Segmentation of colour images using variational expectation-maximization algorithm

Nikolaos Nasios and Adrian G. Bors

Dept. of Computer Science, University of York,York YO10 5DD, UK
E-mail: {nn,adrian.bors}@cs.york.ac.uk

Abstract

The approach proposed in this paper takes into account the uncertainty in colour modelling by employing variational Bayesian estimation. Mixtures of Gaussians are considered for modelling colour images. Distributions of parameters characterising colour regions are inferred from data statistics. The Variational Expectation-Maximization (VEM) algorithm is used for estimating the hyperparameters corresponding to distributions of parameters. A maximum *a posteriori* approach employing a dual expectation-maximization (EM) algorithm is considered for the hyperparameter initialisation of the VEM algorithm. In the first stage, the EM algorithm is applied on the given colour image, while the second EM algorithm is used on distributions of parameters resulted from several runs of the first stage EM. The VEM algorithm is used for segmenting several colour images.

1 Introduction

Most algorithms used to solve various computer vision and image processing problems do not provide an exact solution. Three main statistical approaches have been used in computer vision: maximum likelihood, maximum *a posteriori* (MAP) and Bayesian inference. One of the algorithms that provide a good approximation for the maximum likelihood is the expectation-maximization (EM) algorithm [5, 6]. EM algorithm has been recently used for colour [7, 16] and multidimensional [11] image segmentation.

In a Bayesian approach each parameter is modelled by a probability density function and the solution is provided by integrating over the distributions of parameters. Bayesian approaches do not suffer from overfitting and have very good generalisation capabilities [8, 10]. Prior knowledge can be easily incorporated and uncertainty is manipulated in a consistent manner. However, computations in the Bayesian framework can seldom be performed exactly due to the need to integrate over distributions of models. The most known Bayesian approaches are Markov Chains Monte Carlo (MCMC), Laplace approximations and variational inference. In variational training the complex inferring problem is split in a set of simpler calculations, characterised by decoupling the degrees of freedom in the original problem. Variational Bayes (VB) algorithm has been proposed for estimating the set of hyperparameters that characterises distributions of parameters for various graphical models, as shown in [1, 2, 9, 10, 12, 15].

The graphical model used in this study consists of a mixture of Gaussians [1, 6, 15, 16]. We employ a maximum log-likelihood estimation procedure for the hyperparameter

BMVC 2004 doi:10.5244/C.18.73

initialisation. A dual EM algorithm is considered for approximating the maximum log-likelihood estimates. Parameters estimated from successive runs of a first stage EM are used as inputs for the second stage EM. After this proper initialisation, the proposed variational expectation-maximization (VEM) algorithm is expected to converge fast.

Image segmentation is the first processing stage in many computer vision systems. In colour images the uncertainty is caused by noise, reflectivity properties, textures, and other influences [13]. A survey of various methods used for colour image segmentation is provided in [3]. For segmenting colour images, Comaniciu used a kernel based algorithm in the context of mean-shift data analysis [4]. A variant of the EM algorithm has been used for 3D segmentation of Magnetic Resonance Images (MRI) of the brain [11]. The EM algorithm was used for colour image segmentation in association with Markov Random Fields in [7], and by employing a split-and-merge approach in [16]. In this paper we use a deterministic Bayesian approach such as the VEM algorithm. The number of mixture components is decided by using the Bayesian Information Criterion (BIC), which corresponds to the Minimum Description Length (MDL) [14].

The paper is organised as follows. Section 2 introduces the variational Bayesian methodology, Section 3 outlines the maximum log-likelihood estimation algorithm for initialising the VEM algorithm, while Section 4 describes the Variational Bayes algorithm. Section 5 presents experimental results when applying the proposed algorithm for colour image segmentation, and Section 6 provides the conclusions of the present study.

2 Variational Bayes Methodology

Mixtures of Gaussians have been used in many applications due to their excellent approximation properties. In the case of colour images, after the conversion to an appropriate colour space, we can model the probability of each pixel by :

$$p(\mathbf{x}) = \sum_{i=1}^{N} \frac{\alpha_i}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp[D(\mathbf{x};\mu_i,\Sigma_i)] \tag{1}$$

where d is the dimension, $\theta_i = \{\alpha_i, \Sigma_i, \mu_i\}$ represents a set of parameters, N represents the number of components, while :

$$D(\mathbf{x};\mu_i,\Sigma_i) = -\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1} (\mathbf{x}-\mu_i). \tag{2}$$

Furthermore, we consider that the sum of mixture probabilities is normalised, $\sum_{i=1}^{N} \alpha_i = 1$.

The parameters modelling a probability density function have their probabilities modelled as the conjugate priors [2, 9]. In the case of a mixture of Gaussians (1) we have the following parameters: means, covariances and mixing probabilities. The conjugate prior for means is a Gaussian distribution $\mathcal{N}(\mu|\mathbf{m},\beta\mathbf{S})$, where β is a scaling factor :

$$\mathcal{N}(\mu|\mathbf{m},\beta\mathbf{S}) \sim \frac{1}{\sqrt{(2\pi)^d |\beta\mathbf{S}|}} \exp[D(\mu;\mathbf{m},\beta\mathbf{S})] \tag{3}$$

A Wishart distribution $\mathcal{W}(\Sigma|\nu,\mathbf{S})$ is the conjugate prior for the inverse covariance matrix, where ν are the degrees of freedom :

$$\mathcal{W}(\Sigma|\nu,\mathbf{S}) \sim \frac{|\mathbf{S}|^{-\nu/2} |\Sigma|^{(\nu-d-1)/2}}{2^{\nu d/2} \pi^{d(d-1)/4} \prod_{k=1}^{d} \Gamma\left(\frac{\nu+1-k}{2}\right)} \exp\left[-\frac{Tr(\mathbf{S}^{-1}\Sigma)}{2}\right] \tag{4}$$

where $Tr(\cdot)$ denotes the trace of the resulting matrix (the sum of the diagonal elements) and $\Gamma(\cdot)$ represents the Gamma function :

$$\Gamma(x) = \int_0^\infty \tau^{x-1} \exp(-\tau) d\tau. \tag{5}$$

For the mixture probabilities we consider a Dirichlet distribution $\mathscr{D}(\alpha|\lambda_1, \dots, \lambda_N)$:

$$\mathscr{D}(\alpha|\lambda_1, \dots, \lambda_N) = \frac{\Gamma(\sum_{j=1}^N \lambda_j)}{\prod_{j=1}^N \Gamma(\lambda_j)} \prod_{i=1}^N \alpha_i^{\lambda_i - 1}. \tag{6}$$

The variational learning is expected to provide better data modelling and generalisation by taking into account the uncertainty in the parameter estimation.

3 Hyperparameter initialisation

For certain datasets, EM algorithm may not converge due to an unsuitable initialisation. If we increase the number of parameters used for data modelling, as in a Bayesian inference approach, we are facing an even more challenging problem in choosing their initial values. In this study we adopt a hierarchical approach to the hyperparameter estimation. In the first stage we employ a dual EM algorithm by using a set of random initialisations. After several runs of the EM algorithm on the same data set, we form distributions of its resulting parameters. Afterwards, a maximum log-likelihood criterion is employed by considering a second EM algorithm applied onto the given distributions of parameters.

In the E-step of the first EM algorithm, the *a posteriori* probabilities $\hat{P}^I_{EM}(i|\mathbf{x}_j)$ are estimated. In the M-step we update the parameters $\hat{\alpha}_i, \hat{\mu}_i, \hat{\Sigma}_i$ of the Gaussian mixture model. The two steps E and M are computed alternatively. We run the EM algorithm L times considering various random initialisations. All the parameters estimated in each of the runs are stored individually, forming sample distributions. We assume that these distributions can be characterised parametrically by a set of hyperparameters. The parametric description of these probabilities is given by (3) for means μ, by (4) for covariance matrices Σ, and by (6) for mixing probabilities α.

The next step consists in estimating the hyperparameters characterising the distributions formed in the previous step. This estimation would correspond to a second level of embedding, characterising the initial estimation of the hyperparameters. The distributions of the means resulting from the EM algorithm can be modelled as a mixture of Gaussians. We apply a second EM algorithm onto the distributions of parameters provided by successive runs of the first EM. The equations of the second EM processing are :

$$\hat{P}^{II}_{EM}(i|\mu_j) = \frac{\hat{a}_i |\hat{\mathbf{S}}_i|^{-1/2} \exp[D(\mu_j; \hat{\mathbf{m}}_i, \hat{\mathbf{S}}_i)]}{\sum_{k=1}^N \hat{a}_k |\hat{\mathbf{S}}_k|^{-1/2} \exp[D(\mu_j; \hat{\mathbf{m}}_k, \hat{\mathbf{S}}_k)]} \tag{7}$$

where $D(\mu_j; \hat{\mathbf{m}}_i, \hat{\mathbf{S}}_i)$ is provided by (2). In the M-step of the dual EM we update the parameters of the Gaussian mixture model:

$$\hat{a}_i = \frac{\sum_{j=1}^{LN} \hat{P}^{II}_{EM}(i|\mu_j)}{LN} \tag{8}$$

$$\hat{\mathbf{m}}_{i,EM} = \frac{\sum_{j=1}^{LN} \mu_j \hat{P}^{II}_{EM}(i|\mu_j)}{\sum_{j=1}^{LN} \hat{P}^{II}_{EM}(i|\mu_j)} \tag{9}$$

$$\hat{\mathbf{S}}_i = \frac{\sum_{j=1}^{LN} \hat{P}_{EM}^{II}(i|\mu_j)(\mu_j - \hat{\mathbf{m}}_{i,EM})(\mu_j - \hat{\mathbf{m}}_{i,EM})^T}{\sum_{j=1}^{LN} \hat{P}_{EM}^{II}(i|\mu_j)} \tag{10}$$

In the second stage of the dual EM algorithm we consider randomly picked data samples $\mathbf{x}_j,\ j = 1,\ldots,M$ as the initial values for the hypermeans. The hypermeans $\hat{\mathbf{m}}(0)$ are calculated as the averaging of the resulting means. The corresponding covariance matrices for the Gaussian distribution of means $\mathbf{S}$, are stored as well. The parameter β is calculated as a scaling factor of the covariance matrices corresponding to the initial distributions $\hat{\Sigma}$, obtained by the first EM, to those of the mean distributions $\mathbf{S}$ obtained from (10), respectively. This parameter is initialised as the average of the eigenvalues of the matrix $\hat{\Sigma}\mathbf{S}^{-1}$, which can be calculated as the value of the trace divided by the space dimension:

$$\beta_i(0) = \frac{\sum_{k=1}^{L} Tr(\hat{\Sigma}_{ik}\mathbf{S}_{ik}^{-1})}{dL} \tag{11}$$

where L is the number of runs for the first EM algorithm. The number of degrees of freedom is initialised as equal to the number of dimensions $\nu = d$.

The Wishart distribution $\mathscr{W}(\Sigma|\nu,\mathbf{S})$ characterises the inverse covariance matrix. We initialise the degrees of freedom $\nu_i(0) = d$, while for the initialisation of $\mathbf{S}$ we consider the distribution of $\hat{\Sigma}$. We apply a Cholesky factorisation onto the matrices $\hat{\Sigma}_k$, $k = 1,\ldots,L$ resulted from successive runs of the EM algorithm. The Cholesky factorisation results into an upper triangular matrix $\mathbf{R}_k$ and a lower triangular matrix $\mathbf{R}_k^T$ such that :

$$\hat{\Sigma}_{ik}^{-1} = \mathbf{R}_{ik}\mathbf{R}_{ik}^T \tag{12}$$

We generate L subgaussian random vectors $\mathbf{N}$, each of dimension d, whose coordinates are independent random variables $\mathscr{N}(0,1)$. The matrix $\mathbf{S}$ will be initialised as [8] :

$$\mathbf{S}_i(0) = \frac{\sum_{k=1}^{L} \mathbf{R}_{ik}\mathbf{N}_k(\mathbf{N}_k\mathbf{R}_{ik})^T}{L} \tag{13}$$

For the Dirichlet parameters we use the maximum log-likelihood estimation for (6). After applying the logarithm on (6) and differentiating the resulting expression with respect to the parameters λ_i, $i = 1,\ldots,N$ we obtain the following iterative expression :

$$\psi(\lambda_i(t)) = \psi(\sum_{k=1}^{N} \lambda_k(t-1)) + \log E[\hat{\alpha}_i] \tag{14}$$

where t is the iteration number, $\log E[\hat{\alpha}_i]$ is the expectation of the mixing probability $\hat{\alpha}_i$, and where $\psi(\cdot)$ is the digamma function (the logarithmic derivative of the Gamma function):

$$\psi(\lambda_i) = \frac{\Gamma'(\lambda_i)}{\Gamma(\lambda_i)} \tag{15}$$

where $\Gamma(\cdot)$ function is provided in (5). We consider the mean of mixing probability distributions, estimated in the first EM stage, as an appropriate estimate for $E[\hat{\alpha}_i]$. The hyperparameters λ_i, are calculated by using Newton's method as follows :

$$\lambda_i(t) = \lambda_i(t-1) - \frac{\psi(\lambda_i(t)) - \psi(\lambda_i(t-1))}{\psi'(\lambda_i(t))} \tag{16}$$

Just a few iterations are usually necessary in order to estimate the Dirichlet hyperparameters $\lambda_i(0)$, $i = 1,\ldots,N$.

4 Variational Expectation-Maximization algorithm

Integrating over the entire parameter space would amount to a very heavy computational task, involving multidimensional integration. Variational Bayes algorithm has been used for estimating hyperparameters of mixture models [2, 9]. In our approach we use the initialisation provided by the maximum log-likelihood as described in previous Section. The proposed algorithm is called variational expectation-maximization (VEM) algorithm. The VEM algorithm is iterative and consists of two steps at each iteration: variational expectation (V-E) and variational maximization (V-M). In the first step we compute the *a posteriori* probabilities, given the hidden variable distributions and their hyperparameters. In the V-M step we find the hyperparameters that maximise the log-likelihood, given the observed data and their *a posteriori* probabilities.

In the V-E step we calculate the *a posteriori* probabilities for each data sample $\mathbf{x}_j$, depending on the hyperparameters :

$$\begin{aligned}\hat{P}(i|\mathbf{x}_j) = \quad & \exp\left[-\frac{1}{2}\log|\mathbf{S}_i| + \frac{1}{2}d\log 2 + \frac{1}{2}\sum_{k=1}^{d}\psi\left(\frac{\nu_i+1-k}{2}\right)+\right.\\ & \left. +\psi(\lambda_i) - \psi(\sum_{k=1}^{N}\lambda_k) - \frac{\nu_i}{2}(\mathbf{x}_j-\mathbf{m}_i)^T\beta_i\mathbf{S}_i^{-1}(\mathbf{x}_j-\mathbf{m}_i) - \frac{d}{2\beta_i}\right] \qquad (17)\end{aligned}$$

where i is the mixture component, d is the number of dimensions, j denotes the data index, $\psi(\cdot)$ is the digamma function from (15), and $D(\mathbf{x}_j;\mathbf{m}_i,\beta_i\mathbf{S}_i)$ is provided in (2).

In the V-M step we perform an intermediary calculation of the mean parameter as in the EM algorithm, but considering the *a posteriori* probabilities from (17):

$$\hat{\mu}_{i,VEM} = \frac{\sum_{j=1}^{M}\mathbf{x}_j\hat{P}(i|\mathbf{x}_j)}{\sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)} \qquad (18)$$

The hyperparameters of the mean distribution are updated as follows :

$$\mathbf{m}_i = \frac{\beta_i(0)\mathbf{m}_i(0) + \sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)\mathbf{x}_j}{\beta_i(0)+\sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)} \qquad (19)$$

$$\begin{aligned}\mathbf{S}_i = \quad & \mathbf{S}_i(0) + \sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)(\mathbf{x}_j-\hat{\mu}_{i,VEM})(\mathbf{x}_j-\hat{\mu}_{i,VEM})^T\\ & +\frac{\beta_i(0)\sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)}{\beta_i(0)+\sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j)}(\hat{\mu}_{i,VEM}-\mathbf{m}_i(0))(\hat{\mu}_{i,VEM}-\mathbf{m}_i(0))^T \qquad (20)\end{aligned}$$

while the hyperparameters for Wishart and Dirichlet distributions are updated as :

$$\beta_i = \beta_i(0) + \sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j); \nu_i = \nu_i(0) + \sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j); \lambda_i = \lambda_i(0) + \sum_{j=1}^{M}\hat{P}(i|\mathbf{x}_j) \qquad (21)$$

The effectiveness of the modelling is shown by the increase in the log-likelihood with each iteration. The convergence is achieved when we obtain a small variation in the log-likelihood for the given set of *a posteriori* probabilities. The BIC cost criterion is used to choose the necessary number of mixture components :

$$\mathscr{C}_{VEM}(N) = \sum_{i=1}^{M}\log p(\mathbf{x}_i) - \frac{N}{2}\left[3+d+\frac{d(d+1)}{2}\right]\log M \qquad (22)$$

where the first term is the log-likelihood of the data for $p(\mathbf{x}_i)$, N is the number of components and M of data samples. The number of components N is considered as that corresponding to the largest $\mathscr{C}_{VEM}(N)$.

We assume that each component in the given mixture of Gaussians is assigned to a certain range of colours in the image. Variational segmentation of a colour $\mathscr{V}_k$, $k = 1, \ldots, N$ is obtained following a hard decision onto the resulting *a posteriori* probabilities :

$$\mathscr{V}_k = \{\mathbf{x}_j \mid k = \arg\max_{i=1}^{N} \hat{P}(i|\mathbf{x}_j)\} \quad (23)$$

where $\hat{P}(i|\mathbf{x}_j)$ are the *a posteriori* probabilities obtained at the convergence of VEM algorithm. An important issue in colour image segmentation is the selection of an appropriate colour space and that of a representative set of colours. Image respresentation in the L*u*v* colour space has been found as appropriate to be modelled by Gaussian mixtures [4, 16].

5 Experimental results

The proposed algorithm has been applied for segmenting several colour images. We present the results for three colour images, entitled "Sunset," "Lighthouse," and "Forest" shown in Figure 1. We can observe that "Sunset" image displays a lighting variation in the background, "Lighthouse" contains a mixture of constant colour areas and textures, while "Forest" displays natural textures. The first step consists in transforming the colour coordinate system from RGB to L*u*v* in order to obtain a more adequate colour space for segmentation [4, 7, 13].

The input space is three-dimensional and we consider only pixels resulted from subsampling the images by two on each axis. We apply the variational expectation-maximization (VEM) algorithm as described in Sections 3 and 4. The initialisation is performed by employing the dual EM algorithm. The first EM is run considering 10 different initialisations and a total amount of $10N$ samples are generated. The second EM was initialised with data samples from the given data set. After running the dual EM algorithm on the colour data and onto the resulting EM parameters, we use the maximum likelihood initialisation from Section 3 to initialise the hyperparameters for the VEM algorithm. VEM algorithm provides the set of hyperparameters. We calculate the *a posteriori* probabilities (17) for the entire image. Each image is split in regions based on the colour, considering a Gaussian mixture model for colour images. The hard decision for segmentation is taken by assigning a pixel to that colour region that corresponds to the maximum *a posteriori* probability for that component (23).

Each segmented region $\mathscr{V}_k$, $k = 1, \ldots, N$ is displayed in the colour corresponding to its hypermean in Figures 2, 3 and 4. Segmented "Sunset" image is shown in Figure 2a when considering 7 mixture components, in Figure 2b when considering 10 mixture components and in Figure 2c when using 8 mixture components. In these images we can observe a good separation of the palm-tree from the background as well as the smooth separation of the twilight shadows in the background. Segmented "Lighthouse" image is displayed in Figure 3a when using 8 components, in Figure 3b when using 10 components and in Figure 3c when considering 9 components. In these segmented images we can observe a good separation of the sky from the sea and ground, respectively. In Figure 4a we represent the segmented "Forest" when using 5 components, Figure 4b when using 6 components and in Figure 4c when considering 9 components. In all these images we can observe a good texture segmentation based only on the colour information.

(a) “Sunset” (b) “Lighthouse” (c) “Forest”

Figure 1: Original images to be segmented.

(a) Using 7 components. (b) Using 10 components. (c) Using 8 components.

Figure 2: Segmentation of “Sunset” image with VEM algorithm.

(a) Using 8 components. (b) Using 10 components. (c) Using 9 components.

Figure 3: Segmentation of “Lighthouse” image with VEM algorithm.

(a) Using 5 components. (b) Using 6 components. (c) Using 9 components.

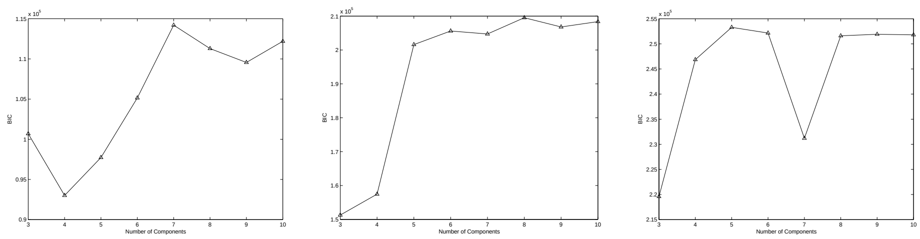
Figure 4: Segmentation of “Forest” image with VEM algorithm.

The number of mixture components has been calculated using Bayesian Information Criterion (BIC) (22). The plots displaying the evaluation of $\mathscr{C}_{VEM}(N)$ for a certain set of components for each image are displayed in Figure 5. From Figure 5a we observe that 7 components are needed to segment the "Sunset" image, from Figure 5b that 9 components would be more appropriate for the "Lighthouse" image, while from Figure 5c we remark that five components would be sufficient for the "Forest" image. As we observe from these plots, according to the BIC criterion, for a small number of components the images are not properly segmented. We can observe that for a certain number of components we reach the saturation in the cost function $\mathscr{C}_{VEM}(N)$.

We compare the proposed variational colour segmentation algorithm with expectation-maximization (EM) algorithm [5, 6, 7, 16]. The average likelihood L_a is given by:

$$L_a = \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \hat{P}(i|\mathbf{x}_j) \qquad (24)$$

For both algorithms we consider the average of ten different runs, considering random initialisation for the EM. Peak signal to noise ratio (PSNR) is calculated between the original image and the segmented image, when assigning the hypermean value to entire segmented regions. The comparison results, considering the average likelihood and the PSNR are shown in Table 1, where the standard deviation is calculated for each result.



(a) "Sunset" image. (b) "Lighthouse" image. (c) "Forest" image.

Figure 5: Estimating the number of mixture of Gaussian components using Bayesian Information Criterion (BIC).

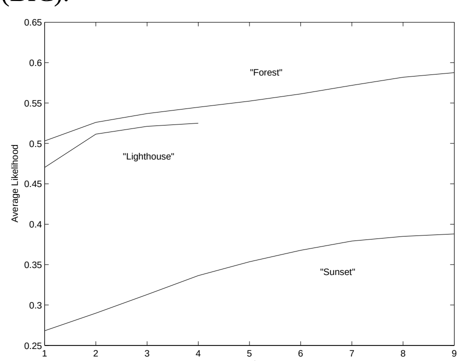


Figure 6: Average likelihood variation with the iteration.

From Table 1 we can observe that the average *a posteriori* probability for "Forest" image is larger than those for "Sunset" and "Lighthouse" images. In all three colour images we have obtained better segmentation results when using VEM compared to the EM. The same convergence criterion has been used for both algorithms, *i.e.* when the log-likelihood L_a from (24) varies with less than 1% from one iteration to another. In

Algorithm	Measure	Colour Images		
		“Sunset” (N=7)	“Lighthouse” (N=8)	“Forest” (N=5)
EM	L_a	0.3676 $\pm$ 0.0051	0.5093 $\pm$ 0.0228	0.4435 $\pm$ 0.0013
	PSNR (dB)	14.76 $\pm$ 2.07	12.08 $\pm$ 2.49	5.35 $\pm$ 0.39
VEM	L_a	0.3759 $\pm$ 0.0063	0.5209 $\pm$ 0.0059	0.5587 $\pm$ 0.0311
	PSNR (dB)	18.64 $\pm$ 0.40	15.88 $\pm$ 1.08	10.96 $\pm$ 0.59

Table 1: Comparison between EM and VEM algorithms in colour image segmentation.

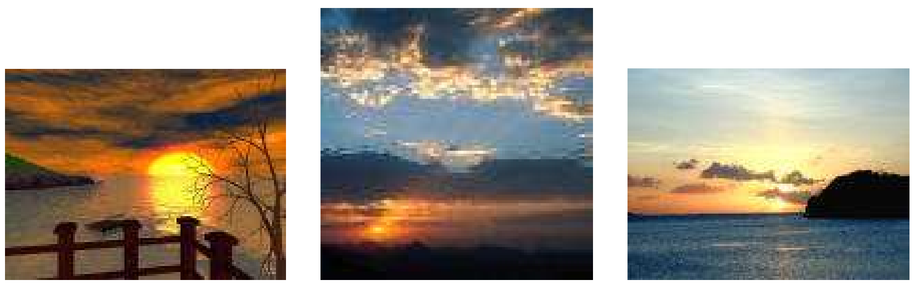


Figure 7: Original images showing sunsets.

Figure 6 we display the variation of L_a with respect to iteration number. Another set of three colour images are shown in Figure 7, while Figure 8 provides their segmentation when using VEM algorithm.

Figure 8: Segmented sunset images using VEM algorithm.

6 Conclusions

We propose a new variational algorithm for colour image segmentation, by considering a mixture of Gaussians model. The variational algorithm is derived from the Bayesian extension of the expectation-maximization algorithm and it is called variational expectation maximization (VEM) algorithm. This paper provides a solution for the initialisation problem in the variational training when representing colour images. The initialisation of the proposed VEM algorithm has two stages. In the first stage we model distributions of parameters resulting from repetitive runs of the EM algorithm on the same image. In the second stage we apply maximum log-likelihood estimation in order to obtain initial hyperparameter estimates for the VEM algorithm. We have used appropriate estimators for the parameter distributions under consideration: Normal for the means, Wishart for the covariance matrix, and Dirichlet for the mixing probabilities. The segmentation results provided by the VEM algorithm are good for various colour images.

References

[1] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *Proc. of Conf. on Uncertainty in Artificial Intelligence*, pages 21–30, Stockolm, Sweden, 1999.

[2] H. Attias. A variational Bayesian framework for graphical models. In *Advances in Neural Infor. Proc. Systems (NIPS) 12*, pages 209–215, 2000.

[3] H.D. Cheng, X.H. Jiang, Y. Sun, and J.L. Wang. Color image segmentation: advances and prospects. *Pattern Recognition*, 34(12):2259–2281, 2001.

[4] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(2):281–288, 2003.

[5] A.P. Dempster, N.M. Laird, and D.B Rubin. Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Stat. Soc., Ser. B*, 39(1):1–38, 1977.

[6] M.A.T. Figueiredo and A.K. Jain. Unsupervised learning of finite mixture models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(3):381–396, 2002.

[7] J.B. Gao, J. Zhang, and M.G. Fleming. Novel technique for multiresolution color image segmentation. *Optical Engineering*, 41(3):608–614, Mar. 2002.

[8] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, 1995.

[9] Z. Ghahramani and M. Beal. Propagation algorithms for variational Bayesian learning. In *Advances in Neural Infor. Proc. Systems (NIPS) 13*, pages 294–300, 2001.

[10] T.S. Jaakkola and M.I. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10:25–37, 2000.

[11] J.L. Marroquin, B.C. Vemuri, S. Botello, and F. Calderon. An accurate and efficient Bayesian method for automatic segmentation of brain MRI. In *Proc. of Eur. Conf. on Comp. Vis. (ECCV), LNCS 2353*, pages 560–574, Copenhagen, Denmark, 2002.

[12] T. Minka and J Lafferty. Expectation-propagation for the generative aspect model. In *Proc. of Conf. on Uncertainty in Artificial Intelligence*, pages 2359–2407, Edmonton, Canada, 2002.

[13] K.N. Plataniotis and A.N. Venetsanopoulos. *Color Image Processing and Applications*. Springer Verlag, Berlin, 2000.

[14] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, 1989.

[15] S.J. Roberts and W.D. Penny. Variational Bayes for generalized autoregressive models. *IEEE Trans. on Signal Processing*, 50(9):2245–2257, 2002.

[16] Z.H. Zhang, C.B. Chen, J. Sun, and K.L. Chan. EM algorithms for Gaussian mixtures with split-and-merge operation. *Pattern Recognition*, 36(9):1973–1983, 2003.