Color Image Segmentation Using a Model-Based Clustering and a MFA-EM Algorithm

Jong-Hyun Park

Department of Computer Science, Chonbuk National University, S. Korea

jhpark@dahong.chonbuk.ac.kr

Abstract. In this paper we present a statistical model-based approach to the color image segmentation. A novel deterministic annealing EM and mean field theory are used to estimate the posterior probability of each pixel and the parameters of the Gaussian mixture model which represents the multi-colored objects statistically. Image segmentation is carried out by clustering each pixel into the most probable component Gaussian. The experimental results show that the mean field annealing EM provides a global optimal solution for the ML parameter estimation and the real images are segmented efficiently using the estimates computed by the maximum entropy principle and men field theory.

1 Introduction

The ultimate goal of image segmentation is to partition a given image into homogeneous regions or segments or pattern classes. This is accomplished by assigning each pixel to the one of K classes. In recent years, a varied segmentation algorithm has been proposed. First, Geman and Geman[1], and Besag[2] have studied the application of the Markov random field(MRF) model and Bayesian methods to segment the intrinsic character of image. Saeed et al.[3], and McLachlan et al.[4] have proposed the Gaussian mixture model for each pixel intensity in an image field and used the expectation-maximization(EM) algorithm to estimate the parameters of the given model. Hofmann et al.[5][6] have formulated the texture segmentation as a data clustering problem based on spare proximity data. They have derived the clustering algorithm using deterministic annealing based on mean field approximation and used this algorithm in the segmentation of texture image.

In many researches, the Gaussian mixture model(GMM) is commonly used to represent the probability distribution of the feature vector observed in each pixel. The EM algorithm is naturally used for the estimation of the parameter in the Gaussian mixture model. However, the estimates of parameters obtained by the EM algorithm are strongly dependent upon their initial values, and the Baum function in the Expectation Step is difficult, if not impossible, to compute since we have to take the summation over all possible configurations of the hidden variables. To overcome this problem, we are going to use the concepts of the deterministic annealing EM algorithm and the mean field theory.

In this paper, we adopt the Gaussian mixture model to represent the probability distribution of the observed feature vector and perform the image segmentation using this model. To solve the initialization problem of the conventional EM algorithm, we are going to consider the concept of deterministic annealing EM algorithm. We also develop the mean field annealing EM(MFA-EM) algorithm for the estimation of the parameters in a mixture model.

2 Deterministic Annealing EM Algorithm

We let denote the incomplete data consisting of the observed feature vector, and z denote the value of unobserved random vector. Also, we let $P(y \mid z; \theta)$ denote the conditional density function of the random vector Y given z, and

$P(z;\pi)$ denote the probability function of Z, where π is a vector of parameters. Then the complete data vector is defined by $x = (y,z)$, and the log likelihood function that could be formed on the basis of the complete data x if it were fully observed is given by

$$LogL_C(\Theta \mid x) = \log P(y \mid z;\theta) + \log P(z;\pi) \qquad (1)$$

where Θ is the vector containing the elements of θ and π. The problem of maximum likelihood estimation of Θ given the observed vector y can be solved by applying the EM algorithm. The EM algorithm starts with some initial guess at the maximum likelihood parameter, $\Theta^{(0)}$, and then proceeds to iteratively generate successive estimates, $\Theta^{(1)}$, $\Theta^{(2)}$, ... by repeatedly applying the following two steps,

E-Step : Compute a posterior density function $P_z^{(t)} = p(z \mid y, \pi^{(t-1)})$, and compute the following expectation: $Q(\Theta \mid \Theta^{(t-1)}) = E_{P_z^{(t)}}(\log L_c(\Theta \mid x))$

M-Step : Find the value, $\Theta^{(t)}$ of the parameter vector,
Θ that maximizes $Q(\Theta \mid \Theta^{(t-1)})$

However the EM algorithm has two kind of disadvantages. The first is hard to avoid unfavorable local maximum of the log-likelihood according to starting value of the parameter and the second is overfitting, i.e. to maximize the performance on unseen future data. Thus we have to think about the method which is able to improve the EM algorithm. It is known as deterministic annealing EM algorithm[6]. This is to use the principle of maximum entropy to estimate the parameter. We consider the complete data log likelihood $LogL_c(\Theta \mid x)$ as a function of the hidden variable z for fixed parameter vector Θ, and define a cost function on the hidden variable space Ω_Z as follows:

$$H(z; y, \Theta) = -\log L_c(\Theta \mid x) \qquad (2)$$

Then we need to minimize $E(H(z; y, \Theta))$ with respect to probability distribution $P(z;\pi)$ over the distribution space subjected to a constraint on the entropy. It yields a quantity, which is known as the generalized free energy in statistical physics. Introducing a Lagrange parameter β, we arrive at the following object function:

$$\vartheta(P_z, \Theta) = E_{P_z}(H(z; y, \Theta)) + \beta \cdot E_{P_z}(\log P_z) \qquad (3)$$

The solution of the minimization problem associated with the generalized free energy in $\vartheta(P_z, \Theta)$ with respect to probability distribution $P(z;\pi)$ with fixed parameter Θ is the following Gibbs distribution:

$$P(z;\pi) = \frac{1}{\sum_{z \in \Omega_z} \exp(-\beta H(z'))} \cdot \exp(-\beta H(z)) \qquad (4)$$

Hence we can obtain a new posterior distribution, $P(z;\pi)$ parameterized by β. For $0 < \beta < 1$, an increase of β means a change in the form of posterior distribution from uniform to the original distribution. Next, we should find the minimum of $\vartheta(P_z, \Theta)$ with respect to Θ with fixed posterior $P(z;\pi)$. It means finding the $\Theta^{(t)}$ that minimizes $\vartheta(P_z, \Theta)$. The generalized free energy, $\vartheta(P_z, \Theta)$ can be

written by the following form:

$$\vartheta(P_z,\Theta) = Q_\beta(\Theta) + \beta \cdot E_{P_z}(\log P_z) \qquad (5)$$

Since the second term on the right hand side of the generalized free energy is independent of Θ, so we should find the value of Θ minimizing the first term

$$Q_\beta(\Theta) = E_{p_z}(H(z;y,\Theta)) \qquad (6)$$

But since an estimate of the parameter value at the early stage of EM-algorithm is not guaranteed to be near the true one, the influence of posterior distribution at the E-step should be weakened at starting point of iteration. Ideally, as the iteration proceeds, the effect should be strengthened. To achieve this purpose, we can add a new β–loop, which is called annealing loop, to the original EM-algorithm and replace the original posterior with the new posterior distribution, $P(z;\pi)$ parameterized by β. Thus we can obtain the deterministic annealing EM-algorithm:

INITIALIZE the parameter $\Theta^{(0)}$ and the prior distribution, $P_z^{(0)}$;
        temperature $T = 1/\beta \leftarrow T_0$;
WHILE $T > T_{FINAL}$; $t = 0,\cdots$;
  REPEAT
    E-Step : Compute a posterior density function
      $P_z^{(t)} = p(z\,|\,y,\pi^{(t-1)})$, and compute the following expectation:
      $Q_\beta(\Theta\,|\,\Theta^{(t-1)}) = E_{P_z^{(t)}}(-\log L_c(\Theta\,|\,x))$
    M-Step : Find the value, $\Theta^{(t)}$ of the parameter vector, Θ that minimize
      $Q_\beta(\Theta\,|\,\Theta^{(t-1)})$
    $t = t+1$;
  UNTIL convergence ;
  $T \leftarrow \eta \cdot T$; $\Theta^{(0)} \leftarrow \Theta^{(t)}$

3 Segmentation of Images Using MFA-EM Algorithm

Suppose that a color image consists of a set of disjoint pixel labeled 1 to N, and that each pixel is assumed to belong to one of K distinct regions. We let Y_i denote the finite dimensional feature vector observed from i th pixel($i = 1,\cdots,N$). Also we let K groups $G_1,\cdots,G_K$ represent the K possible regions. Further, we let $Z_1,\cdots,Z_n$ denote the unobservable group indicator vectors, where the k th element Z_{ik} of Z_i is taken to be one or zero according as the i th pixel does or does not belong to the k th group. We usually adopt a finite mixture model to represent the marginal distribution of the feature vector Y_i observed from the i th pixel($i = 1,\cdots,N$). We let $P_k(y_i;\theta_k)$ denote the conditional probability density function of Y_i given that the i th pixel comes from the k th group G_k ($i = 1,\cdots,N$, $k = 1,\cdots,K$). Then unconditionally with respect to the group of origin, Y_i has the finite mixture form

$$P(y_i;\Theta) = \sum_{k=1}^{K} \pi_{ik} P_k(y_i;\theta_k), \quad (i = 1,\cdots,N) \qquad (7)$$

where π_{ik} is the prior probability that the i th pixel belongs to the k th group.

Given the data and with knowledge of parameter vector Θ, the maximum a posterior estimate, $\hat{z}_i$ of the class indicator vector Z_i at pixel i is defined as :

$$\hat{z}_i = \arg\max_{1 \le k \le N} P(Z_{ik} = 1 | y_i, \Theta) \tag{8}$$

We can proceed to segment an image by assigning class memberships to each pixel individually using the above MAP estimate of the pixel class. In general, each pixel in homogeneous regions of most natural images is correlated with one another. Markov random fields(MRF) have been commonly used to model this correlation. MRF models are not computationally tractable, thus we propose a simplified new model which incorporates neighboring pixel correlation to yield improved segmentation. We let N_i be some specified neighborhood of the i th pixel, containing s pixels. We will assume that two kinds of the prior probability density of Z are given as

$$P(z;\pi) = \prod_{i=1}^{N}\prod_{k=1}^{K} \pi_{ik} \sum_{s \in N_i} z_{sk} \tag{9}$$

where π_{ik} is the prior probability that the i th pixel belong to the k th group.

On the other hand, concerning the joint probability density function of Y given z, a common assumption in image analysis is to take the feature vector Y_i to be independently distributed given their group membership.

$$P(y | z;\theta) = \prod_{i=1}^{N}\prod_{k=1}^{K} P_k(y_i;\theta_k)^{Z_{ik}} \tag{10}$$

In common, $P_k(y_i;\theta_k)$ is taken to be the multivariate normal density $\phi(y_i;\mu_k,\Sigma_k)$ with mean μ_k and covariance matrix Σ_k, so that θ_k contains the elements of μ_k and the distinct element of Σ_k. Hence the two kinds of log likelihood function of the complete data are given by

$$LogL_{C1}(\Theta | x) = \sum_{i=1}^{N}\sum_{k=1}^{K}(Z_{ik}\log\phi(y_i;\theta_k) + \sum_{s \in N_i} Z_{sk} \cdot \log\pi_{ik}) \tag{11}$$

Here, we define a cost function for the segmentation as follows:

$$H_m^{cs}(z;\Theta) = -LogL_{Cm}(\Theta | x), \quad m = 1,2 \tag{12}$$

If we apply the maximum entropy principle discussed in Section 2 for these cost functions, we can obtain the posterior distribution of assignments random vector Z of each pixel for the cost functions. These are given as the Gibbs distribution

$$P(H_m^{cs}(z;\Theta)) = \frac{1}{\sum_{Z' \in \Omega_z}\exp(-\beta H_m^{cs}(z'))} \cdot \exp(-\beta H_m^{cs}(z)), \quad m=1,2 \tag{13}$$

But, since the assignment variables Z_{ik} in each cost functions of the segmentation are statistically dependent with all neighboring pixel variables, the Gibbs distribution cannot be exactly rewritten in factorized form. So, we need the cost contribution that is able to reduce the influence of correlation on individual data assignments. The mean field approximation is a well-known method to approximate the potential energy of an interacting many particles with the average interaction of interacting particles, which is called a mean field[7]. We define an approximating family of distributions with potentials, ε_{ik} which represents the partial cost for independently assigning the pixel to each group. The approximate cost function is defined as

$$H_m^0(z;\Theta) = -\sum_{i=1}^{N}\sum_{k=1}^{K} Z_{ik}\varepsilon_{ik}^m, \qquad m=1,2 \tag{14}$$

To use the mean field approach as an approximation of the segmentation problem, we split the original cost function and write

$$H_m^{cs} = H_m^0 + V_m^p, \qquad m=1,2 \tag{15}$$

where V^p represents a perturbation term due to neglected interactions. The free energy of the cost functions can be rewritten as

$$F(H_m^{cs}) = F(H_m^0) - (1/\beta)\cdot\log < \exp(-\beta V_m^p) >_{H^0} \tag{16}$$

Here, the average bracket $<>$ denote the average with respect to $P(H_m^0(z))$. By the Jensen's inequality, we can obtain the following result,

$$< \exp(-\beta V_m^p) >_{H_m^0} \geq \exp(-\beta < V_m^p >_{H_m^0}) \tag{17}$$

It yields the well-known upper bound

$$F(H_m^{cs}) \leq F(H_m^0) - < V_k >_{H_m^0} \tag{18}$$

Hence, the optimal mean fields come from a variation approach to minimizing the upper bound under free energy. It yields the optimal potential for assigning the i th pixel to group k.

$$\frac{\partial}{\partial \varepsilon_{ik}^m}(F(H_m^0) + < H_m^{cs} - H_m^0 >)\big|_{\varepsilon_{ik}^m = \varepsilon_{ik}^*} = 0 \tag{19}$$

The calculation of the partial derivatives gives us the two optimal potentials

$$\varepsilon_{ik}^{*1} = \log\hat{\pi}_{ik}\phi_k(y_i;\hat{\theta}_k) \tag{20}$$

Thus, the resulted optimal assignment on the E-step of the t th iteration using these potentials is given by

$$\tau_k^{m(t)}(y_i) = E\{Z_{ik} \mid y_i, \Theta^{(t-1)}\} = \frac{\exp(-\beta\varepsilon_{ik}^{*m})}{\sum_{h=1}^{K}\exp(-\beta\varepsilon_{ih}^{*m})}, \qquad m=1,2 \tag{21}$$

and if we replace the assignment variable, $\mathbf{Z}_{ik}$ by its optimal assignment $\tau_k^{m(t)}(y_i)$, we can obtain the Q function.

$$Q_{\beta}^1(\Theta \mid \Theta^{(t-1)}) = -\sum_{i=1}^{N}\sum_{k=1}^{K}\{\tau_k^{1(t)}(y_i)\log\phi(y_i;\theta_k) + \sum_{s\in N_i}\tau_k^{1(t)}(y_i)\cdot\log\pi_{ik}\} \tag{22}$$

On the M-step on the t th iteration, we should choose the value of Θ that maximizes $Q_{\beta}(\Theta \mid \Theta^{(t-1)})$. If we solve the likelihood equation using LaGrange's multiplier method, we can obtain the current fit for the mixing proportions in the two kinds of models,

$$\hat{\pi}_{ik}^1 = \frac{\sum_{s\in N_i}\tau_k^{1(t)}(y_s)}{\sum_{k=1}^{K}\sum_{s\in N_i}\tau_k^{1(t)}(y_s)} \tag{23}$$

We can also obtain the current estimators of the means, and covariance matrices. These are given explicitly by

$$\hat{\mu}_k^{m(t)} = \frac{1}{\sum\limits_{i=1}^{N} \tau_k^{m(t)}(\mathbf{y}_i)} \sum_{i=1}^{N} \tau_k^{m(t)}(\mathbf{y}_i) \cdot \mathbf{y}_i \qquad (24)$$

4 Experimental Results

To assess the performance of the proposed segmentation algorithm, we have conducted the simulated experiments using real images. Fig. 1(a) is an original image that has 100 x 100 pixels and 24bit RGB color levels and consists of multi-colored objects and background. The conversion of the RGB color image to HSI model is carried out and the hue and saturation components are only used as features. Fig. 1(b) shows the color distributions of the block image in HS-space. Here we can see the data consist of four overlapped clusters which represent objects with three color components and one colored background, respectively.
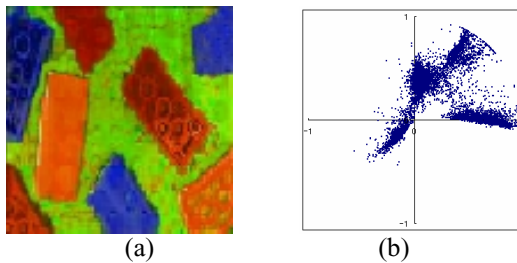

(a)   (b)

Fig. 1 Real image used in the experiment:
(a) original block image, (b) colour distributions in HS-space.

Fig. 2 shows three different initial positions and the convergence results of applying the conventional EM and MFA-EM algorithms to the data in Fig. 1(b) to examine their initialization effects on the parameter estimation. The initial and final component Gaussians are represented by circles and ellipses, respectively, drawn as points with the unit distance from the mean as $(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})=1$. In Fig. 2(a), the initial positions were chosen close to the centers while they were chosen as randomly selected data points in Fig 2(b) and (c).

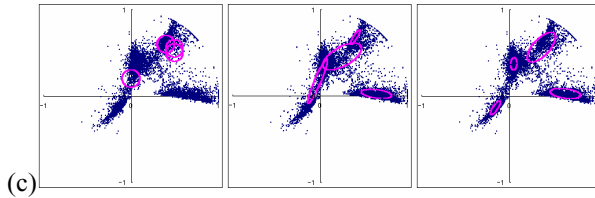Initial position   EM   MFA-EM

(c)
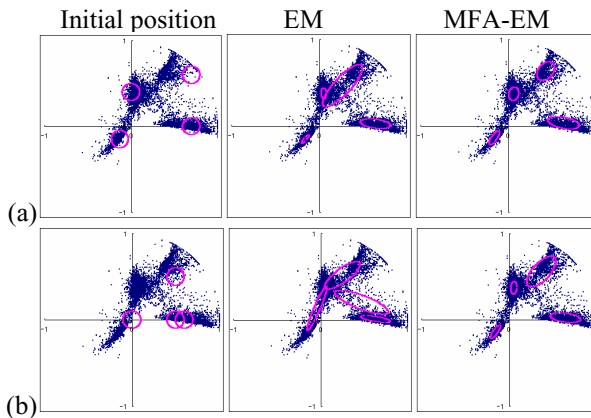
Fig. 2 Results of applying EM and MFA-EM algorithms with three sets of initial conditions.

We can observe that the four component Gaussians estimated by using the conventional EM algorithm converge to each different pattern while those of the MFA-EM algorithm converge to the same pattern regardless of the initial positions. Therefore the MFA-EM algorithm can partition the data points into four clusters more effectively than the conventional EM algorithm. The segmentation results for the block image are shown in Fig. 3(a) and the superpositioned results on the original image are shown in Fig. 3(b) for the boundary examination. The boundaries of the objects are preserved by the MFA-EM algorithm whereas they are degraded by the conventional EM algorithm. Especially the background is regarded as the same region in the segmented result of the MFA-EM algorithm while that of the conventional EM contains lots of small regions.
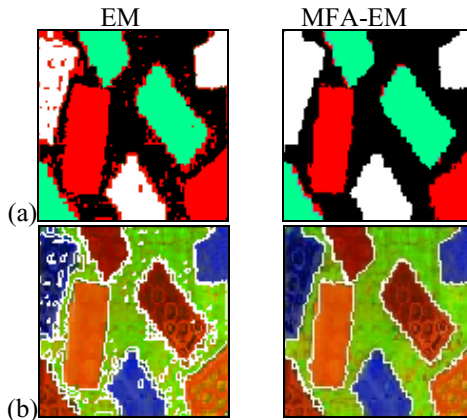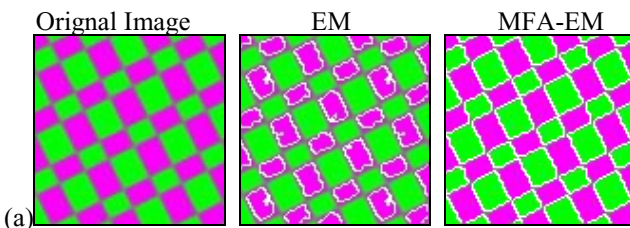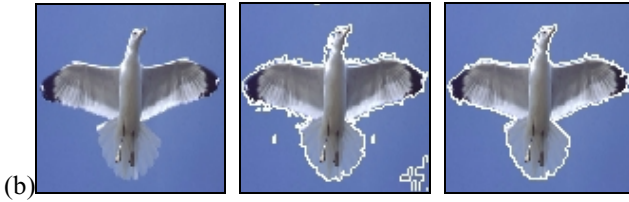
EM   MFA-EM

(a)

(b)

Fig. 3 Segmentation results of block image : (a) segmentation results, (b) segmentation results superpositioned on the original image

Orignal Image   EM   MFA-EM

(a)

(b)

Fig. 4 Segmentation results by the EM and MFA-EM algorithms:
(a) checker board image, (b) bird image

The results of the second experiments for the real images with simple color distributions are shown in Fig. 4. The MFA-EM has been shown to provide a subjectively superior segmentation in that the homogeneous objects are partitioned into the same region accurately and the fine structure is preserved.

5 Conclusions

We have proposed a mean field annealing EM algorithm for the image segmentation. The proposed algorithm was derived using the principle of maximum entropy to overcome the local maximal problem associated with the conventional EM algorithm. For the image segmentation, first, image color distributions were modeled using Gaussian mixture models in hue-saturation space. Then the maximum likelihood estimates of the GMM were obtained by employing the MFA-EM algorithm. Finally, assigning each sample to the regions with the most probable component Gaussian segmented the image.

We conclude from the experiments for the real images that MFA-EM algorithm is robust to initial conditions and a global optimal solution for the ML parameter estimation. It provided a superior segmentation in that each object in homogeneous region was partitioned into the same region accurately. Further work will investigate the effect of the initial temperature value of T and its decreasing rate on the performance of the deterministic annealing algorithm.

References

[1] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," IEEE Transactions on PAMI, vol. 6, pp. 721-741, 1984.

[2] J. Besag, "Towards Bayesian image analysis," Journal of Applied Statistics, vol. 16 , pp. 395-407, 1989.

[3] M. Saeed, W. C. Karl, T. Q. Nguyen, and H. R. Rabiee, "A new multiresolution algorithm for image segmentation," International Conference on ASSP, vol. 5, pp. 2753-2733, 1998.

[4] G. J. McLachlan, S.K. Nguyen, G.J. Galloway, and D. Wang, "Clustering of magnetic resonance images," Technical Report, Department of Mathematics, University of Queensland, 1998.

[5] T. Hofmann and J. M. Buhman, "Pairwise data clustering by deterministic annealing," IEEE Transactions on PAMI, vol. 19, no. 1, pp. 1-13, 1998.

[6] N. Ueda and R. Nakano, "Deterministic annealing EM algorithm", Neural Networks, vol. 11, pp. 271-282, 1998.

[7] J.Zerubia and R.Chellappa, "Mean field annealing using compound Gauss Markov random fields," IEEE Transactions on Neural Networks, vol. 4, no. 4, pp. 703-709, 1993.