

## Image Segmentation with the EM and the BYY Learning

Kai Tian and Hongzhang Jin

College of Automation, Harbin Engineering University, Harbin, China

**Abstract:** The aim of the study is to present an image segmentation method based on feature space clustering with the Gaussian Mixture Model (GMM) and the Expectation Maximization (EM) algorithm. To address the issue of selection of determining the number of clusters, the Bayesian Ying-Yang (BYY) learning is employed. Moreover, to improve the performance of the segmentation on noisy images, both intensity and spatial position information are employed as features to describe a pixel in the image. The simulation results on images with and without noises validate the performance of the proposed method.

**Keywords:** Bayesian ying-yang learning, clusters, feature space, images segmentation

### INTRODUCTION

One crucial step in image processing and pattern recognition is the segmentation which continues to be a challenging research area and greatly determines the quality of many existing techniques including object tracking, image recognition and object-based image compression. Image segmentation is a process of dividing an image into different regions such that each region is, but the union of any two adjacent regions is not, homogeneous (Cheng *et al.*, 2001). In the past decades, many techniques have been proposed to deal with the image segmentation problem, including thresholding, feature space clustering, edge-based techniques, region-based techniques, graph-based techniques and hybrid techniques (Cheng *et al.*, 2001; Tao *et al.* 2007).

The image segmentation based on feature space clustering are termed as unsupervised classification methods which organize unlabeled feature vectors into clusters or “natural groups” such that samples within a cluster are more similar to each other than samples belonging to different clusters (Tang *et al.*, 2009). One popular technique for feature space clustering is the model-based technique which has received much attention during the last decades. The most commonly used statistical models are Markov random field models (Cohen *et al.*, 1991; Zoltan and Pong, 2006; Li, 2009), hidden Markov models (Ibrahim *et al.*, 2006) and Gaussian mixture models (Deng and Clausi, 2004; Carreira-Perpinan, 2007; Kim and Tang, 2007; Liu *et al.*, 2008; Tang *et al.*, 2009; Nguyen and Wu, 2012).

In this study the Gaussian mixture model (GMM) is adopted and discussed. Let  $\mathbf{x}$  be a  $d$ -dimensional random variable, the goal of clustering is to assign  $\mathbf{x}$  to a cluster according to the criterion of minimizing the

measure between  $\mathbf{x}$  and the center of the cluster, with  $\mathbf{x}$  has the following probability density:

$$p(\mathbf{x}) = \sum_{y=1}^k \alpha_y p(\mathbf{x} | \theta_y) \quad (1)$$

where, the weights  $\alpha_y > 0$  and  $\sum_{y=1}^k \alpha_y = 1$  and the finite mixture  $p(\mathbf{x} | \theta_y)$ ,  $y = 1, 2, \dots, k$  is a Gaussian density given by:

$$p(\mathbf{x} | \theta_y) = \frac{1}{(2\pi)^{d/2} |\Sigma_y|} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_y)^T \Sigma_y^{-1} (\mathbf{x} - \mathbf{m}_y)\right\} \quad (2)$$

with  $\mathbf{m}_y$  and  $\Sigma_y$  are the mean and the covariance, respectively and  $\theta_y = (\mathbf{m}_y, \Sigma_y)$ .

It can be clearly seen from Eq. (1) and Eq. (2) that two key issues in GMM-based image segmentation are the evaluation of parameters  $\theta_y = (\mathbf{m}_y, \Sigma_y)$  and the selection of the number of clusters. Generally speaking, the first issue can be solved by an approximate Maximum A Posteriori (MAP) estimation or the maximum likelihood estimation which can be obtained by the Expectation Maximization (EM) (Dempster *et al.*, 1977; Redner and Walker, 1984). However, the second problem has not been solved efficiently yet. In addition, it is inevitable to map an image into a feature space. Thus, besides the above two issues, another problem that have to be solved for GMM-based segmentation is the selection of features to represent an image.

First proposed Xu (1995) and developed for over a decade, Bayesian Ying-Yang (BYY) learning provides a general framework that accommodates typical learning algorithms from a unified perspective and

improved model selection criteria. BYY learning consists of two subcategories. One is featured with Ying-Yang best matching for developing typical learning algorithms, which is one major focus of Xu (1995) and Xu (1998), while the other is Ying-Yang best harmony featured with its favorable nature for model selection (Xu, 2004).

Since the studies have shown that BYY is capable of selecting the clusters of GMM (Xu, 1997), this study adopts the BYY learning to solve the problem of the selecting the number of clusters associated to the GMM-based image segmentation. Considering that one important difference between the objects in an image may be the intensity (or color), adopting intensity information in image segmentation is useful and effective. Besides, the spatial position is also an important factor when human beings perform segmentation. Thus, this study adopts the features consisting of both intensity (color) and spatial coordinates to represent a pixel.

### TEXT IMAGE SEGMENTATION BASED ON EM

Before performing the segmentation, we have to map a gray-level image to a  $d$ -dimensional feature space via representing each pixel by a feature vector  $\mathbf{x} \in \mathbb{R}^d$ . Assume  $\mathbf{x}$  has the probability density given by Eq. (1) and (2). To evaluate the parameters  $\theta_y = (m_y, \Sigma_y)$ , the following EM algorithm which is capable of updating the parameters of the GMM with the sample data algorithm can be used. The EM algorithm consists of two iterative steps, i.e. the E-Step and the M-Step.

**E-Step:** Computing the posterior probability with the equation:

$$p^*(y | x, \Theta) = \frac{\alpha_y p(x | \theta_y)}{\sum_{y=1}^k \alpha_y p(x | \theta_y)} \quad (3)$$

where, the parameter set  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$  and  $k$  is the number of clusters.

**M-Step:** maximizing the likelihood function, obtaining the new value of parameter set  $\Theta$ . For a GMM, the new value of  $\Theta$  are given by:

$$\begin{aligned} \alpha_y^{new} &= \frac{1}{N} \sum_{y=1}^k \alpha_y p^*(y | \mathbf{x}, \mathbf{m}_y, \Sigma_y) & (4) \\ \mathbf{m}_y^{new} &= \frac{\sum_{j=1}^N x_j p(y | x_j, \mathbf{m}_y, \Sigma_y)}{\sum_{j=1}^N p(y | x_j, \mathbf{m}_y, \Sigma_y)} \\ \Sigma_y^{new} &= \frac{\sum_{j=1}^N p(y | \mathbf{x}_j, \mathbf{m}_y, \Sigma_y) (\mathbf{x}_j - m_y^{new})(\mathbf{x}_j - m_y^{new})^T}{\sum_{j=1}^N p(y | \mathbf{x}_j, \mathbf{m}_y, \Sigma_y)} \end{aligned}$$

where,  $N$  is the number of pixels in the image.

The EM algorithm iteratively runs E-step and M-step until the following condition being stratified:

$$\log(L(\Theta | \mathbf{x})) = \log \prod_{j=1}^N p(\mathbf{x}_j | \Theta) < \varepsilon \quad (5)$$

where,  $\varepsilon > 0$  is a small real number.

After getting the optimal parameter set  $\Theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_k^*)$  of the GMM by using the above described EM algorithm, each pixel which is featured by vector  $\mathbf{x}$  will be classified into a class  $\hat{y}(\mathbf{x}) \in \{1, 2, \dots, k\}$  according to:

$$\hat{y}(\mathbf{x}) = \arg \max_y p(y | \mathbf{x}) = \arg \max_y \alpha_y p(\mathbf{x} | \Theta) \quad (6)$$

Finally, a image is segmented into different regions by labeling each pixel according the class it assigned.

Though above EM-based segmentation algorithm is theoretically efficient and easy to be implemented, one critical parameter that have to be predefined is that of the number of clusters (or number of classes), i.e., the parameter of  $k$ . A bad estimate of  $k$  will lead to serious problems (Xu *et al.*, 1993).

### IMAGE SEGMENTATION WITH EM AND BYY

**Selection of the optimal number of clusters with by learning:** In the BYY learning (Xu, 1997), there are two primary elements: the external observations  $\mathbf{x}$  and the output action  $y$ . The  $\mathbf{x}$  are known (visible), but the  $y$  is unknown (invisible). All of these elements are treated as random variants and the joint distribution  $p(\mathbf{x}, y)$  can be calculated in two ways:

$$\begin{cases} p_{ying}(\mathbf{x}, y) = p(y)p(\mathbf{x} | y) \\ p_{yang}(\mathbf{x}, y) = p(\mathbf{x})p(y | \mathbf{x}) \end{cases} \quad (7)$$

Practically, the results of these two equations are always not equal unless we can find the optimal solution of  $p(y)$ ,  $p(\mathbf{x}|y)$ ,  $p(\mathbf{x})$  and  $p(y|\mathbf{x})$ . The core of the BYY learning is that the specification of a Ying-Yang pair above enhances best the so-called Ying-Yang harmony by minimizing a harmony measure,  $F_s(M_{yang}, M_{ying})$ , which is defined as follows:

$$\begin{aligned} F_s(M_{yang}, M_{ying}) &= F_s(p_{ying}(\mathbf{x}, y), p_{yang}(\mathbf{x}, y)) \geq 0 & (8) \\ F_s(M_{yang}, M_{ying}) &= 0 \Leftrightarrow p_{ying}(\mathbf{x}, y) = p_{yang}(\mathbf{x}, y) \end{aligned}$$

Here, the Kullback divergence is adopted and getting:

$$F_s(M_{yang}, M_{ying}) = \int_{x,y} p(y | \mathbf{x}) p(\mathbf{x}) \ln \frac{p(y | \mathbf{x}) p(\mathbf{x})}{p(\mathbf{x} | y) p(y)} dx dy \quad (9)$$

Since the architecture and the parameter set  $\Theta$  of the EM based GMM has been fixed, the remaining task is to find the optimal number of clusters (or the optimal number of classes). According to the BYY learning, the selection of the number of clusters, i.e., the parameter  $k$ , can be determined in two ways. The first one is according to the following function:

$$k^* = \arg \min_k J_1(k) = \arg \min_k (F_s(M_{yang}, M_{ying})) \quad (10)$$

Since the architecture has been fixed,  $F_s(M_{yang}, M_{ying})$  is only the function of  $k$  and the parameter set  $\Theta$ . Assume the parameter set obtained by EM algorithm is  $\Theta^*$ , then we have  $F_s(M_{yang}, M_{ying}) = F_s(k, \Theta^*)$ . Thus Eq. (1) can be rewritten as:

$$k^* = \arg \min_k J_1(k) = \arg \min_k (F_s(k, \Theta^*)) \quad (11)$$

The other way to get the optimal  $k$  is:

$$k^* = \arg \min_k J_2(k) = \arg \min_k \left\{ -\int_{x,y} p_{yang}(\mathbf{x}, y) \Big|_{\Theta^*} \ln p_{ying}(\mathbf{x}, y) \Big|_{\Theta^*} d\mathbf{x}dy \right\} \quad (12)$$

In the special case for GMM, the above function  $J_1$  and  $J_2$  degrade to

$$J_1(k) = -O_N(p(y|x) | \Theta^*) + J_2(k) \quad (13)$$

and

$$J_2(k) = \sum_{y=1}^k \alpha_y^* \ln \sqrt{|\Sigma_y^*|} - \sum_{y=1}^k \alpha_y^* \ln \alpha_y^* \quad (14)$$

where,  $\alpha_y^*$ ,  $\Sigma_y^*$  and  $\Theta^*$  are the parameters of the GMM determined with the EM algorithm.

**Image segmentation with EM and BYY:** Before performing the segmentation, we have to map a gray-scale image into  $d$ -dimensional feature, i.e. representing each pixel by a vector  $\mathbf{x} \in \mathbb{R}^d$ . Considering that one major difference between the objects in an image may be the intensity (or color), adopting intensity (or color) information in image segmentation is useful and effective. Besides, the spatial position information is also an important factor when human beings perform segmentation. Thus, this study adopts the features consisting of both intensity (color) and spatial coordinates to represent a pixel. Given a gray-scale image  $I$ , we use the following function to map  $I$  into 3-dimensional feature space.

$$\begin{aligned} x_{i1} &= I(\text{mod}(i, W), \lfloor i/W \rfloor + 1) / 256 \\ x_{i2} &= \text{mod}(i, W) / H \\ x_{i3} &= (\lfloor i/W \rfloor + 1) / W \end{aligned} \quad (15)$$

where,  $I(m,n)$  is the intensity of the gray-scale image  $I$  at the position  $(m,n)$ ,  $N=W \times H$  is the total pixels of  $I$  and  $W$  and  $H$  are the width and height of the image, respectively. The feature vector  $\mathbf{x}_i$  is constructed by:

$$\mathbf{x}_i = \{\lambda_1 x_{i1}, \lambda_2 x_{i2}, \lambda_3 x_{i3}\} \quad (16)$$

where, the weight  $\lambda_i > 0$  is used to emphasis the contribution of the  $i$ th feature component to the classification. Since the criterion function  $J_2(k)$  is U-shape and the  $k$  corresponding to the bottom of the U-Shape is the optimal number of clusters.

With Eq. (15) and Eq.(16), a gray-scale image  $I$  is represented by  $\{\mathbf{x}_i \in \mathbb{R}^3 | i=1,2,\dots,N\}$ . Assume the probability density of  $\mathbf{x}_i$  is the GMM given by Eq. (1) and Eq. (2). Then, the following iterating procedure is employed to perform the segmentation.

**Step 1:** Let the number of classes  $k = 1$ .

**Step 2:** Employ the EM algorithm described by Eq. (3)~Eq. (5) to estimate the parameter set:

$$\Theta^* = \{\alpha_y^*, m_y^*, \Sigma_y^* | y = 1, 2, \dots, k\}$$

**Step 3:** Compute the criterion function  $J_2(k)$  according to Eq.(14).

**Step 4:** If  $k > 1$ , compute:

$$\Delta J(k) = J_2(k) - J_2(k-1) \quad (17)$$

**Step 5:** If  $k > 2$  and

$$\text{sgn}(\Delta J(k)) \neq \text{sgn}(\Delta J(k-1)). \quad (18)$$

Then record the optimal number of clusters  $k^* = k-1$  and stop the procedure. Otherwise,  $k = k+1$ , go to step 2.

## SIMULATION RESULTS AND ANALYSIS

**Results on noise-free images:** The proposed image segmentation method based on the EM and the BYY learning has been implemented in MatlabR2007b. Two images have been used to evaluate the performance of the proposed method. The first one is a synthetic image consisting of three gray levels as shown in Fig. 1a. The second image is the bench mark ‘‘House’’ picture as shown in Fig. 1b. The image size is 128 by 128.

The feature weights  $\lambda_i$  are 10, 5 and 5 for the features of intensity, y-position and x-position, respectively. The  $J_2(k)$  obtained from Eq.(14) for the three images shown in Fig. 1 are plotted in Fig. 2. From Fig. 2 one can see that the  $J_2(k)$  has U-shape. Figure 3 draws the segmentation results corresponding to the  $k^*$

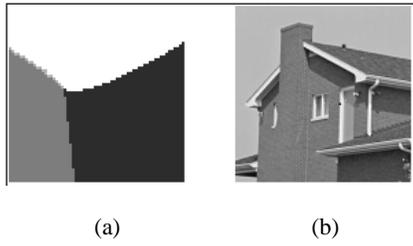


Fig. 1: Test images, (a) Synthetic image, (b) House

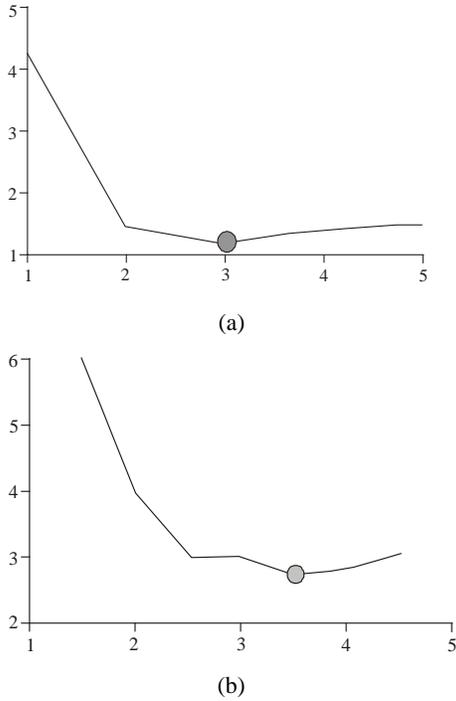


Fig. 2:  $J_2(k)$  with (a) and (b) corresponding to Fig. 1a and b



Fig. 3: Segmentation results

where  $J_2(k)$  has a minimum value. From Fig. 3, one can see that by using the proposed method given by Eq. (17) and Eq. (18), a satisfied segmentation results can be obtained.

**Results on noise images:** Different level of Gaussian noises are added to Fig. 1a as shown in Fig. 4. The weights for different feature components are also 10, 5 and 5. the  $J_2(k)$  are shown in Fig. 5.

Comparing Fig. 5 and 4a, one can see that for the noise images, the shape of  $J_2(k)$  is almost the same as

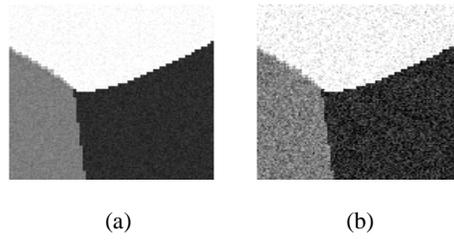


Fig. 4: Images with Gaussian noises (a) Local variance is 0.001 and (b) Local variance is 0.01

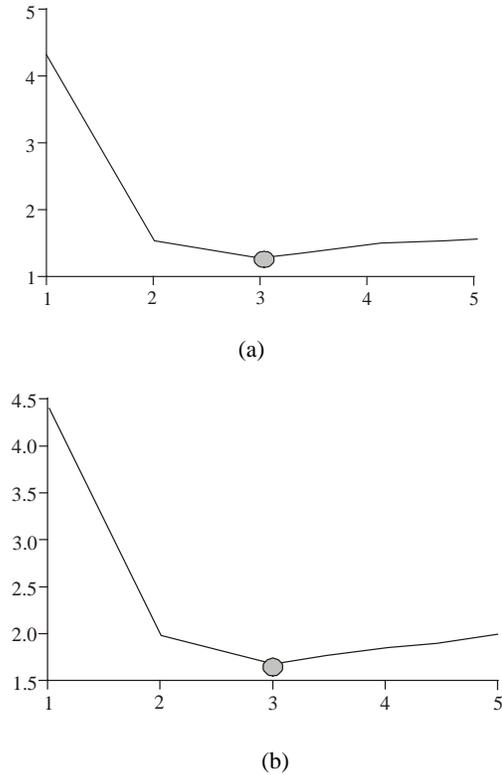


Fig. 5:  $J_2(k)$  for images with Gaussian noise, (a) Local variance is 0.001 and (b) Local variance is 0.01

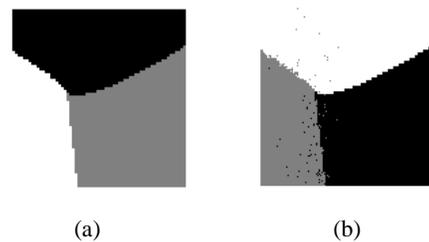


Fig. 6: Segmentation results on noisy image, (a) Local variance is 0.001 and (b) Local variance is 0.01

that for the noise-free images.  $J_2(k)$  is also U-shape, meaning that the Eq.(17) and Eq.(18) are still effective. The segmentation results with  $k^*$  corresponding to the minimum of  $J_2(k)$  are shown in Fig. 6, which clear shows that the proposed method is robust to noise.

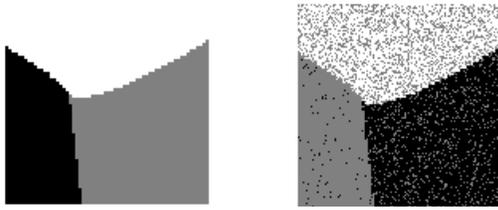


Fig. 7: Segmentation results on noisy image with intensity feature, (a) Local variance is 0.001 and (b) Local variance is 0.01

**Evaluation on the proposed feature representation method:** One highlight of the proposed method is that not only the intensity of a pixel, but the x-position and y-position of the pixel are also used to construct a feature vector as described in Eq. (15) and Eq. (16). Figure 7 shows the segmentation results of the images shown in Fig. 5 with the optimal  $k^*$  obtained from the proposed method but use only the intensity feature, i.e., the feature vector for each pixel  $x_i = (x_{i1})$ . Comparing Fig. 6 to 7, one can see that though the method employing intensity feature can find the correct  $k^*$ , it is non-robust to noise.

## CONCLUSION

The EM and BYY based image segmentation has been proposed. Two contributions can be claimed. One is that the BYY learning is introduced to the EM-based image segmentation to address the selection of the number of clusters. Another contribution is that both the intensity and the spatial position information are employed as features to describe a pixel in the image, with which the proposed method has a better performance on the noisy images.

## ACKNOWLEDGMENT

The work of this study is partly supported by the National Natural Science Foundation of China (Grant No. 61003128).

## REFERENCES

Carreira-Perpinan, M.A., 2007. Gaussian mean-shift is an EM algorithm. *IEEE T Pattern Anal.*, 29(5): 767-776.

Cheng, H.D., X.H. Jiang, Y. Sun and J. Wang, 2001. Color image segmentation: Advances and prospects. *Pattern Recogn.*, 34(12): 2259-2281.

Cohen, F.S., Z. Fan and M.A. Patel, 1991. Classification of rotated and scaled textured images using Gaussian Markov random field models. *IEEE T. Pattern Anal.*, 13(2): 192-202.

Dempster, A.P., N.M. Larid and D.B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39(1): 1-38.

Deng, H. and D.A. Clausi, 2004. Gaussian MRF rotation-invariant features for image classification. *IEEE T. Pattern Anal.*, 26(7): 951-955.

Ibrahim, M., N. John, M. Kabuka and A. Younis, 2006. Hidden Markov models-based 3D MRI brain segmentation. *Image Vision Comput.*, 24(10): 1065-1079.

Kim, S.C. and T.J. Tang, 2007. Texture classification and segmentation using wavelet packet frame and Gaussian mixture model. *Pattern Recogn.*, 40: 1207-1221.

Li, S.Z., 2009. *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, London.

Liu, X., H. Fu and Y. Jia, 2008. Gaussian mixture modeling and learning of neighboring characters for multilingual text extraction in images. *Pattern Recogn.*, 41(2): 484-493.

Nguyen, T.M. and Q.M.J. Wu, 2012. Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem. *IEEE T. Syst. Man Cy. B*, 42(1): 193-202.

Redner, R. and H. Walker, 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.*, 26(2): 195-239.

Tang, H., J.L. Dillenseger, X.D. Bao and L.M. Luo, 2009. A vectorial image soft segmentation method based on neighborhood weighted Gaussian mixture model. *Comput. Med. Imag. Grap.*, 33(8): 644-650.

Tao, W., H. Jin and Y. Zhang, 2007. Color image segmentation based on mean shift and normalized cuts. *IEEE T. Syst. Man Cy. B*, 37(5): 1382-1389.

Xu, L., 1995. Bayesian-Kullback coupled Ying-Yang machines: Unified learning and new results on vector quantization. *Proceeding of the International Conference on Neural Information Processing*.

Xu, L., 1997. Bayesian Ying-Yang machine, clustering and number of clusters. *Pattern Recogn. Lett.*, 18(11-13): 1167-1178.

Xu, L., 1998. RBF nets, mixture experts and bayesian Ying-Yang learning. *Neurocomputing*, 19(1-3): 223-257.

Xu, L., 2004. Advances on BYY harmony learning: Information theoretic perspective, generalized projection geometry and independent factor autodetermination. *IEEE T. Neural Networ.*, 15: 885-902.

Xu, L., A. Krzyzak and E. Oja, 1993. Rival penalized competitive learning for clustering analysis, RBF net and curve detection. *IEEE T. Neural Networ.*, 4(4): 636-649.

Zoltan, K. and T.C. Pong, 2006. A Markov random field image segmentation model for color textured images. *Image Vision Comput.*, 24(10): 1103-1114.