ENERGY MINIMIZATION-BASED MIXTURE MODEL FOR IMAGE SEGMENTATION

Zhiyong Xiao[†] *Mouloud Adel*[⋆] *Salah Bourennane*[†]

[†] Ecole Centrale Marseille, Institute Fresnel, Marseille 13013, France
[⋆] Aix Marseille Université, Institute Fresnel, Marseille 13013, France
Corresponding Author: Tel: +33 491288076, Email:adzhiyxiao@gmail.com

ABSTRACT

A novel mixture model with spatial constraint is proposed for image segmentation. This model assumes that the pixel label prior probabilities are similar if the pixels are geometric close. An energy function is defined on the spatial space for measuring the spatial information. We also derive an energy function on the observed data space from the log-likelihood function of the standard mixture model. We estimate the model parameters and posterior probability by minimizing the combination of the two energy functions, using the gradient descent algorithm. Numerical experiments are presented where the proposed method is tested on synthetic and real world images. These experimental results demonstrate that the proposed method achieves competitive performance compared to spatially variant finite mixture model.

Index Terms— Energy minimization, mixture model, spatial information, gradient descent algorithm, image segmentation.

1. INTRODUCTION

Image segmentation is an important step in image processing and computer vision. Its goal is to classify image pixels based on the coherence of certain features such as intensity, color, texture, motion, location.

The mixture model [1] is one of the Bayesian-based methods. It is a flexible and powerful technique for image segmentation. The advantages of the standard mixture model is that it has a simpler form, and requires a small number of parameters. However, the main drawback is that the pixels are considered independent in the mixture model. In order to take into account the spatial dependence between image pixels, a set of spatially constrained mixture model have been proposed for image segmentation [2–6].

The spatially variant finite mixture model (SVFMM) [2,3, 6] assumes that the prior distribution form a Markov random field. In [4], the prior probability is based on Gauss-Markov random field, which controls the degree of smoothness for each cluster.

In this paper, we introduce an energy function on feature space based on mixture model. This energy function measures the disagreement between the labels and the observed data. Furthermore, in order to incorporate the spatial relationships into the prior distribution, we propose a energy function on spatial space. This spatial energy function measures the spatial smooth, and it uses the locally invariant idea [7], i.e., the nearby pixels are likely to have similar labels. It has been shown that segmentation performance can be significantly enhanced if the geometrical structure is exploited and the local invariance is considered. We use the gradient descent method [8] to solve the proposed model.

The remainder of this paper is organized as follows. Section 2 describes the details of the proposed model. The details of the gradient descent algorithm are presented in Section 3. In Section 4, we show the experimental results.

2. THE PROPOSED METHOD

Let $\mathcal{X} = \{x_p,\ p \in \mathcal{P}\}$ denote an observed image, where x_p is the observation of pixel p and $\mathcal{P}$ is the set of pixels in the scene. We also denote a label set $\mathcal{F} = \{1, 2, ...K\}$, where K is the total number of classes. The goal of the image segmentation/classification problem is to assigns each pixel p a label $f_p \in \mathcal{F}$.

The image segmentation/classification problem can be formulated in terms of energy minimization. The input is a set of pixels $\mathcal{P}$ and a set of labels $\mathcal{F}$. The goal is to find a mapping from $\mathcal{P}$ to $\mathcal{F}$ which minimizes a particular energy function. A standard form of the energy function [9, 10] is

$$E(f) = E_{data}(f) + E_{spatial}(f) \tag{1}$$

where $E_{data}(f)$ measures the quality of a particular segment of the feature space, while $E_{spatial}(f)$ measure the spatial smooth of class labels.

2.1. Energy Function of Feature

In the proposed model, the prior distribution $P(f_p = k)$ is defined in the following form:

$$P(f_p = k) = \pi_{pk} \tag{2}$$

The prior distribution π_{pk} of pixel p belonging to the class k should satisfies the constraints $0 \le \pi_{pk} \le 1$ and $\sum_{k=1}^{K} \pi_{pk} = 1$.

The proposed model also assumes that the density function at a pixel observation x_p is given by [1]

$$P(x_p) = \sum_{k=1}^{K} P(f_p = k) P(x_p | f_p = k) \tag{3}$$

where $P(x_p | f_p = k)$ is a conditional density on the class label k, i.e.

$$\begin{aligned} P(x_p | f_p = k) &= P(x_p | \theta_k) \\ &= \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left(-\frac{(x_p - \mu_k)^2}{2\sigma_k^2}\right) \end{aligned} \tag{4}$$

where θ_k is the parameter of the kth distribution component. In general, we assume that the probability distribution is a Gaussian one.

The log-likelihood function of the proposed model is given by

$$\begin{aligned} \mathcal{L}(\pi, \theta) &= \sum_{p=1}^{N} \log \sum_{k=1}^{K} P(f_p = k) P(x_p | f_p = k) \\ &= \sum_{p=1}^{N} \log \sum_{k=1}^{K} \pi_{pk} P(x_p | \theta_k) \end{aligned} \tag{5}$$

The log-likelihood function is considered as a function of the parameters π and θ.

Then, we introduce the Jensen's inequality [11], which states that given a set of numbers $\lambda_k \ge 0$ and $\sum_{k=1}^{K} \lambda_k = 1$, we have

$$\log \left(\sum_{k=1}^{K} \lambda_k z_k \right) \ge \sum_{k=1}^{K} \lambda_k \log z_k \tag{6}$$

Note that the posterior probability $P(f_p = k | x_p)$ always satisfies the conditions: $P(f_p = k | x_p) \ge 0$ and $\sum_{k=1}^{K} P(f_p = k | x_p) = 1$. We apply the Jensen's inequality to the log-likelihood function (5), and we have

$$\begin{aligned} \mathcal{L}(\pi, \theta) &= \sum_{p=1}^{N} \log \sum_{k=1}^{K} \pi_{pk} P(x_p | \theta_k) \qquad (7) \\ &= \sum_{p=1}^{N} \log \left(\sum_{k=1}^{K} \pi_{pk} P(x_p | \theta_k) \times \frac{P(f_p = k | x_p)}{P(f_p = k | x_p)} \right) \\ &\ge \sum_{p=1}^{N} \sum_{k=1}^{K} P(f_p = k | x_p) \log \left(\frac{\pi_{pk} P(x_p | \theta_k)}{P(f_p = k | x_p)} \right) \end{aligned}$$

We define

$$\mathcal{L}(f, \pi, \theta) = \sum_{p=1}^{N} \sum_{k=1}^{K} P(f_p = k | x_p) \log \left(\frac{\pi_{pk} P(x_p | \theta_k)}{P(f_p = k | x_p)} \right) \tag{8}$$

Thus, maximizing the log-likelihood function $L(\pi, \theta)$ is equivalent to maximizing the log-likelihood function $L(f, \pi, \theta)$. Since the logarithm is a monotonically increasing function, it is more convenient to consider the negative logarithm of the likelihood function as an energy function. Thus, the energy function of the feature space can be expressed as

$$\begin{aligned} E_{data}(f, \pi, \theta) &= -L(f, \pi, \theta) \qquad (9) \\ &= \sum_{p=1}^{N} \sum_{k=1}^{K} P(f_p = k | x_p) \log \left(\frac{P(f_p = k | x_p)}{\pi_{pk} P(x_p | \theta_k)} \right) \end{aligned}$$

In order to account for the spatial dependence between image pixels, the proposed method present a spatial energy function to measure the spatial information.

2.2. Spatial Energy Function

In this subsection, we describe a different modeling strategy to take the spatial information of the priors into account. A natural assumption could be that if two pixels p, q are close in the geometry, then the prior probabilities π_p and π_q are also close to each other.

In this paper, we use the Gaussian radial basis function to measure the pairwise closeness. The geometric closeness h is a Gaussian function of the magnitude of the relative position vector of pixel p from pixel q, $\| u_p - u_q \|$. The geometric closeness function is given as a decreasing function h when the distance $\| u_p - u_q \|$ increases.

$$h_{pq} = \exp\left(-\gamma \|u_p - u_q\|^2\right) \tag{10}$$

where γ is a free parameter, which defines the desired structural location between neighboring pixels. u_p and u_q are the spatial location of the pixel p and q, respectively. We set $\gamma = 10$.

We can use either Euclidean distance

$$d(\pi_p, \pi_q) = \|\pi_p - \pi_q\|^2 \tag{11}$$

to measure the dissimilarity between prior probabilities π_p and π_q.

With the above defined closeness function h, we can use the following term to measure the smoothness of the prior probability [12].

$$R(\pi) = \frac{1}{2} \sum_{p,q=1}^{N} \|\pi_p - \pi_q\|^2 h_{pq} \tag{12}$$

By minimizing the above term R, we expect that if two pixels p and q are close, the priors π_p and π_q are similar to each other. Thus, we consider this term R as the

$E_{spatial}(f,\pi)$. Combining this spatial energy function with the energy function $E_{data}(f,\pi,\theta)$, the objective energy function becomes

$$E(f,\pi,\theta) = \sum_{p=1}^{N}\sum_{k=1}^{K} P(f_p = k|x_p)\log(P(f_p = k|x_p)) - \sum_{p=1}^{N}\sum_{k=1}^{K} P(f_p = k|x_p)\log(\pi_{pk}P(x_p|\theta_k)) + \frac{1}{2}\sum_{p,q=1}^{N}\|\pi_p - \pi_q\|^2 h_{pq} \quad (13)$$

3. GRADIENT DESCENT ALGORITHM

In this section, we use the gradient method for adjusting the parameters and posterior probability to minimize the above energy function E (13). The proposed algorithm for energy function E (13) can be summarized as follows:

Step 1: Initialize the parameters $\Theta = (\mu, \sigma, \pi)$ by the following substeps.

a) Initialize the mean μ_k and covariance σ_k using the K-means algorithm. The initial value of π_{pk} is set to $1/K$.

b) Calculate the Gaussian distribution $P^{(t)}(x_p|\theta_k)$ from (4). Then, calculate the posterior probability $P(f_p = k|x_p)$, which is given

$$P^{(t)}(f_p = k|x_p) = \frac{\pi_{pk}^{(t)}P^{(t)}(x_p|\theta_k)}{\sum_{k=1}^{K}\pi_{pk}^{(t)}P^{(t)}(x_p|\theta_k)} \quad (14)$$

c) Update the parameters $\mu_k, \sigma_k, \pi_{pk}$ using the following rule

$$\mu_k^{(t+1)} = \frac{\sum_{p=1}^{N} P^{(t)}(f_p = k|x_p)x_p}{\sum_{p=1}^{N} P^{(t)}(f_p = k|x_p)} \quad (15)$$

$$[\sigma_k^2]^{(t+1)} = \frac{\sum_{p=1}^{N} P^{(t)}(f_p = k|x_p)[x_p - \mu_k^{(t+1)}]^2}{\sum_{p=1}^{N} P^{(t)}(f_p = k|x_p)}$$

$$\pi_{pk}^{(t+1)} = \frac{P^{(t)}(f_p = k|x_p)}{\sum_{k=1}^{K} P^{(t)}(f_p = k|x_p)}$$

d) Set $\Theta^t = \Theta^{t+1}$, and return to step b) until $t < T$, $T = 3$.

e) Calculate the posterior probability $P^t(f_p = k|x_p)$.

Step 2: Update parameters $\Theta = (\mu, \sigma, \pi)$ to obtain the new parameters $\Theta^{(t+1)}$, which can be calculated and updated using the gradient method [8]

$$\Theta^{(t+1)} = \Theta^{(t)} - \eta\nabla E(\Theta^{(t)}) \quad (16)$$

where η is the learning rate and its value is sufficiently small. In this paper, we have selected $\eta = 10^{-5}$, and $\nabla E(\Theta^{(t)}) = [\frac{\partial E}{\partial \mu_k}, \frac{\partial E}{\partial \sigma_k}, \frac{\partial E}{\partial \pi_k}]$.

The derivative of E with respect to μ_k, σ_k, π_{pk} are respectively given by

$$\frac{\partial E}{\partial \mu_k} = -\sum_{p=1}^{N} P(f_p = k|x_p)\left(\frac{x_p - \mu_k}{\sigma_k^2}\right) \quad (17)$$

$$\frac{\partial E}{\partial \sigma_k} = -\sum_{p=1}^{N} P(f_p = k|x_p)\left(-\frac{1}{\sigma_k} + \frac{(x_p - \mu_k)^2}{\sigma_k^3}\right)$$

$$\frac{\partial E}{\partial \pi_{pk}} = -\frac{P(f_p = k|x_p)}{\pi_{pk}} + \sum_{q=1}^{N} h_{pq} \parallel \pi_{pk} - \pi_{qk} \parallel$$

Using the above formulas to update the parameters μ, σ, π, and we can obtain the new parameters $\mu_k^{t+1}, \sigma_k^{t+1}, \pi_{pk}^{t+1}$.

Step 3: Update $P(f_p = k|x_p)$ to obtain the new posterior probability. The updating rule is

$$P^{(t+1)}(f_p = k|x_p) = P^{(t)}(f_p = k|x_p) - \eta\nabla E(P^{(t)}(f_p = k|x_p)) \quad (18)$$

The derivative of E, with respect to the posterior probability $P(f_p = k|x_p)$ are given by

$$\frac{\partial E}{\partial P(f_p = k|x_p)} = \log P(f_p = k|x_p) - \log(\pi_{pk}P(x_p|\theta_k)) + 1 \quad (19)$$

Step 4: Check for convergence of the energy function. If the convergence criterion is not satisfied, then set $\Theta^t = \Theta^{t+1}$, $P^{(t+1)}(f_p = k|x_p) = P^{(t)}(f_p = k|x_p)$ and return to Step 2.

Table 1. comparison of the proposed method with other methods (three-class image), Misclassification ratio (%)

Gaussian Noise (0 mean, var)	GMM	SVFMM	ProposedMethod
var=0.01	11.75	3.71	1.16
var=0.02	19.24	10.18	2.33
var=0.03	25.95	16.62	4.51
var=0.04	30.15	22.10	7.74
var=0.05	32.88	26.04	9.85

4. EXPERIMENTS

In this section, we provide experimental results on synthetic and real-world images for evaluating the proposed algorithm.

In the first experiment, we use a synthetic image similar to the one used in [3, 13] (Fig. 1.(a)). Fig. 1.(b) shows the same

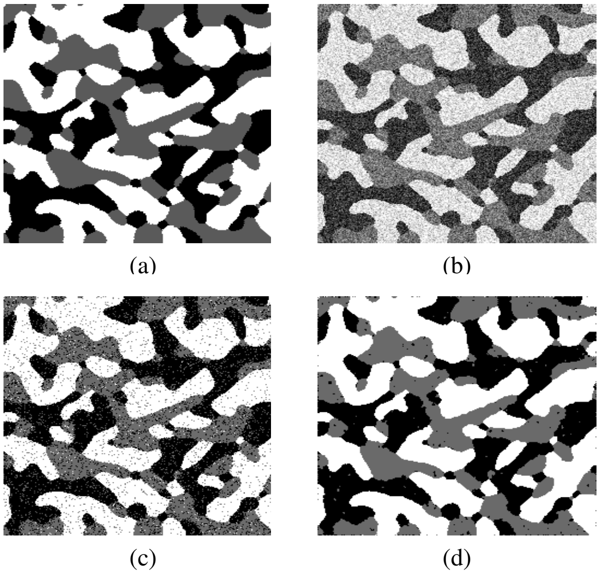

Fig. 1. Experiment on a synthetic image. (a) Original image. (b) Corrupted original image with Gaussian noise(0 mean, 0.02 variance). Results obtained by using (c) SVFMM (MCR=10.18%), (d) the proposed method (MCR=2.33%).

image with added Gaussian noise (mean=0, variance=0.02). We analysis the noise robustness and compare the segmentation result with SVFMM. Fig. 1.(c) and Fig. 1.(d) show the segmentation results obtained by SVFMM and the proposed method, respectively. Compared with these two methods, it is easy to view that the proposed method obtains the best result in noise environment. To further examine its robustness to noise, the evaluation of the proposed method within noisy environment is presented. The results obtained with varying levels of Gaussian noise are presented in Table. 1. As can be seen, the proposed method has a lower MCR compared with Gaussian Mixture Model (GMM) and SVFMM. The misclassification ratio (MCR) [13] is used to measure the segmentation accuracy, which is computed by the ratio between number of misclassified pixels and total number of pixels.

In order to further test the accuracy and determine the efficiency of the proposed method in noise environment, we do experiment on real world image. We use the probabilistic Rand (PR) index [14] to evaluate the performance of the proposed algorithm of the Berkeley database image. It contains values in the range $[0, 1]$, with values closer to 1 indicating a good result. Fig. 2.(a) shows a color image. We tried to segment the image into three classes. The image in Fig. 2.(b) shows the segmentation result obtained by the proposed method. The image shown in Fig. 2.(c) is obtained by corrupting the original image in Fig. 2.(a) with Gaussian noise (0 mean, 0.02 variance). Fig. 2.(d) presents the segmentation result obtained by using the proposed method. For visual in-

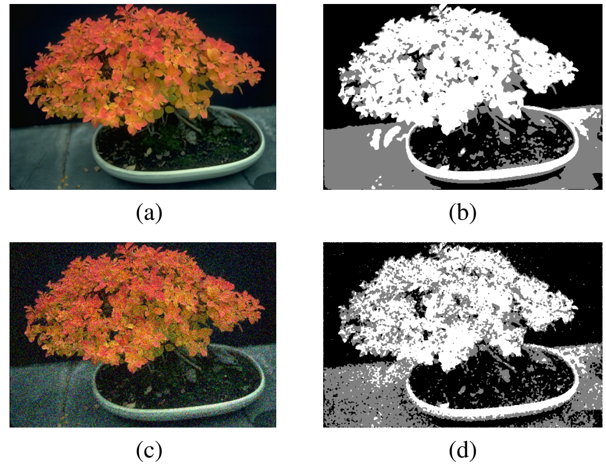

Fig. 2. Experiment on a real world image. ((a) Original image. (b) Segmentation result obtained by the proposed method (PR=0.722). (c) Corrupted original image with Gaussian noise (0 mean, 0.02 variance) and its three-class segmentation by (d) Segmentation result obtained by the proposed method (PR=0.669).

spection of the results, the proposed method also produces a good segmentation result in a noisy environment.

5. CONCLUSION

In this paper, we presented a energy minimization-based mixture model for image segmentation. The proposed method considers both the energy of the data space and the energy of the spatial space. The proposed model postulates that the unobserved pixel labels generated by prior distributions are similar if the pixels are close in geometry of the data distribution. The kernel function is used to measure the geometry closeness of the pixels. We use the gradient descent algorithm to estimate the parameters of the model and the posterior probabilities. The proposed method has been tested on synthetic and real world images. The experimental results show excellent performance of the proposed model in segmenting the images compared to the GMM and SVFMM.

6. REFERENCES

[1] Franck Picard, "An introduction to mixture model," *Statistics for Systems biology Research Report No.7*, 2007.

[2] S. Sanjay-Gopal and T. J. Hebert, "Bayesian pixel classification using spatially variant finite mixture and the generalized *EM* alogirhm," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1014–1028, 1998.

[3] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, 2005.

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

[5] C. Nikou, A. Likas, and N. P. Galatsanos, "A bayesian framework for image segmentation with spatially varying mixtures," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2278–2289, 2010.

[6] Aritseidis Diplaros, Nikos Vlassis, and Theo Gevers, "A spatilly constrained generative mode and an *EM* algorithml for image segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 798–808, 2007.

[7] D. Cai, X. He, and J. Hani, "Document clustering using locality preserving indexing," *IEEE Trans. Knowledge and Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.

[8] C. M. Bishop, "Neural networks for pattern recognition," *Oxford, U.K.:Oxford University Press*, pp. 59–73,257–292, 1995.

[9] Yuri Boykov, Olga Veksler, and Ramin Zabin, "Fast approximate energy minmization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1–18, Nov. 2001.

[10] Vladimir Kolmogorov and Ramin Zabin, "What energy functions can be minimized via graph cuts ?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2000.

[11] W. Rudin, "Real and complex analysis," *3rd ed. New York: McGraw-Hill*, pp. 62–65, 1987.

[12] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, 2011.

[13] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain *MR* images through a hidden markov random field model and the epectation-maximiztion algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, 2001.

[14] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "A measure for objective evaluation of image segmentation algorithms,," *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 3, pp. 34–41, Jun. 2005.