Unsupervised Image Segmentation: A Bayesian Approach

Li Cheng

Dept. of Computing Science
Univ. of Alberta
Edmonton, A.B. T6G 2E8

Terry Caelli *

Dept. of Computing Science
Univ. of Alberta
Edmonton, A.B. T6G 2E8

Abstract

In this paper we explore a new two-stage approach to unsupervised image segmentation. First we cluster pixels in the image colour histogram space with a Minimum Discription Length parametric clustering method to approximate hypothesized densities. Second, we utilize a variant of a MRF model called EM-HMRF as well as loopy belief propagation to update the region densities, and so, segmentation. The system is autonomous with the number of clusters being estimated at run time. Results are promising with both coloured and grey-scale images.

Keywords: Unsupervised Image Segmentation, MArkov Random Fields, Bayesian inference, Expectation-Maximization.

1 Introduction

Image Segmentation is difficult to solve partly due to it's ill-proposed nature and the variations that occur over many different types of images. Much research effort ([1], [2], [3], [10]) have been devoted to this issue partly due to its practical uses in such tasks as image retrieval. Consequently, in this paper we restrict interest to classes of images (grey-level or coloured) that are (nearly) piecewise continuous as considered by Geman and Geman [6].

Three key problems arise in unsupervised image segmentation:

- how to extract features from image data to build the appropriate feature space;
- how to estimate the clusters(modes) from the feature space of image data;
- given the estimated clusters, how to perform spatial grouping to decide which pixel belong to which cluster.

*This project was funded by a grant from the Natural Science and Engineering Research Council of Canada.

Our approach to these three problems is to use Bayesian parameter estimation methods of current interest in the computer vision literature, in general. With textures, for example, Comaniciu et. al. [3] have interleaved nonparametric (Mean Shift) methods to model the feature space using sliding windows and outlier detection iteratively till convergence, clusters are then constructed in the 'cleaned' data; Carson et. al. [2]'s 'blob world' approach integrates human's prior knowledge as multiple cues and then combines those cues to form the feature space. The EM (Expectation Maximization) algorithm is employed in the feature space to estimate MDL-based (Minimum Description Length) clustering.

Our approach addresses image segmentation task in a different way. Here, the feature space is formed by a nonlinear transformation of the image raw data. An EM-type mixture of Gaussian (MoG) algorithm is then run in the feature space to extract sufficient statistics for the image data, with the estimated number of clusters based on the MDL criteria. Similar to Zhang et. al [10]'s method, the image domain is modeled as a HMRF (Hidden Markov Random Field) and the EM method is utilized to iteratively estimate the parameters of clusters and infer the hidden cluster label for each pixel.

The paper is organized as follows: First, the MDL clustering algorithm is introduced in Sec. 2. Second, the image model - the EM-HMRF (Expectation Maximization - Hidden Markov Random Field) model is derived in detail in Sec. 3. Third, the LBP (Loopy Belief Propagation) algorithm is introduced as the inference engine for the MRF layer in Sec. A. Finally, experimental results are shown in Sec. 5 with conclusions in Sec. 6.

2 MDL-based clustering algorithm

Assuming that the image feature space can be effectively approximated by a mixture of K Gaussians, the first problem is how to estimate the Gaussian parameters $\theta = \{\theta_k = (\pi_k, \mu_k, \Sigma_k); k = 1, \ldots, K\}$ with priors, π_k, means, μ_k, and variance-covariance matrices, Σ_k. To achieve this, the

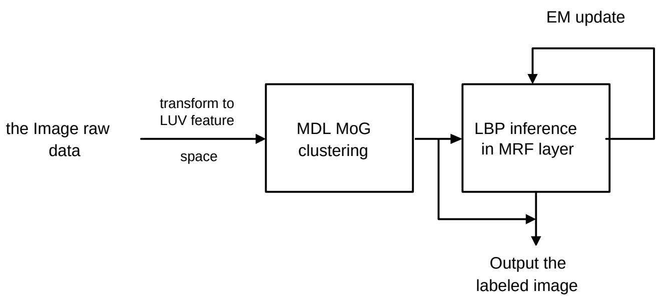

Figure 1: Flow chart illustration of the proposed two-stage (clustering and inference) framework.

EM algorithm is typically used to determine the optimal parameters of a mixture of K Gaussian in the maximum likelihood sense [4]. In this approach, however, what is missing is the derivation of an optimal inference procedure for pixel clustering, even given optimal feature space clustering.

For clustering, we can compute the complete data likelihood as $p(X, Y|\theta)$ where X denotes the evidence for image pixels, Y is the set of observed image data, where θ refers to the set of Gaussian parameters that define an objective function Q which corresponds to the expected complete log likelihood (Fig. 2):

$$Q(q, \theta) = E_{q(X|Y,\theta)}[\ln(p(X, Y|\theta))].$$

The EM algorithm iterates through computing the E step:

$$q(X|Y, \theta) = \arg\max_{q^*} Q(q, \theta)$$

and the M step:

$$\theta = \arg\max_{\theta^*} Q(q, \theta)$$

till converge to local optimal (see [4] for more details of the EM algorithm).

The problem of determining K, the number of mixture components, is a model selection problem [7]. Ideally K is chosen to best fit the 'natural' number of clusters present in the feature space by using the EM algorithm over different numbers of clusters. However, Maximum Likelihood estimators, to which the EM algorithm belongs to, tend to overfit the data converging on the most complex model. So, to regularize the ML estimator, one can use model selection criteria, such as AIC (Akaike Information Criteria), BIC (Bayesian Information Criteria), etc [7]. The objective we adopted here is to minimize the MDL criteria given by [1]:

$$MDL(K, \theta) = -\sum_{n=1}^{N} \log\left(\sum_{k=1}^{K} p_{y_n|x_n}(y_n|k, \theta)\pi_k\right) + \frac{1}{2} S \log(NM) \quad (1)$$

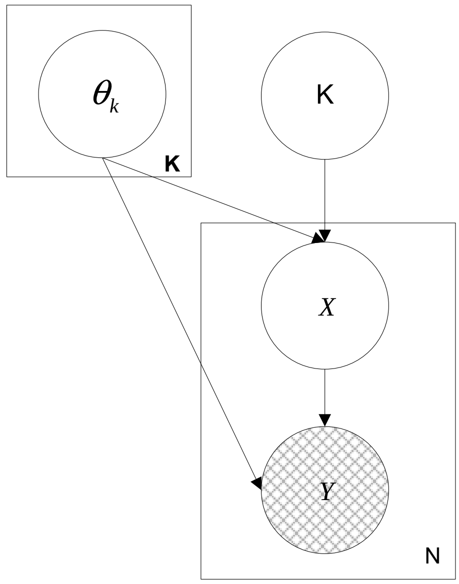

Figure 2: Graphical model for mixture of Gaussian clusters. Here the circles denote variables and boxes for those which are iid's (independent identical distributions). The shaded circle corresponds to the observed variables and unshaded for the hidden variables. K is the number of Gaussians, N is the number of data points; x is the cluster label, y is the observed data vector. θ_k represents the Gaussian parameters for a given cluster.

and where

$$S = K\left(1 + M + \frac{(M+1)M}{2}\right) - 1.$$

Notice that from a Bayesian viewpoint, Eq.(1) is composed of two parts, a likelihood part and a prior part. For computational simplicity we choose to maximize the augmented Q function:

$$Q(K, \theta) \propto MDL(K, \theta)$$

where:

$$Q(K, \theta) = \sum_{k=1}^{K} \bar{T}_k(-\frac{1}{2} trace[\bar{\Sigma}_k \Sigma_k^{-1}] - \frac{1}{2}(\bar{\mu}_k - \mu_k)^t \Sigma_k^{-1}(\bar{\mu}_k - \mu_k) - \frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_k|) + \log(\pi_k)) - \frac{1}{2} L \log(NM).$$

$\bar{T}_k$, $\bar{\mu}_k$, and $\bar{\Sigma}_k$ are iteratively updated using the following EM algorithm:

1. E step:

$$\bar{T}_k = \sum_{n=1}^{N} p_{x_n|y_n}(k|y_n, K, \theta)$$

2. M step:

$$\bar{\pi}_k = \frac{\bar{T}_k}{N} \quad (2)$$

$$\bar{\mu}_k = \frac{1}{\bar{T}_k} \sum_{n=1}^{N} y_n p_{x_n|y_n}(k|y_n, K, \theta) \quad (3)$$

$$\bar{\Sigma}_k = \frac{1}{\bar{T}_k} \sum_{n=1}^{N} (y_n - \bar{\mu}_k)(y_n - \bar{\mu}_k)^t p_{x_n|y_n}(k|y_n, K, \theta). \quad (4)$$

As a result of using this MDL principle, when several models using different values of K fit the data equally well, the algorithm chooses the simplest method. In our experiment, we have set the range of K to between 1 and 6, as from experimentation, this resulted in the best performance.

3 EM-HMRF image modelling

During the first stage of the proposed framework, MDL-based mixture of Gaussian clustering was performed in the feature space. In the second inference stage, following Zhang et. al.'s [10] approach, a Hidden Markov Random Field model is utilized to incorporate pixel-wise Markovian interactions on the two dimensional image lattice, as illustrated in Fig.3 (Note the close resemblance of the model in this second stage to that of the first stage). In this context, the clustering labels x and the Gaussian parameters are unknown variables.

As a consequence, we, again, resort to the EM algorithm for parameter estimation. The EM algorithm is very similar to that discussed in the previous section, with the only difference being that, instead of computing the expected complete log likelihood in the E step directly, we approximate it via loopy belief propagation, in the HMRF model. In other words, consider we have observed a random field, denote as Y, with each node defined by y_i. There is also a HMRF X, where each node, x_i, 'causes' one unique node y_i, and via LBP (Loopy belief propagation, see Appendix A for details), we can approximate the MAP of the posterior distribution of $p(X|Y, \theta)$. Meanwhile the M step is exactly the same as Eq.2, 3 and 4, and of course the number of clusters K is now a fixed value.

Currently we model the pixel-wise interaction in the MRF simply using Potts model [4], that is:

$$p_\beta(X|Y, \theta) = \frac{1}{z(\beta)} \exp\left\{ -\sum_{i \sim j} \beta\delta(x_i \neq x_j) \right\}$$

where i and j are two neighbouring sites in the Markov Random Field X; $\delta(\cdot) = 1$ if $(\cdot) = 1$ and 0 otherwise; β is a positive value in the Potts model used to encode the strength of the interaction between neighbouring sites in X; $z(\beta)$ is the normalizing constant. Note that the β value is obviously adaptive to the input image. For proof of concept here we handcrafted this value with different images, but plan in the future to let the model autmatically adapt the β value to current images by Markov Chain Monte Carlo method as, for example, described in [8].

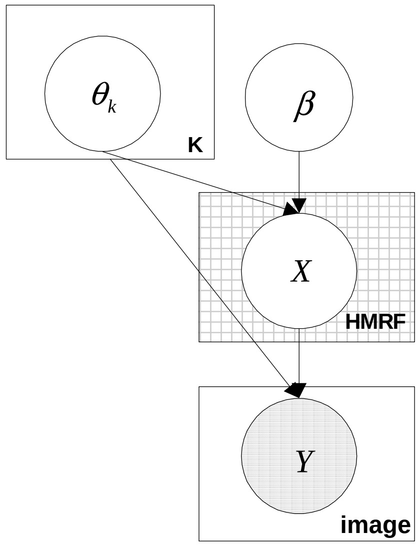

Figure 3: The EM-HMRF segmentation model. Here the circles denote variables and boxes for iid's (independent identical distributions), the grid-box is for the HMRF. The shaded circle is for observed variables and unshaded for the hidden variables; K is the number of Gaussians, X contains, for each site, one segmentation label, Y contains the corresponding transformed pixel values; β denote the parameter for the Potts model; θ_k represents the Gaussian parameters for given cluster, respectively.

4 The segmentation framework

Image raw data are typically stored in RGB space. To make the Eclidean distance calculation less sensitive to the local manifolds of colour space, the feature space is formed by nonlinear transformation of the image raw data into the *LUV* [1] space. The transformed image data is then processed using the following two-stage approach.

AS already mentioned, the first part, the MDL-based clustering algorithm is performed on the feature space to characterize the densities as mixture of Gaussians, in the maximum likelihood sense. The image data is initially clustered in the feature space using this model accordingly to form the initial segmented image X; In the second part, the image pixel segmentation labels are updated iteratively,

[1] We chose to use the LUV colour space due to its property of perpetual uniformity: two colours are equally distant in the colour space. See [5] for more details about the CIE LUV colour space.

based on the EM-HMRF model, via the LBP inference in the MRF layer. The algorithms stops when the EM-HMRF algorithm converges to an equilibrium state. The basic steps are the follows:

1. Nonlinear transformation of the image data from the RGB space into the LUV space.
2. let $t = 0$. In the feature space, the MDL clustering algorithm is employed for density estimation (obtaining the parameters of the Gaussians θ_0) and initial image segmentation, X_0.
3. let $t = t + 1$. In the image domain
   (a) based on current image segmentation X_t, run the EM algorithm to update the parameters of the Gaussians, θ_{t+1}.
   (b) Based on the parameters θ_{t+1} and the image segmentation X_t, the LBP algorithm is employed for approximate inference of the new image segmentation X_{t+1}.
4. repeat step 3 till convergence.

5 Experimental results

We have tested our algorithm on a variety of images. In this section we show results on four images: *hand*, *table tennis*, *pentagon* and *forest*, as shown in Figs. 4(a), 6(a), 8(a) and 10(a), respectively. The *hand* results are shown in Fig. 4. In Fig. 4(b), after the MDL clustering stage, the estimated model captures the image data with significant noise due to the textured background. The segmentation of the first EM-HMRF iteration is quite similar to Fig. 4(b). However, as the EM-HMRF algorithm iterates, more noise elements are smoothed out as seen in Fig. 4(c) - finally reaching the equilibrium state in Fig. 4(d). The final result was determined from 3 Gaussian clusters, with parameters shown in Fig. 5. The first iteration results of the *table tennis* image, Fig. 6(b) are quite similar to that of the *hand*, again, due to the textured regions. Notice that in further EM-HMRF iterations the segmentation results are quite stable with only minor changes (see Fig. 6(c),(d)). In this case four clusters were derived with parameters shown in Fig. 7. Fig. 8 shows results for the *pentagon* image. The first iteration resulted in slightly incorrect color positions (see Fig. 8(a)). With iterations, the image statistics were adjusted gradually to the optimal values (Fig. 8(c),(d)) based on 6 clusters with parameters shown in Fig. 9. Fig. 10 contains results for the *forest* image. Here significant noise remained in the first iteration (Fig. 10(b)) largely due to the wide dispersion of colours in the colour space. This noise is eliminated quickly at the second stage of the EM-HMRF; and then the changes among iterations reach equilibrium similar to the *table tennis* image (Fig. 10(c)) with 5 estimated clusters and the parameters are shown in Fig. 11.

¿From those simulations we have made three observations. First, the second inference stage overcomes the noise-sensitive nature of the first stage (the MDL clustering model) by the EM-HMRF spatial model. Second, this framework can also be applied to images only with 'weak' textural features (see Fig. 4(a), Fig. 6(a)) - consistent with the assumed piecewise constant image model. Third, the chance of success for this framework heavily depends on the expressiveness of the feature space and also, due to the large hypothesis space for θ and the possible multi-modal nature for the posterior distribution of θ, we need to be careful that the estimated θ_0 of the first stage has to be as close to the optimal (true) θ as possible so as to avoid only local optimal solutions. Currently we have used multiple initial values to avoid this situation.

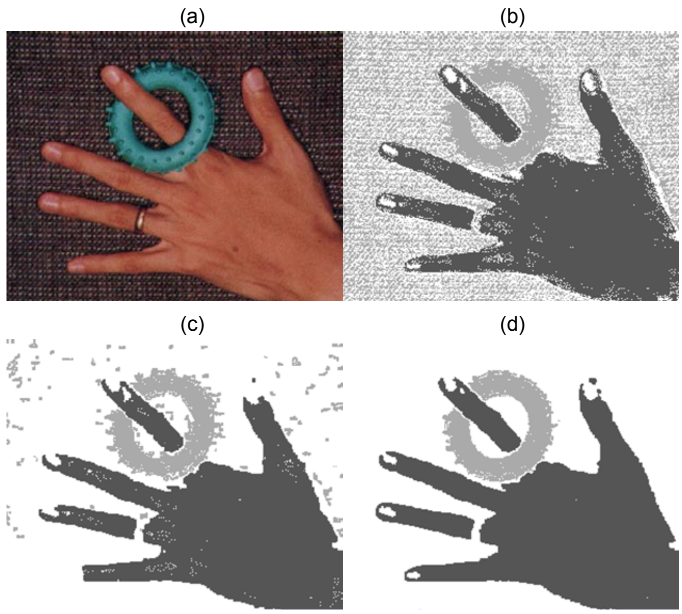

Figure 4: (a) The *hand* image, 303x243 pixels colour image. (b) the result after the first stage (classification via Mixture of Gaussians densities). (c) the result of the second EM-HMRF iteration; (d) the result after the 10th EM-HMRF iteration. $\beta = 5$ is used in this image.

	Cluster 1	Cluster 2	Cluster 3
Prior	0.3613	0.0620	0.5767
Means	70.1035	66.0011	48.4535
	51.3934	0.0000	16.2684
	22.9182	0.0711	5.3237
Variance	12.3045 -16.4833 2.3298	78.1949 -0.0001 -0.0367	121.8399 56.9382 39.1876
	-16.4833 47.9287 -0.2006	-0.0001 0.0001 -0.0000	56.9382 249.1410 -0.5515
	2.3298 -0.2006 17.1457	-0.0367 -0.0000 0.2072	39.1876 -0.5515 49.1891

Figure 5: The mixture of Gaussians parameters for *hand* image.

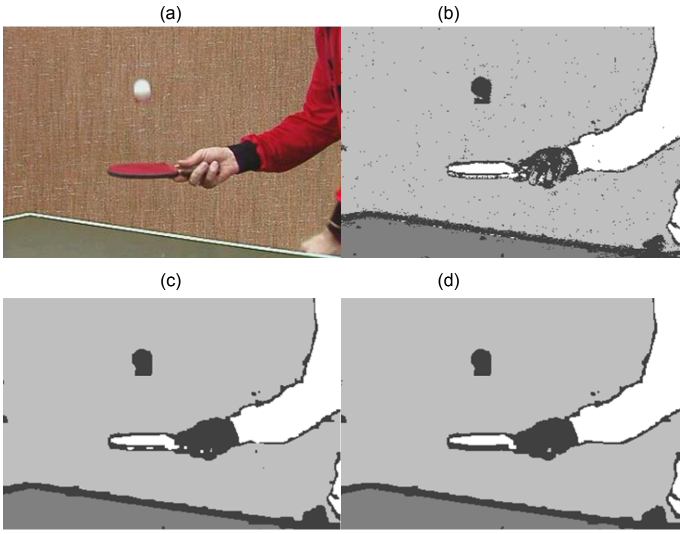

Figure 6: (a) the *tabletennis* 352x240 pixels colour image. (b) the result after the first stage (classification via Mixture of Gaussian densities). (c) the result of the first EM-HMRF iteration; (d) the result after the 4th EM-HMRF iteration. β is set to 3.5 here.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Prior	0.0858	0.7289	0.0939	0.0914
Means	72.6170 3.3274 13.5868	75.5340 20.9906 18.7001	54.7385 85.6684 11.8647	74.6857 17.4750 12.6804
Variance (1e+3)	0.0014 -0.0007 -0.0013 -0.0007 0.0049 0.0020 -0.0013 0.0020 0.0072	0.0140 -0.0071 -0.0048 -0.0071 0.0103 0.0037 -0.0048 0.0037 0.0063	0.0516 0.1434 0.0014 0.1434 1.1676 0.1128 0.0014 0.1128 0.0391	0.2317 -0.0987 -0.0754 -0.0987 0.2699 0.0447 -0.0754 0.0447 0.0567

Figure 7: The mixture of Gaussians parameters for the *pingpang* image.

6 Conclusions and future work

A new two-stage framework is proposed for unsupervised image segmentation which employs the MDL clustering algorithm for initial density estimation and image segmentation, and then the EM-HMRF model for inferring, enforcing spatial constraint, via the LBP algorithm which infers the optimal segmentation results from the Markov Random Field. Promising results are obtained on some well-known images. To continue this line of work, instead of the simple model which focuses on assumed piecewise constant images, kernel methods (such as Fourier/wavelet-based convex kernels) could be used to model the more complicated textural components; meanwhile, instead of handcraft a value for specific image, the parameter β can be estimated with the bayesian sampling techniques [8].

A Inference using Loopy Belief Propagation (LBP)

In this section we describe how to apply Loopy Belief Propagation (LBP) approximation in this setting. First we derive the full conditional joint distribution over all MRF sites $x_i \in X$, and assume *iid* distributions amongst nodes y_i:

$$\begin{aligned} p(X|Y,\theta) &\propto p(Y|X)\, p(X|\theta) \\ &= \prod_i p((y_i|x_i) \prod_{\partial i} p(x_i|x_{\partial i},\theta) \end{aligned}$$

where θ refers to the aforementioned Mixture of Gaussians parameters.

The belief for MRF X at node i is:

$$b(x_i) \propto p(y_i|x_i) \prod_{j\in\partial i} m_i^j(x_i) \tag{5}$$

where the message update rules are:

$$m_i^j(x_i) \propto \sum_{x_j} p(x_i|x_j) p(y_j|x_j) \prod_{k\in\partial j\backslash i} m_j^k(x_j).$$

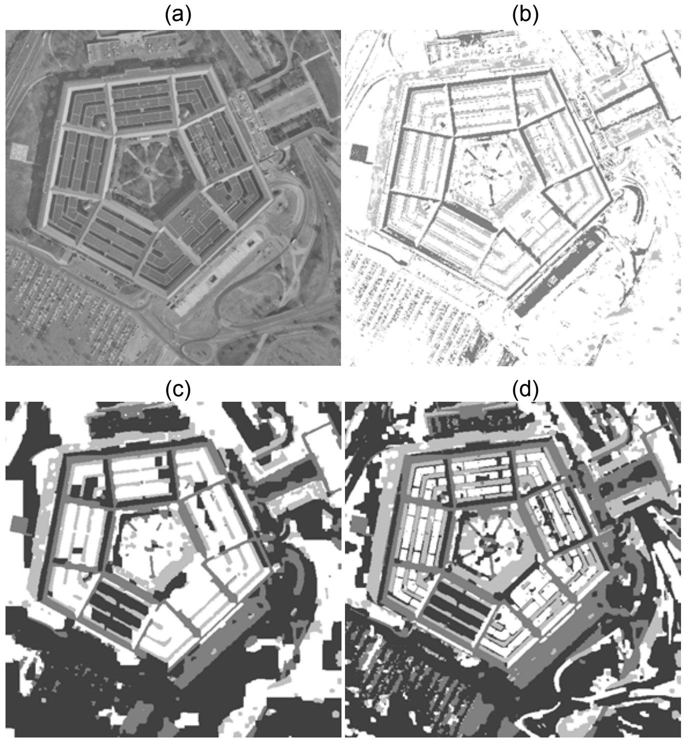

Figure 8: (a)initial clustering and segmentation of the *pentagon* image; (b) the result after first EM-HMRF iteration; (c) result after 9 EM-HMRF iterations; (d) the final result converges after 18 EM-HMRF iterations: $\beta = 3.5$.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Prior	0.2168	0.1139	0.0647	0.1886	0.2519	0.1641
Means	121.0634	176.9440	142.2758	150.2287	135.6298	101.2054
Variance	60.4923	299.9312	722.9201	106.5095	64.7907	122.6017

Figure 9: The mixture of Gaussians parameters for *pentagon* image.

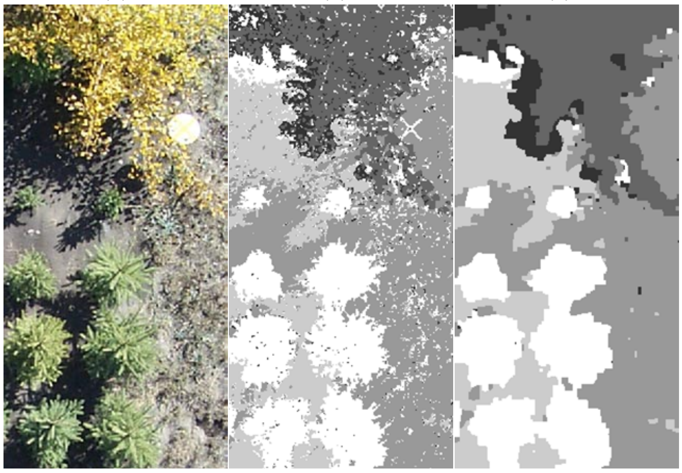

Figure 10: (a) the *forest* 203x420 pixel colour image. (b) the result after the first stage (classification via Mixture of Gaussians densities). (c) the result after the 4th EM-HMRF iteration. $\beta = 6$ in this image.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Prior	0.0495	0.1912	0.3401	0.1809	0.2382
Means	58.9406 12.6989 23.5591	84.2029 14.1137 35.8005	82.0475 2.1750 5.4147	46.8941 0.5604 0.6711	74.1190 0.1444 18.4276
Variance (diagnal)	321.1871 131.1732 167.6395	119.6599 60.6052 205.3584	89.6460 4.8920 13.7159	131.7911 13.4334 7.2348	245.9378 3.6627 60.3040

Figure 11: The mixture of Gaussians parameters for *forest* image. Due to the space limit, only the diagonal elements of the variance-covariance matrix are shown.

After computing the belief at site i, we have:

$$
\begin{aligned}
b(x_i) &\propto \sum_{1,..,i-1,i+1,..,n} \prod_i p(y_i|x_i) \prod_{\partial i} p(x_i|x_{\partial i}) \\
&\propto p(x_i|y_i,\theta).
\end{aligned}
$$

This implies that, in this MRF network, the single belief $b(x_i)$ approximates the marginal probability $p(x_i|y_i,\theta)$; and in networks without loops, the beliefs are equal to the exact marginal probabilities [9].

It is simple to show that the joint probabilities over the hidden MRF X is just the product of beliefs over X, defied by:

$$
\begin{aligned}
p(X|Y,\theta) &\propto \prod_i p(y_i|x_i) \prod_{j\in\partial i} m_i^j(x_i) \\
&\propto \prod_i b(x_i).
\end{aligned}
$$

References

[1] C. Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, March 1994.

[2] Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*. Springer, 1999.

[3] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. In *IEEE Conf. on Comp. Vis. and Pattern Recognition*, pages 750–755, 1997.

[4] A. P. Dempster, N. M. Laird, and D.B.Rubin. Maximum likelihood from incompletedata via the em algorithm. *Journal of the Royal Statistical Society, B*, 39:1–38, 1977.

[5] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, 2 edition, 1996.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721 – 724, 1984.

[7] Mark H. Hansen and Bin Yu. Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774, 2001.

[8] Descombes X., Morris R. D., Zerubia J., and Berthod M. Estimation of markov random field prior parameters using markov chain monte carlo maximum likehood. *IEEE Trans. on Image Processing*, 8(7):954–963, 1999.

[9] J. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. *International Joint Conferences on Artificial Intelligence*, 20, 2001.

[10] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation maximization algorithm. *IEEE Trans. Medical Imaging*, 20(1), 2001.