

# Image Segmentation via Variational Mixture of Gaussians

ZHANG Yuan-yuan, ZHONG Yi-wei

( Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China )

**Abstract:** Gaussian mixture model (GMM) has been effectively used in image segmentation. In this case, the features of an image are described by a mixture model with  $K$  different components. However, how to choose the number of mixture components  $K$  and estimate model parameters are still short of solutions. Current algorithms such as maximum likelihood and sampling methods are known for their own limitations. So we present an alternative algorithm based on Bayesian variational method and apply it in image segmentation. This method works at less computational cost than sampling methods, and can also naturally handle the model selection problem. In the model's iterative process, the algorithm can automatically determine the number of mixture components in view of the data collected. By comparing our method against other classical segmentation methods on natural images acquired from Berkeley Segmentation Data Set, it suggests that our method provides better performance on image segmentation.

**Key words:** image segmentation; variational inference; Gaussian mixture models; expectation-maximization

**CLC number:** TP391.41

**Document code:** A

**Article ID:** 1001-5132 (2014) 01-0023-06

Image segmentation is an important problem in computer vision which is aim to partition an image into disjoint regions, after that the pixels in the same region have similar visual characteristics. Recently, Gaussian mixture model is such a useful mathematical model used for unsupervised image segmentation<sup>[1-3]</sup>. This method models the image with mixture of Gaussians, and the segmentation is a process of assigning each pixel to the component which it most likely belongs to. In such case, image segmentation is converting to an inference problem that optimizes model parameters to fit the data. A common technique used to inference is Expectation-Maximization (EM) algorithm which is an iterative technique to get the maximum likelihood estimates<sup>[2,4-5]</sup>. The key issue of EM algorithm is model selection problem that is how to determine the number of compo-

nents in order to avoid over-fitting or under-fitting. If the choice of number is inappropriate, the speed of convergence will be very slow and the result of clustering may be poor. Many information rules are selected to try to solve this problem<sup>[6-10]</sup>, such as Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) which compute relative loss information. But they do not know what the true model is like, so with these rules we cannot explain the accuracy in an absolute sense. Even if all the candidate models are not fit well, these rules will not give any hints. Another crucial problem is how to guarantee EM algorithm to get the global optimum which is depends on the initial value of models. Currently improved method is to use  $K$ -means to get a rough estimate, and then use this estimate as the initial value of EM

**Received date:** 2013-06-09.

JOURNAL OF NINGBO UNIVERSITY (NSEE): <http://journalg.nbu.edu.cn/>

**Foundation items:** Supported by the National Natural Science Foundation of China (61175026); Discipline Project of Ningbo University (XKL09154).

**The first author:** ZHANG Yuan-yuan ( 1981- ), female, Ninghai Zhejiang, PhD, experimentalist, research domain: machine learning and computer vision.  
E-mail: zhangyuanyuan@nbu.edu.cn

algorithm<sup>[11]</sup>, but it also increases the computation time.

To tackle above problems in inference problem of Gaussian mixture model, we proposed to use variational inference techniques to find its posterior estimation. Variational inference is a deterministic approximate method, which uses an iterative technique to shrinking  $KL$  divergence between approximate probability distribution and target distribution<sup>[5,12-14]</sup>. We use this mean to determine the appropriate number of components from the data and successfully solve the model selection problem. And if the corresponding distributions belong to exponential family distribution, this algorithm will converge to a global optimum<sup>[15]</sup>. Furthermore, in order to avoid singularity, we extend the standard Gaussian mixture model to a full Bayesian framework, where the model parameters are treated as random variables.

The rest of this paper is organized as follows. Section 1 gives a brief introduction of Gaussian mixture models and extends it to a full Bayesian model. Section 2 presents a mathematical analysis of variational inference, focusing on how to find an optimal estimate for Gaussian mixture model with full Bayesian treatment. Section 3 shows the experimental result on image segmentation. Section 4 concludes the paper and gives some directions for future works.

## 1 Gaussian mixture models

In this section, we give a brief review of the Gaussian mixture models. A Gaussian mixture models is a generative model which is a weighted sum of  $K$  Gaussian distributions, then each observation  $y_i$  has following distributions<sup>[5]</sup>:

$$p(y_i | \pi, \mu, \gamma) = \sum_{k=1}^K \pi_k \mathcal{N}(y_i | \mu_k, \gamma_k^{-1}), \quad (1)$$

where  $\mu = \{\mu_k\}$  are the mean vectors and  $\gamma = \{\gamma_k\}$  are precision matrixes. The parameters  $\pi = \{\pi_k\}$  are mixing coefficient, satisfying  $\sum_{k=1}^K \pi_k = 1, 0 \leq \pi_k \leq 1$ . Furthermore, let  $c_1, \dots, c_N$  denote the labels of components which the observations  $y_1, \dots, y_N$  are generated from. Then we can generate each observation

from following process:

$$\begin{aligned} c_i &\sim \text{categorical}(\pi_1, \dots, \pi_K), \\ y_i | c_i &\sim \mathcal{N}(\mu_{c_i}, \gamma_{c_i}^{-1}). \end{aligned} \quad (2)$$

The graphical model is showed in Fig.1.

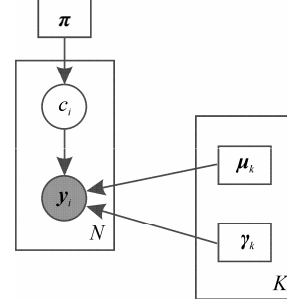


Fig.1 Graphical representation for GMM

In order to make the model more smooth and flexible, we extend Gaussian mixture models with simple representation (see Fig.1) to a full Bayesian approach showing in Fig.2. We treat the weights of the mixture and the parameters as latent variables, where they will be placed prior distributions. In such a case, the weights are often regarded as a  $K$ -dimensional random vector independently drawn from following symmetrical Dirichlet distribution,

$$p(\pi_1, \dots, \pi_K | \alpha_0) = \frac{\Gamma(K\alpha_0)}{[\Gamma(\alpha_0)]^K} \pi_1^{\alpha_0-1} \dots \pi_K^{\alpha_0-1}, \quad (3)$$

where  $\alpha_0$  is the concentrate parameter and  $\Gamma(x)$  is Gamma function. We use a Normal-Wishart distribution as the conjugate prior for multivariate normal likelihood,

$$p(\mu_k, \gamma_k) = \mathcal{N}(\mu_k | \beta_0, (\eta_0 \gamma_k)^{-1}) \mathcal{W}(\gamma_k | \nu_0, \sigma_0). \quad (4)$$

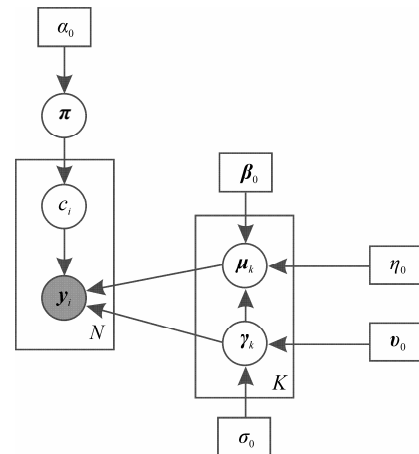


Fig.2 Graphical representation for Bayesian Gaussian mixture model

Thus, a Bayesian Gaussian mixture model assumes the following generative process for each observation:

$$\begin{aligned} \pi_1, \dots, \pi_k &\sim \text{Dir}(\alpha_0), \\ c_i &\sim \text{categorical}(\pi_1, \dots, \pi_k), \\ (\boldsymbol{\mu}_k, \boldsymbol{\gamma}_k) &\sim \text{NM}(\boldsymbol{\beta}_0, \eta_0, \mathbf{v}_0, \sigma_0), \\ \mathbf{y}_i | c_i &\sim \mathcal{N}(\boldsymbol{\mu}_{c_i}, \boldsymbol{\gamma}_{c_i}^{-1}). \end{aligned} \quad (5)$$

When Gaussian mixture model is used in image segmentation, each component is corresponding to an image region which may correspond to some object. So we turn image segmentation into an inference process calculating which segment each pixel is most likely to belong to. EM algorithm is such a standard inference solution, but it needs to specify the number of segments beforehand. Obviously, for an unknown image, it is clearly inappropriate. So we use variational inference instead which is introduced in next section.

## 2 Variational inference

### 2.1 Variational inference

In a probability graphical model, if visible variables are  $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n)$  and latent variables are  $W = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ , the purpose of inference is to estimate its posterior distribution  $p(W|Y)$ . But it is always intractable to calculate posterior distributions in deed, so variational inference is proposed to find an approximate distribution  $q(W)$  instead of the real posterior distribution  $p(W|Y)$ . We measure the degree of approximation between  $q(W)$  and  $p(W|Y)$  with KL divergence, which is

$$KL(q \parallel p) = \int q(W) \ln \frac{q(W)}{p(W|Y)} dW, \quad (6)$$

and there are

$$\ln p(Y) = \mathcal{L}(q) + KL(q \parallel p), \quad (7)$$

$$\mathcal{L}(q) = \int q(W) \ln \frac{p(Y, W)}{q(W)} dW. \quad (8)$$

We can see that  $\mathcal{L}(q)$  is the lower bound of the logarithm of likelihood  $\ln p(Y)$ . So Minimization KL divergence between  $q(W)$  and  $p(W|Y)$  is equal to find an optimal  $q^*(W)$  to maximize the lower bound  $\mathcal{L}(q)$ .

Furthermore, we assume the collection of hidden variables  $W$  can be divided into  $M$  disjoint subsets  $W_1, \dots, W_M$ , then the factorization of  $q(W)$  can be written in the form:

$$q(W) = \prod_{i=1}^M q_i(W_i). \quad (9)$$

After a mathematical transform, we can get the best distribution  $q_j^*$  for each factor as follows:

$$\ln q_j^*(W_j) = E_{i \neq j} [\ln p(Y, W)] + \text{const}, \quad (10)$$

$E_{i \neq j} [\ln p(Y, W)]$  is the expectation of the logarithm of joint probability over observations and latent variables, taken with respect to all factors except  $q_j$ . Thus each factor can be updated given observation and other factors iteratively. In addition, if the posterior distributions belong to exponential family, the convergence of variational inference will be a global maximum.

### 2.2 Variational inference for Gaussian mixture models

In this section, we consider a variational approach for Gaussian mixture models described in Fig.2. We assume the variational distribution is  $q(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\mu}, \boldsymbol{\gamma})$ . Firstly, we factorize the variational distribution as follows<sup>[4]</sup>:

$$q(\boldsymbol{\pi}, \mathbf{c}, \boldsymbol{\mu}, \boldsymbol{\gamma}) = q(\boldsymbol{\pi}) \prod_{i=1}^N q(c_i) \prod_{k=1}^K q(\boldsymbol{\mu}_k, \boldsymbol{\gamma}_k). \quad (11)$$

According to the inference procedure which introduced in section 2.1, we obtain the following variational distribution for each latent variable:

$$q^*(c_i) = \text{categorical}(c_i | \phi_{i1}, \dots, \phi_{iK}), \quad (12)$$

$$q^*(\boldsymbol{\pi}) = \text{Dirichlet}(\boldsymbol{\pi} | \alpha_1, \dots, \alpha_K) \quad (13)$$

$$q^*(\boldsymbol{\mu}_k, \boldsymbol{\gamma}_k) = \mathcal{N}(\boldsymbol{\mu}_k | \boldsymbol{\beta}_k, (\eta_k \boldsymbol{\gamma}_k)^{-1}) \mathcal{W}(\boldsymbol{\gamma}_k | \mathbf{v}_k, \sigma_k). \quad (14)$$

The variational distribution on the label index for observation  $\mathbf{y}_i$  is a categorical distribution with parameter  $\phi_{i1}, \dots, \phi_{iK}$ . The update of each element is

$$\begin{aligned} \ln \phi_{ik} = & \left( \psi(\alpha_k) - \psi\left(\sum_{j=1}^K \alpha_j\right) \right) + \\ & \frac{1}{2} \left( \sum_{j=1}^D \psi\left(\frac{\sigma_k + 1 - j}{2}\right) + D \ln 2 + \ln |\mathbf{v}_k| \right) + \\ & \left( -\frac{D}{2\eta_k} - \frac{\sigma_k}{2} (\mathbf{y}_i - \boldsymbol{\beta}_k)^\top \mathbf{v}_k (\mathbf{y}_i - \boldsymbol{\beta}_k) \right) + \text{const}. \end{aligned} \quad (15)$$

The variational distribution  $q^*(\boldsymbol{\pi})$  is a Dirichlet

with parameters  $\alpha_1, \dots, \alpha_K$ . Each of these is given by

$$\alpha_k = \alpha_0 + N_k. \quad (16)$$

The variational distribution  $q^*(\mu_k, \gamma_k)$  is a Normal-Wishart distribution which has four parameters, the updates are

$$\eta_k = \eta_0 + N_k, \quad (17)$$

$$\beta_k = \frac{1}{\eta_k}(\eta_0 \beta_0 + N_k \bar{y}_k), \quad (18)$$

$$\mathbf{v}_k^{-1} = \mathbf{v}_0^{-1} + N_k \mathbf{S}_k + \frac{\eta_0 N_k}{\eta_0 + N_k} (\bar{y}_k - \beta_0)(\bar{y}_k - \beta_0)^T, \quad (19)$$

$$\sigma_k = \sigma_0 + N_k. \quad (20)$$

To compact the formulas, we introduce four new variables in above equations:

$$N_k = \sum_{i=1}^N \phi_{ik}, \quad (21)$$

$$\bar{y}_k = \frac{1}{N_k} \sum_{i=1}^N \phi_{ik} y_i, \quad (22)$$

$$\mathbf{S}_k = \frac{1}{N_k} \sum_{i=1}^N \phi_{ik} (y_i - \bar{y}_k)(y_i - \bar{y}_k)^T. \quad (23)$$

We can get an approximate posterior distribution on  $\pi, c, \mu, \gamma$  by iterating these equations between two

steps. In the first step, we fix other parameters and optimize  $\{\phi_{ik}\}$  using Eq.(15); in the second step we fix  $\{\phi_{ik}\}$  optimize  $\{\alpha_k\}$  and  $\{\mu_k, \gamma_k\}$  using Eq.(17)-(20).

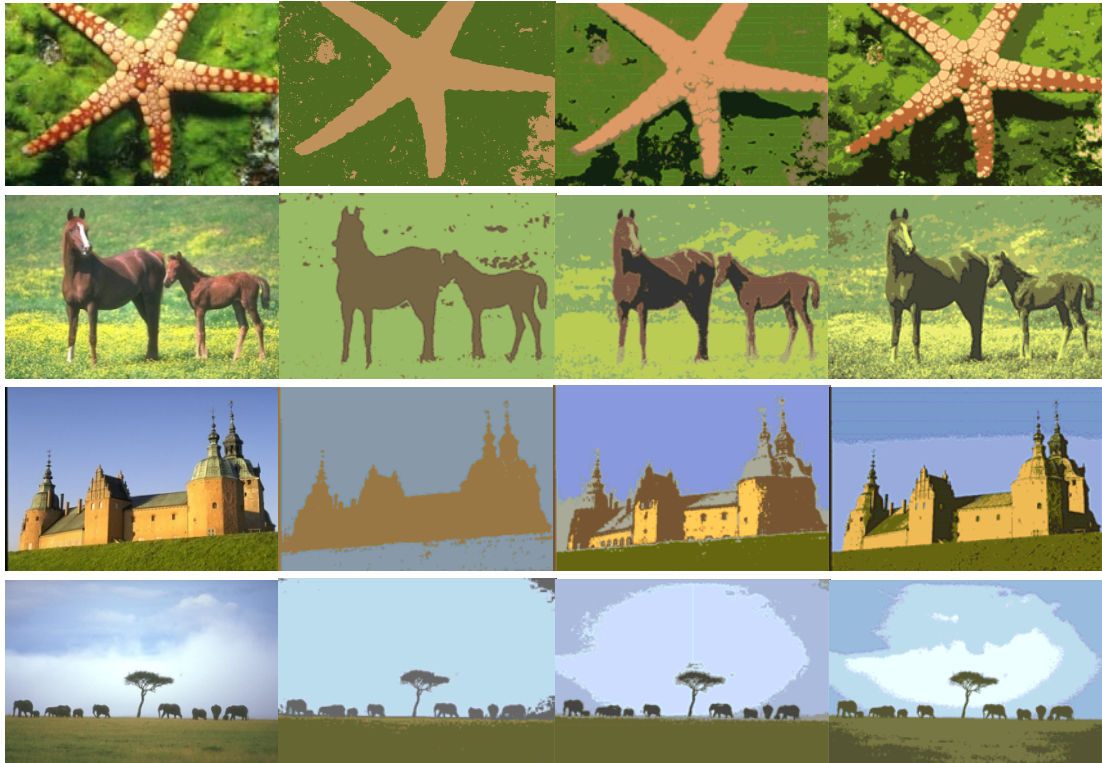
### 3 Experimental result in image segmentation

#### 3.1 Feature extraction

In preprocessing stage, each image is represented as  $L \times a \times b$  color matrix. We use  $7 \times 7$  cut-off window feature to capture local texture. The texture window is smoothed by convolving with a 2D Gaussian kernel before stacking the color values inside the window into a vector. Finally, to reduce the cost of computation, we use PCA to project the feature vectors into 8-dimensional space<sup>[1]</sup>.

#### 3.2 Result

The proposed algorithm has applied in segmenting color images from Berkeley segmentation dataset 500, (<http://www.eecs.berkeley.edu/Research/Projects/CS/>)



the first column: original images; the second column: variational mixture of Gaussians;  
the third column: standard EM algorithm; the fourth column: FCM algorithm

**Fig.3 Segmentation results**

vision/grouping/resources.html#bsds500). The initial number of components is set to 10, which is relatively larger than the true number of segments. In the iterative process, redundant components will be deleted out of the model.

In order to evaluate the usefulness of the proposed algorithm, we compared our segmentation algorithm against standard EM algorithm and the Fuzzy c-means clustering (FCM) algorithm. In Fig.3, the first column is original images. The final result of our algorithm is showed in the second column. The last two columns are comparison results based on EM and FCM. Different segments are labeled with distinguished colors. The speed of these three algorithms is similar. However, we can find that our method has discovered the proper structures of the images, and it can efficiently separate the objects from background. Relatively, the segmentation results based on both EM and FCM are worse because optimization may poor when the initialization of algorithms is not proper enough.

## 4 Conclusion and future work

This paper has presented an unsupervised image segmentation approach using variational mixture of Gaussians. This inference algorithm overcomes model selection problems and brings the global optimum result. Using variational Bayesian inference, we showed obviously improvement over the EM and the FCM in performance of image segmentation application. Although this article is focused on solving the image segmentation, the idea can be extended to other applications as well.

However, there are still some limitations of this method. The main limitation of our algorithm is ignoring the spatial dependence between pixels, so how to add spatial information to the model is a crucial problem. We can put space constraints on the label assignment using Markov Random Field<sup>[16]</sup>. Furthermore, the variational inference described above is operated in a batch mode, because it updates all

variational parameters in a “batch” before updating the model parameters. This inference form may increase the running time of the algorithm, and is not suitable for large-scale problems. Stochastic variational inference has been proposed to aim to solve this problem<sup>[17-19]</sup>. Eventually, this paper only focuses on unsupervised image segmentation, so it is matter of ongoing research that how to join supervised information in the model structure to make more accurate and meaningful segmentation result.

## References:

- [1] Yang A Y, Wright J, Ma Yi, et al. Unsupervised segmentation of natural images via lossy data compression[J]. *Computer Vision and Image Understanding*, 2008, 110(2):212-225.
- [2] Haim P, Joseph F, Ian J. A study of Gaussian mixture models of color and texture features for image classification and segmentation[J]. *Pattern Recognition*, 2006, 39(4):695-706.
- [3] Konstantinos B, Aristidis L, Galatsanos N P, et al. A spatially constrained mixture model for image segmentation[J]. *Neural Networks*, 2005, 16(2):494-498.
- [4] Omer R, Hayit G, Jacob G. Combining region and edge cues for image segmentation in a probabilistic Gaussian mixture framework[C]. *Computer Vision and Pattern Recognition*, MN: Minneapolis, 2007.
- [5] Bishop C M. *Pattern recognition and machine learning* [M]. New York: Springer Presses, 2007.
- [6] Constantinos C, Titsias M K, Aristidis L. Bayesian feature and model selection for Gaussian mixture models[J]. *Pattern Analysis and Machine Intelligence*, 2006, 28(6): 1013-1018.
- [7] Cathy M, Bertrand M. Data-driven penalty calibration: a case study for Gaussian mixture model selection[J]. *ESAIM: Probability and Statistics*, 2011, 15:320-339.
- [8] Figueiredo M A T, Jain A K. Unsupervised learning of finite mixture models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(3):381-396.
- [9] Steele R J, Raftery A E. Performance of Bayesian model selection criteria for Gaussian mixture models[M]. Washington: Dept Stat Univ Washington D C Tech Rep, 2009: 559.
- [10] Alexander A, Thomas L. On image segmentation using information theoretic criteria[J]. *The Annals of Statistics*,

- 2011, 39(6):2912-2935.
- [11] Bradley P S, Fayyad U M. Refining initial points for K-Means clustering[C]. Proceedings of the 15th International Conference on Machine Learning, Canada: Montreal, 1998:91-99.
- [12] Jaakkola T S, Jordan M I. Bayesian parameter estimation via variational methods[J]. Statistics and Computing, 2000, 10(1):25-37.
- [13] Ghahramani Z, Beal M J. Variational inference for Bayesian mixtures of factor analysers[C]. In Advances in Neural Information Processing Systems, 2000:449-455.
- [14] Tzikas D G, Likas A C, Galatsanos N P. The variational approximation for Bayesian inference[J]. Signal Processing Magazine, 2008, 25(6):131-146.
- [15] Wainwright M J, Jordan M I. Graphical models, exponential families, and variational inference[J]. Foundations and Trends in Machine Learning, 2008, 1(1/2):1-305.
- [16] Orbanz P, Buhmann J M. Nonparametric Bayesian image segmentation[J]. International Journal of Computer Vision, 2008, 77(1/3):25-45.
- [17] Blei D M, Jordan M I. Variational inference for Dirichlet process mixtures[J]. Bayesian Analysis, 2006, 1(1):121-144.
- [18] John P, David B, Jordan M I. Variational Bayesian inference with stochastic search[C]. Proceedings of the 29th International Conference on Machine Learning, UK: Edinburgh, Scotland, 2012.
- [19] Hoffman M, Blei D M, Wang Chong, et al. Stochastic variational inference[J]. Journal of Machine Learning Research, 2012, 14:1303-1347.

## 基于变分高斯混合模型的图像分割算法

张媛媛, 钟意伟

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

**摘要:** 提出了一种基于变分推断的高斯混合模型的图像分割算法. 该算法首先用贝叶斯混合高斯模型对图像的特征进行建模, 并针对模型的参数学习问题, 利用变分推断算法估计模型的参数及其后验概率; 这种方法比采样法的计算量更少, 而且能够根据图像数据自动优化混合个数, 实现了模型的自动选择. 最后, 该算法在 Berkeley 的自然图像集上进行的实验结果与经典的图像分割算法进行了比较, 结果表明此方法得到的图像分割结果精度较高, 具有较好的性能.

**关键词:** 图像分割; 变分推断; 高斯混合模型; 期望最大化算法

(责任编辑 章践立)