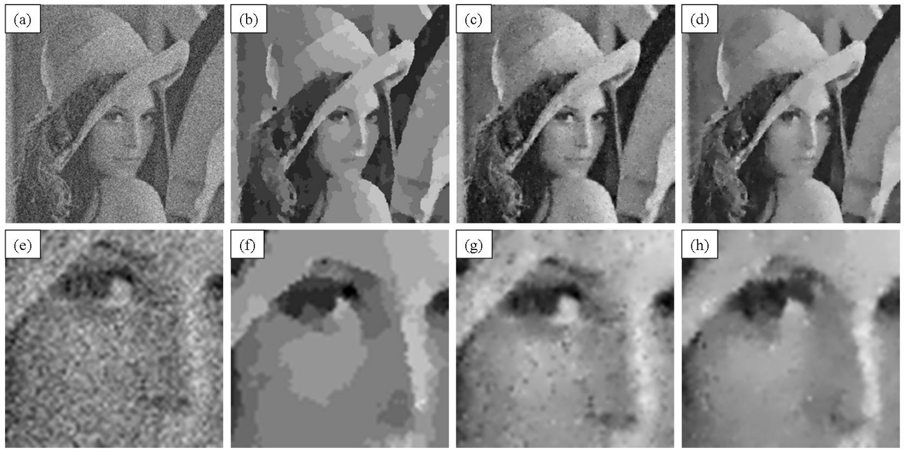

Fig. 5. First row: Explicit entropy control allows us restore images by using estimators other than the MPM see text). Second row: Details.

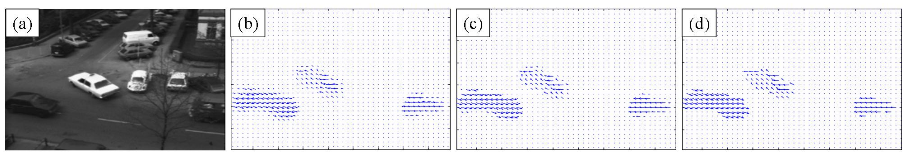

Fig. 6. Optical flow computation (a) Frame 8 of the Hamburg taxi and optical flow magnitude estimator with (b) $n = 1$, (c) $n = 2$ and (d) $n = \infty$ (mode).

compute an optimal estimator that interpolates smooth velocities between neighboring models, while preserving the motion discontinuities. Note that since in this case the models are only partially ordered, one cannot compute the median. Instead, one may compute a family of estimators that include the mode and the mean as special cases:

$$\bar{f}_r = \sum_{ij} \theta_{ij} w_{ijr} \tag{16}$$

with $w_{ijr} = p^n_{ijr} / \sum_{ij} p^n_{ijr}$ The parameter n controls the sharpness of the transitions between models: for large n the estimator corresponds to the mode, while for $n = 1$ one gets the mean.

Since the θ parameters are fixed, one finds the optimal p as the unique minimizer of U_{EC} using the Gauss-Seidel algorithm with projection that was described above. The only change that is necessary is in the definition of m_{kr} in (13), which changes to:

$$m_{ijr} \stackrel{def}{=} [f_1(r) - f_r(r - \theta_{ij})]^2 - \mu + \lambda \text{card}(\mathcal{N}_r). \tag{17}$$

Note that since (17) does not depends on p, m_{ijk} can be precomputed, and the computational cost is the same as that for the case of image restoration with fixed constant models.

Figure 6 shows the computed optical flow with the EC-GMMF method for different estimators. Panel (a) shows the frame 8 of the Hamburg taxi sequence (image f_1) the other image corresponds to the frame 9 (no shown). Panel (b) shows the mode optical flow, i.e. the vector field θ_{ij} corresponding to the largest marginal at each pixel. Panel (c) show the mean flow and panel(d) shows a robust mean computed with (16) and $n = 2$. The mean optical flow is obtained with a sub-pixel resolution and the edges are better preserved by the robust mean.

4 Summary and Conclusions

We have proposed an efficient parametric segmentation method based on Bayesian estimation with prior MRF models and the Expectation-Maximization (EM) procedure. This method estimates the model parameters and the posterior marginals in successive steps as minimizers of QPD energy functions subject to linear constraints, so that each step in the EM procedure has a unique minimum. We also showed that it is possible to implement the estimation process as a Generalized EM algorithm, in which one performs the minimization of the posterior energy with respect to the model parameters and the posterior marginals simultaneously, which decreases significantly the computational cost.

The key point for the superior performance of our method is the introduction of a quadratic term (derived from the Gini coefficient) that controls the entropy of the posterior marginals. This performance is demonstrated by numerical experiments that compare our approach with other state-of-the-art algorithms, such as minimum graph-cut and HMMF methods. The numerical experiments performed demonstrate that the proposed algorithm is more robust to noise and to the initial values for the parameters and significantly more efficient from a computational viewpoint.

It is important to remark that for the case of fixed models, the algorithm reduces to a single E step, and the posterior marginals computed with a simple and efficient Gauss-Seidel procedure correspond to the global optimum. Since these marginals are entropy-controlled, they approximate very well the true ones, and may be used to compute estimators different from the posterior mode (MPM), such as the posterior mean or median. These estimators have the property of interpolating smoothly the estimated feature between neighboring constant models, while preserving the discontinuities in the solution. Two early vision applications that take advantage of this fact are presented: piecewise smooth image reconstruction and optical flow computation.
This work was partially supported by CONACYT, Mexico (grants 40722 and 46270).

References

1. S. Birchfield and C. Tomasi: Multiway cut for stereo and motion with slanted surfaces. ICCV (1999) 489-495.
2. Y. Boykov, O. Veksler, and R. Zabih: Fast approximate energy minimization via graph cuts. IEEE–PAMI, 23 (2001) 1222–1239.
3. Y. Boykov and M.-P. Jolly: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. ICCV, I, (2001) 105112.

4. A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Roy. Statist. Soc. B 39 (1977) 138 .
5. R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, 2nd. ed., John Wiley & Sons, Inc., New York, 2001
6. S. Geman and D. Geman: Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. IEEE–PAMI, 6 (1984) 721–741, .
7. D. Greig, B. Porteous, and A. Seheult: Exact maximum a posteriori estimation for binary images. J. R. Statist. Soc. B, 51 (1989) 271279.
8. A.K. Jain and R.C. Dubes:Algorithm for Clustering Data. Englewood Cliffs, NJ: Prentice Hall (1998).
9. V. Kolmogorov and R. Zbih: What energy functions can be minimized via graph cuts. ECCV III (2002), 65–81.
10. S.Z. Li: Markov Random Field Modeling in Image Analysis. Springer-Verlag, Tokyo (2001).
11. J. Marroquin, S. Mitter and T. Poggio: Probabilistic solution of ill–posed problems in computational vision. J. Am. Stat. Asoc., 82 (1987) 76-89.
12. J.L. Marroquin, S. Botello, F. Calderon and B.C. Vemuri: The MPM-MAP algorithm for image segmentation. ICPR (2000).
13. J. L. Marroquin, F. Velazco, M. Rivera and M. Nakamura: Gauss-Markov Measure Field Models for Low-Level Vision. IEEE–PAMI, 23 (2001) 337–348.
14. J. L. Marroquin, E. Arce and S. Botello: Hidden Markov Measure Field Models for Image Segmentation. IEEE–PAMI, 25, (2003) 1380-1387.
15. R. Neal and R. Barry: A vew of the EM algorithm that justifies incremental, sparse, and others variants. in Learning in Graphical Models, M. Jordan (ed.) Dordrecht: Kluwer Academic Publishers, Boston MA. (1998) 355-368.
16. J. Nocedal and S. Wright: Numerical Optimization (Springer Series in Operational Resarch) Springer Verlag, New York (1999).
17. O. Picher, A. Teuner and B. Hosticka: An unsupervised texture segmentation algorithm with feature space reduction and knowledge feedback. IEEE Trans. Image Process. 7 (1998) 53–61.
18. M. Rivera and J.C. Gee:Image segmentation by flexible models based on robust regularized networks. ECCV, III (2002) 621–634.
19. M. Rivera and J.C. Gee, ”Two-level MRF models for image restoration and segmentation,” BMVC, 2 (2004) 809-818.
20. Tsai, J Zhang and A Willsky: Expectation-Maximization Algorithms for Image Processing Using Multiscale Methods and Mean Field Theory, with Applications to Laser Radar Range Profiling and Segmentation. Opt. Engineering, 40, 7, (2001) 1287-1301.
21. Z. Tu, S.C. Zhu and H.Y. Shum: Image Segmentation by Data Driven Markov Chain Monte Carlo. ICCV (2001) 131–138.
22. Y. Weiss and E.H. Adelson: A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. CVPR (1996) 321–326.
23. Z. Wu and R. Leaby: An optimal graph theoretical approach to data clustering: Theory and its applications to image segmentation. IEEE–PAMI, 11 (1993) 1101–1113.
24. J. Zhang: The mean field theory in EM procedures for Markov random fields. IEEE Trans. Signal Processing, 40 (1992) 2570–2583.
25. A. Tsai, J Zhang, A Wilsky: Multiscale Methods and Mean Field Theory in EM Procedures for Image Processing. Eight IEEE Digital Signal Processing Workshop (1998).