ON OPTIMIZATION OF PROBABILITY VECTOR RANDOM FIELDS USED FOR IMAGE SEGMENTATION

G. Sfikas[1,2], C. Nikou[1], C. Heinrich[2], N. Galatsanos[3]

[1]Department of Computer Science, University of Ioannina, Greece,
[2] LSIIT (UMR 7005 CNRS–UDS), University of Strasbourg, France,
[3]Department of ECE, University of Patras, Greece.

ABSTRACT

In the context of image segmentation, Markov random fields (MRF) are extensively used. However solution of MRF-based models is heavily dependent on how successfully the MRF energy minimization is performed. In this light two methodologies, complementary to each other, are proposed for optimization of the special class of models comprising of a random field imposed on label priors. This class of segmentation models poses a special optimization problem, as the variables constituting the MRF in this case are continuous and are subject to probability constraints (positivity, sum-to-unity). The proposed methods are evaluated numerically in terms of objective function value and segmentation performance, and compare favorably to existing corresponding optimization schemes.

1. INTRODUCTION

Markov random fields have been successfully incorporated in various applications in the field of image processing, as image segmentation [1, 2], image restoration [3] or super-resolution [4]. The foremost reason for their popularity is that they can elegantly and formally model the spatial coherence trait of images.

In this paper we are interested in Markov random fields in the context of image segmentation, *i.e.* clustering a given image in J non-overlapping *meaningful* regions. A mesh of J-variate zero-one vectors is typically assumed ($\mathbf{Z}$) that serves to relate each image pixel to a unique cluster. Thus effectively, mesh $\mathbf{Z}$ conveys all information necessary about the assumed segmentation. Observed feature vectors are thence assumed to by generated independently given knowledge of the cluster they belong to. Inference of the segmentation is thence performed by estimating the posterior of $\mathbf{Z}$ conditioned on the observed image.

According to the way MRF constraints are implemented and based on the previously described general generative model correspond two distinct generative model types. The first model type involves assuming a Markov random field directly on class labels. However, inference of the posterior field distribution is typically intractable and estimation algorithms such as the computationally expensive family of the Markov chain Monte Carlo techniques [5] have to be employed. Other inference methodologies propose convenient approximations for the posterior random field, such as the pseudo-likelihood [6] or the simulated-field approximation [2]. Imposing a discrete MRF on the pixel labels is successfully used in [7], [8], [9] among others.

An alternative to avoid the computational cost of the pixel label MRF estimation is to model the *contextual mixing proportions*, that is probabilities of the pixel labels (or the mixing proportion vector for each distinct pixel), as a Markov random field [10, 11, 12, 13, 14]. In such models, MAP estimation of the contextual mixing proportions is possible, and the computational cost is transformed from a hard posterior inference problem, as in the discrete *MRF-on-labels* model family, to a difficult constrained optimization problem. In that case, the constraint is that the contextual mixing proportions corresponding to a pixel must always sum up to unity as they must be probability vectors. However, as conjectured and experimentally observed in [15], an advantage for the second model would be a less sharply peaked likelihood function.

The resulting estimated segmentation, for whichever of the aforementioned MRF schemes, relies heavily on parameter initialization and optimization [16]. There exist powerful optimization schemes suitable for discrete MRFs, as graph cuts [8] or loopy belief propagation [17]. However such schemes are not applicable in the case where the MRF is imposed implicitly on the pixel label priors, as in the second model described in the preceding paragraph. The added difficulty is that the quantities to be optimized are probabilities, which means that the search space is continuous, and secondly that each probability vector must always sum to unity.

In this paper, we propose and evelute two distinct methods on optimization of the Markov random field for segmentation probabilistic models that lie in the MRF-on-contextual

mixing proportions family. An outline of the model considered is described in section 2. The first method proposed, described in section 3, introduces a novel strategy in updating field sites, as opposed to the standard sequential raster scan of MRF sites. In section 4, we propose a method to deal with the sum-to-unity constraint posed in every site to be updated. Note that the proposed methods are complementary to each other, and may thus be used in tandem.

2. MRF FORMAL DESCRIPTION

Let $\mathbf{Z}$ be a mesh of zero-one $Jx1$ vectors, $\{z^n\}_{n=1}^N$, controlling to which of the J classes of the image, the corresponding pixel indexed n belongs to. Thus $\mathbf{Z}$ defines a segmentation on the observed image. We define the vector mesh $\mathbf{\Pi} = \{\pi^n\}_{n=1}^N$ of contextual mixing proportions according to $p(z_j^n = 1) = \pi_j^n$; in other words the prior probability of pixel n belonging to class indexed j is π_j^n.

We assume an MRF on the mesh of the class priors $\mathbf{\Pi}$. Consequently [18], $\mathbf{\Pi}$ must be Gibbs-distributed:

$$p(\Pi) \propto \prod_C e^{-\psi_c(\Pi)} \qquad (1)$$

where ψ_c is a function on clique c, called *clique potential* function in the literature, and the product is over all minimal cliques of the Markov random field.

In the present work, we shall also assume that local differences in contextual mixing proportion values are normally distributed:

$$\pi_j^n - \pi_j^k \sim \mathcal{N}(0, 1/\sigma_j^{nk}), \quad \forall n, j, \forall k \in \gamma(n)$$

where $\gamma(n)$ stands for the set of neighbouring pixels of pixel indexed n. Parameter σ_j^{nk} controls Gaussian distribution tightness; consequently as it tends to $+\infty$, contextual mixing proportion smoothness is forced and vice versa. Note that with indices j, n, k on σ_j^{nk} we imply that this smoothness factor may be dependent on class (j) as well as position on the MRF mesh (n and k). In this manner, the MRF model presented here can represent models where the smoothness parameter is class-dependent [12] or not [11], or models where an edge-preserving line process is introduced [13, 14].

In this context, segmentation involves MAP estimation of the likelihood function $p(X, \Pi; \Psi)$, with X comprising of the observed image feature vectors and Ψ including deterministic model parameters, including σ. Assuming that the posterior $p(Z|\Pi, X)$ is tractable, this problem is transformed to optimization of the expectation with regard to $Z|X, \Pi$ of the surrogate function $p(X, Z, \Pi, \Psi)$ [19]. Rewriting this latter in a more convenient manner, our objective function with regard to the MRF-governed contextual mixing parameters Π is

$$<\ln p(Z|\Pi)> + \ln p(\Pi|\sigma) + const. \qquad (2)$$

where $<\cdot>$ denotes expectation with regard to $Z|X, \Pi$.

3. OPTIMIZATION OF THE MRF MESH

A simple and straightforward implementation for optimization of (2) with regard to $\mathbf{\Pi}$ would be to perform a raster scan for each pixel $n \in [1..N]$ in order to update the sites sequentially; this involves solving J quadratic equations for each site and then projecting the resulting π_j^n vector onto the constraints $\sum_{j=1}^J \pi_j = 1$ and $\pi_j \geqslant 0$ [11]. This scheme would typically lead to a local maximum.

However, in practice, this local maximum is often far from the desirable segmentation result both quantitatively and visually (a related work with a detailed discussion on this issue is presented in [16]). This is due to the fact that the values of Π have a direct impact on the segmentation as the hidden variables Z depend on them.

In order to illustrate the importance of Π and its optimization, we have have performed segmentations on a test image [1] using the Expectation Maximization -derived algorithm in [14], by applying two different initialization schemes. At first, we have used a standard *k*-means algorithm which is common in initializing mixture models. The second approach consisted in using as initial condition the ground truth of the image. Although it is impossible to perform the latter initialization in a real segmentation scenario, we applied it in the sense of the best initialization a segmentation method could potentially attain.

A raster scan was applied to both initialization approaches in order to sequentially optimize the parameters Π for each pixel. The results in table 1 and figure 1 validate that the ground truth is indeed a local optimum for our edge-preserving algorithm. However, *k*-means initialization and standard raster scan MRF optimization lead to a solution that is optimal neither in terms of likelihood nor visually.

Let us consider now the Markov random field example in fig. 2. Each site represents a vector of *contextual mixing proportions* for a certain pixel location. Consider the white sites having mixing proportion vectors equal to $\pi^n = z^n = [0.5 + \epsilon, 0.5 - \epsilon]^T$, with $0 < \epsilon < 0.5$ and the gray sites have $\pi^n = z^n = [0.5 - \epsilon, 0.5 + \epsilon]^T$.

Observe, that each gray site is surrounded by exactly two gray and two white neighbors and that all white sites have at most one gray neighbor each. Hence, given appropriate values for σ_1, σ_2 and ϵ the gray sites may have their π parameters updated to values closer to the values of the white sites. On the contrary, this will not be the case if σ_j^{nk} are such that the MRF smoothing effect is tight enough. In that case, each individual update for the gray sites will naturally leave their weights unaffected. Therefore, if the gray sites are optimized jointly higher values for the data likelihood could be obtained. Intuitively, this can be achieved by optimizing groups of pixels with the constraint of being

Table 1. The RAND index [20] for the segmentations of the degraded versions of the 3-class artifical image of [1] along different iterations of the EM algorithm are presented. Method names followed by "Π" refer to the hypothetical segmentations computed using Π instead of the hidden variables Z to classify pixels. The average data log-likelihood at the $1000th$ iteration is also shown.

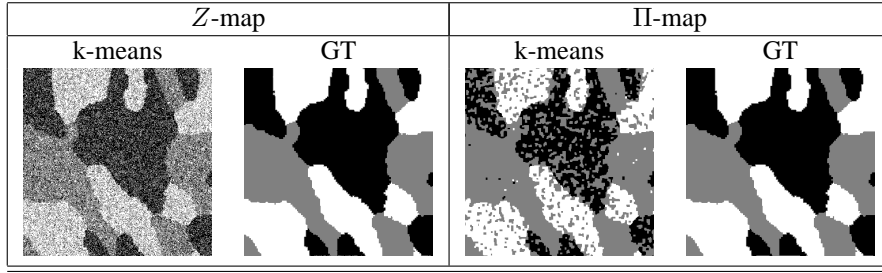| $Initialization$	2	5	10	20	200	500	1000	Av.Lhood
k-means	.70	.64	.62	.62	.62	.62	.62	59.0
k-means (Π)	.70	.73	.75	.76	.77	.77	.77	
Ground truth	.99	.99	.99	.99	.99	.99	.99	129.0
Ground truth (Π)	.99	.99	.99	.99	.99	.99	.99	

Fig. 1. Segmentation results of the 3-class synthetic image of [1] degraded by 2 dB additive white Gaussian noise after 1000 iterations. On the left are shown the segmentations computed using the labels distribution Z to classify the pixels. On the right are shown the hypothetical segmentations computed using the contextual mixing proportions Π instead of Z for classification.
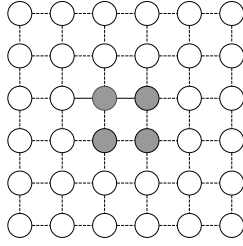

Fig. 2. An example of Markov random field of 6x6 sites. The color of each site corresponds to the image class the pixel is more likely to belong.
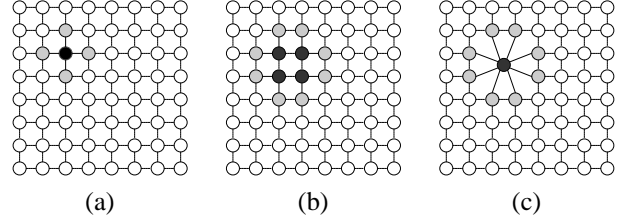

Fig. 3. Grid-scan updates on an example lattice with 8x8 elements and 1st order neighborhoods. Black color shows the elements whose contextual mixing proportions need to be updated. Gray color shows their neighboring pixels. (a) Single element to be optimized and its neighbors. (b) Elements to be co-optimized by a step of grid scan and their neighbors. (c) The same elements to be co-optimized redrawn as one.

all set to the same value. In view of this conjecture, we extend the standard raster scan procedure to a new *grid scan* strategy which is described in Algorithm 1.

The update equations in step 5 of the proposed algorithm are justified as follows. In each update step of a single grid S, we need to optimize (2):

$$<\ln p(Z|\Pi)> + \ln p(\Pi|\sigma) + const. =$$

$$\sum_{j=1}^{J} \left\{ \ln \pi_j \sum_{n \in S} (<z_j^n>) + \right.$$

$$\left. + \sum_{n \in S} \sum_{k \in \gamma(n), k \notin S} \left(-\sigma_j^{nk} (\pi_j - \pi_j^k)^2 \right) \right\} + const. \quad (4)$$

with respect to $\pi_j, \forall j \in [1..J]$. We can easily conclude that the resulting second-order equation

$$a_j^n \left(\pi_j^{n(t+1)} \right)^2 + \beta_j^n \left(\pi_j^{n(t+1)} \right) + c_j^{n(t+1)} = 0 \quad (5)$$

to be solved has coefficients given by:

$$a_j^n = - \left\{ \sum_{n \in S} \sum_{k \in \gamma(n), k \notin S} \sigma_j^{nk(t)} \right\},$$

$$\beta_j^n = \left\{ \sum_{n \in S} \sum_{k \in \gamma(n), k \notin S} \sigma_j^{nk(t)} \pi_j^{k(t)} \right\},$$

Algorithm 1 Grid scan

1 Calculate the initial grid size, $maxLevel$. This is empirically set to

$$maxLevel \leftarrow max(\llcorner log_2 max(dimX, dimY) \lrcorner - 3, 3) \tag{3}$$

2 For each $L \leftarrow maxLevel$ to 1 iterate:

3 Let $subsetLength \leftarrow 2^L$. Let G denote the set of sites, with $|G| = dimX \times dimY$.

4 Partition the $dimX \times dimY$ sites into L subsets $\{S^i\}_{i=1}^L$. Also we require $\cup_{i=1}^L S_i = G$ and $S_i \cup S_j = \varnothing, \forall i \neq j$.

5 For each site subset S_i, $i = 1, \ldots, L$, repeat steps 5.1, 5.2.

5.1 Define a set of sites $\tilde{\gamma}(S_i)$ as

$$\tilde{\gamma}(S_i) \triangleq \{\cup_{s \in S_i} \gamma(s)\} \setminus S_i$$

5.2 Optimize the sites in S_i by solving the quadratic equation (5), as if we had used raster-scan, with $<z_j^n>$ and $\gamma(n)$ being replaced by

$$<\tilde{z}_j> \leftarrow \sum_{n \in S_i} <z_j^n>$$

$$\gamma \leftarrow \tilde{\gamma}(S_i)$$

6 End.

$$c_j^n = \frac{1}{2} \sum_{n \in S} <z_j^n>^{(t)} .$$

which makes the derivation of algorithm step 5 straightforward.

To evaluate the proposed MRF optimization strategy, we computed a number of segmentations using the grid-scan versus the raster-scan optimization method. All tests were performed on noisy versions of a synthetic 3-class image [1] using the MRF model of [14]. In table 2 we present a comparison of raster-scan and grid-scan algorithms in terms of model likelihood and ratio of misclassified pixels (MCR). Likelihood scores are consistently better for grid-scan for all tested noise levels. Visual result as represented with the segmentation MCR however worsens with grid-scan optimization on low-noise levels. This is justified since as the noise level decreases, the need for smoothing decreases as well and higher probability model states may well be corresponding to undesirable smoothing in the resulting segmentation. However, this is an issue of a MRF prior in general.
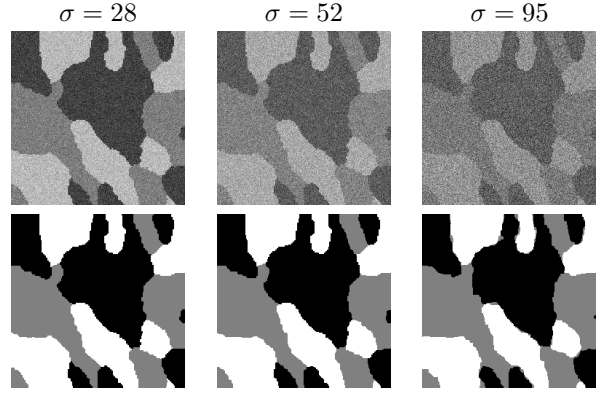

Fig. 4. Top row: A synthetic 3-class image degraded by white Gaussian noise, with varying standard deviations $\sigma = \{28, 52, 95\}$. Bottom row: Corresponding segmentations using the proposed optimization schemes.

Table 2. Comparison in terms of likelihood and misclassification ratio (MCR) for the continuous line process model, between raster-scan and grid-scan optimization methods.

	Raster-scan		Grid-scan	
σ	Av.Likelihood	MCR	Av.Likelihood	MCR
25	43.9	.1%	51.9	.13%
28	40.5	.17%	47.5	.18%
47	27.8	.5%	34.6	.5%
52	28.3	.8%	33.5	.6%
95	28.9	3.7%	31.5	3.2%

4. PROJECTION ON CONSTRAINTS HYPERPLANE

We have already discussed that we need a maximizer for (4) also satisfying the constraints:

$$\sum_{j=1}^J \pi_j^n = 1, \quad \pi_j^n >= 0, \quad \forall j \in [1..J], \quad \forall n \in [1..N].$$

In the general case, the solution of (5) does not satisfy the above constraints, that is, the computed contextual mixing proportion π_j^n, $j = 1, ..., J$ for a given pixel n are not the components of a probability vector.

It can be easily seen, that, for a particular site n, eq. (4) has the form:

$$x^T A x + x^T b + c \ln x + d \tag{6}$$

where we have denoted $[\pi_1^n \pi_2^n \cdots \pi_J^n]$ as x for convenience. Also, note that the above function is concave and the $J \times J$ matrix A is diagonal and negative definite.

An approximation of the objective function (4) is ob-

tained by dropping the term involving the logarithm:

$$\sum_{n=1}^{N} \sum_{k \in \gamma(n)} \left\{ -\sigma_j^{nk} (\pi_j^n - \pi_j^k)^2 \right\} + const. \qquad (7)$$

Let $y^\star$ be the constrained maximizer of the objective function (7), and t a point on the constraints plane other than $y^\star$. It can be shown that $y^\star$ will have to satisfy $(y^\star - \alpha)^T A(t - y^\star) = 0$ for any plane point t. This can be expressed otherwise, as looking for y such that the projection of $\alpha' \equiv A^{\frac{1}{2}} \alpha$ on the transformed plane defined by $t' \equiv A^{\frac{1}{2}} t$ will be $y' \equiv A^{\frac{1}{2}} y$. Thus, formally, we have the following quadratic programming problem to solve:

$$\arg \min_{y'} \| \alpha' - y' \|, \quad \sum_j y_j = 1, \quad y_j \geqslant 0, \quad j = 1, ..., J.$$

We now employ an active set type method as suggested in [11], allowing to derive closed form expressions for the Lagrange multipliers. The associated Lagrange function is given by:

$$L(y, \lambda_0, \lambda) =$$

$$\frac{1}{2} \sum_{j=1}^{J} (b_j y_j - b_j \alpha_j)^2 - \lambda_0 \left(\sum_{j=1}^{J} y_j - 1 \right) - \sum_{j=1}^{J} b_j^2 \lambda_j y_j$$

where λ_0 is the multiplier for the equality, and $\lambda_j, j = 1 \cdots J$ are the multipliers for the inequality constraints. Parameters b_j are the diagonal elements of the Hessian matrix A:

$$b_j = \sqrt{\sum_{k \in \gamma(n)} \sigma_j^{nk}}$$

where we have omitted the n data index from b and u for convenience. First-order necessary conditions imply:

$$y_j = \alpha_j + \frac{\lambda_0}{b_j^2} + \lambda_j \qquad (8)$$

and injecting it into the equality constraint yields:

$$\lambda_0 = \frac{1}{\sum_j b_j^{-2}} - \frac{\sum_j \alpha_j}{\sum_j b_j^{-2}} - \frac{\sum_j \lambda_j}{\sum_j b_j^{-2}} \qquad (9)$$

Finally, by combining (8) and (9) we obtain:

$$y_j = \alpha_j - c_j + c_j \sum_{l=1}^{J} \alpha_l + c_j \sum_{l=1}^{J} \lambda_l + \lambda_j \qquad (10)$$

where $c_j \equiv -\frac{b_j^{-2}}{\sum_{l=1}^{J} b_l^{-2}}$.

Let us notice that the vector $\alpha_j - c_j + c_j \sum_{l=1}^{J} \alpha_l$ is the projection of α on the constraints hyperplane $\sum_{j=1}^{J} y_j = 1$. The set of Lagrange multipliers $\lambda_j, j = 1, ..., J$ must satisfy the inequality constraints. Karush-Kuhn-Tucker conditions
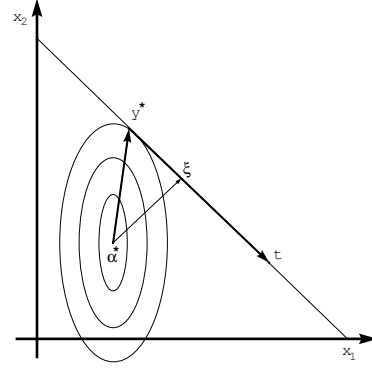

Fig. 5. Example projection to the constraints plane, in the two-dimensional case $J = 2$. Ellipses represent contours of the quadratic approximation to the objective function; the line joining the x_1 and x_2 axes is the linear constraints plane, here $x_1 + x_2 = 1, x_1, x_2 \geqslant 0$. The unconstrained maximizer is $a^\star$, the constrained maximizer is $y^\star$ and t is a point on the constraints plane. Point ξ shows the location of the solution proposed in [11].

[21] state that at the minimizer $y^\star$ we must have $\lambda_j \geqslant 0$ and $\lambda_j > 0$ if $y_j^\star = 0$ which is the active constraint.

Contrary to the projection with [11], we can point out that in our proposal we have constructed our reasoning based on the sole hypothesis that the logarithm in eq. (6) is a negligible quantity with respect to the other terms; this provided, our method will necessarily give the correct constrained optimum. Note also that the projection in [11] could be seen as a subcase of our own proposal, for $b_1 = b_2 = \cdots b_J$.

To evaluate the proposed algorithm we have compared it to the projection algorithm in [11]. We have segmented the color *Church* image (Berkeley database [22], #118035) using the segmentation model proposed in [14]. The resulting comparison revealed that the new algorithm provides consistently higher values for the data likelihood (fig. 6).

5. CONCLUSION

While MRF-driven image analysis and in particular image segmentation can yield excellent results, it is always very dependent on the manner the field itself is optimized with regard to the likelihood function. In view of this fact, we have proposed a strategy for optimizing an MRF mesh, as well as a novel optimizing projection method, applicable in cases where the Gibbs clique potentials are Gaussian distributed. Numerical results show that our proposals work favorably compared to approaches used for similar-structured MRF comprising segmentation models.

Algorithm 2 Projection on constraints hyperplane

1 Let y denote the vector at the current iteration. Initially, we set $y_j \leftarrow b_j, \forall j = 1, 2, ..., J$. In the general case, there exist m negative components y_j. The corresponding set of indices $S = \{j, \text{ with } y_j < 0\}$ constitutes the active set of constraints for the current vector y.

2 $\forall j \notin S$, set $\lambda_j \leftarrow 0$.

3 $\forall j \in S$, set $y_j = y_j^\star \leftarrow 0$ and we compute the corresponding λ_j by solving an $m \times m$ linear system that forces the inequalities to be satisfied as equalities, namely $y_j + \lambda_j + c_j \sum_{l=1}^{J} \lambda_l = 0$, written in matrix form as $(I + \mathbf{1}c^T)\lambda = y$. The Sherman-Morisson formula [21] gives:

$$\lambda_j \leftarrow y_j + \frac{\sum_{l \in S} c_l y_l}{\sum_{l \notin S} c_l}$$

4 Compute the updated y_j values for $j \notin S$ by (10), using the new vector λ.
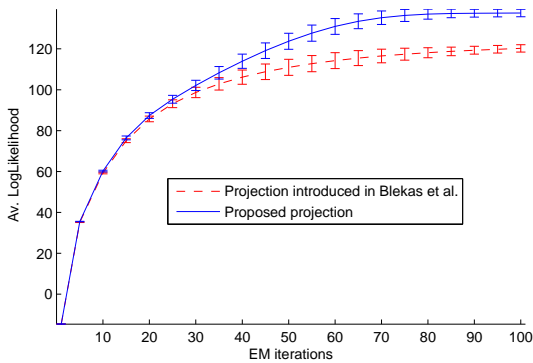
5 Return to step 2 until convergence.

Fig. 6. Comparison of data likelihood values for the projection method in [11] and the projection algorithm proposed in this paper: The test image ([22], see main text) was segmented into three classes using the model of [14]. The solid curve shows our results using the proposed projection against the results using the projection proposed in [11], shown by the dashed curve. For each configuration, we ran the segmentation 10 times using k-means initialization perturbed by additive white Gaussian noise of 0.2 units standard deviation. Likelihood values (averaged over number of pixels N and over the 10 different initializations) are shown for the first 100 EM iterations.

6. REFERENCES

[1] 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 Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.

[2] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for Markov model-based image segmentation," *Pattern Recognition*, vol. 36, pp. 131–144, 2003.

[3] R. Molina, J. Mateos, A.K. Katsaggelos, and M. Vega, "Bayesian multichannel image restoration using compound Gauss-Markov random fields," *IEEE Transactions on Image Processing*, vol. 12, pp. 1642–1654, 2003.

[4] G. Chantas, N.P.Galatsanos, and N.Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1821–1830, 2007.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[6] J. Besag, "Statistical analysis of non-lattice data," *Statistician*, vol. 24, pp. 179–195, 1975.

[7] J. Marroquin, E. Arce, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1380–1387, 2003.

[8] R. Zabih and V. Kolmogorov, "Spatially coherent clustering using graph cuts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 2004, vol. 2, pp. 437–444.

[9] W. Qian and D.M. Titterington, "Estimation of parameters in hidden Markov models," *Philosophical Transactions of the Royal Society: Physical and Engineering Sciences (1990-1995)*, vol. 337, no. 1647, pp. 407–428, 1991.

[10] S. Sanjay-Gopal and T. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[11] K. Blekas, A. Likas, N. Galatsanos, and I. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 494–498, 2005.

[12] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1121–1130, 2007.

[13] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, "MR brain segmentation using an edge-preserving spatially variant Bayesian mixture model," in *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, New York, USA, 2008.

[14] G. Sfikas, C. Nikou, and N. Galatsanos, "Edge preserving spatially varying mixtures for image segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Alaska, USA, 2008.

[15] A. Diplaros, N. Vlassis, and T. Gevers, "A spatially constrained generative model and an EM algorithm for image segmentation," *Neural Networks, IEEE Transactions on*, vol. 18, no. 3, pp. 798–808, May 2007.

[16] R.Szeliski, R.Zabih, D.Scharstein, O.Veksler, V.Kolmogorov, A.Agarwala, M.Tappen, and C.Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, 2008.

[17] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Generalized belief propagation," *Advances in Neural Information Processing Systems (NIPS)*, pp. 689–695, 2000.

[18] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 721–741, 1984.

[19] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, November 2008.

[20] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.

[21] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer-Verlag, 1999.

[22] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of the 8th International Conference on Computer Vision (ICCV '01)*, July 2001, vol. 2, pp. 416–423.