ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 17, No. 2, 2022, *pp*.118-137

A skew Normal Mixture Model with Noise Estimation for image segmentation

Kaili Zhang[1]
School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China
(Received October 07, 2022, accepted November 21, 2022)

Abstract: Accurate image segmentation is an essential step in image processing. Gaussian mixture model (GMM) has been widely used for image segmentation due to its low complexity and high accuracy. However, the model assumes that the intensity distributions of images are symmetric, which makes it hard to obtain ideal results for images with asymmetric distributions. In addition, the model does not consider any noise, which makes it difficult to obtain ideal distribution fitting results when the image contains severe noises. Furthermore, the model only considers the distribution information without any spatial information, so it is sensitive to noise when segmenting images. To address these issues, we model noise with a Gaussian distribution and couple it into a skewed normal mixture model to reduce the effect of asymmetric distributions and noise and can obtain more accurate distribution fitting results. To further reduce the effect of noise, we propose a new anisotropic spatial information constraint term that preserves detailed information while reducing the effect of noise. Finally, an improved EM algorithm is proposed to solve the parameters of the model. Experimental results on synthetic and natural images show that our method achieves better segmentation results compared to other models.

Keywords: Skew normal distribution, Noise estimation, Anisotropic spatial information, Improved EM algorithm

1. Introduction

In most areas of digital image processing, image segmentation has a wide range of applications, such as industrial automation, production process control, online product inspection, image coding, document image processing, remote sensing and biomedical image analysis, security monitoring, as well as military, sports and other aspects. In the processing and analysis of medical images, image segmentation plays an effective role in guiding the three-dimensional display of diseased organs in people's bodies or in determining and analyzing the location of lesions; In the analysis and application of road traffic conditions, image segmentation technology can be used to separate the target vehicle to be extracted from the fuzzy and complex background such as monitoring or aerial photography; Remote sensing image segmentation is also widely used in military fields, such as strategic and tactical investigation, military marine mapping, etc. High-resolution remote sensing image segmentation data can be used for natural disaster monitoring and evaluation, map drawing and updating, forest resources and environment monitoring and management, agricultural product growth detection and yield estimation, urban and rural construction and planning, coastal area environmental monitoring The development of archaeological and tourism resources provides

[1] Corresponding Author *E-mail*:15063169817@163.com

detailed ground information. The segmentation of target houses and roads plays an indispensable role in urban construction and land planning. In the process of transforming data into information, the segmentation of remote sensing images plays a very important role

The methods of image segmentation are to separate the image into non-overlapping regions with the same properties. Image segmentation is a very important and difficult problem in many fields such as image processing and understanding, pattern recognition and artificial intelligence. It is a key step in computer vision technology, so accurate image segmentation is particularly important.

At present, many image segmentation methods have been proposed1, mainly including threshold based image segmentation algorithm, clustering based image segmentation algorithm, finite mixture model(FMM) and deep learning based image segmentation algorithm, etc. The finite mixture model is widely used because of its simple algorithm and small sample demand. We mainly study image segmentation based on finite mixture model. Gaussian mixture model(GMM) is widely used in image segmentation because of its simple algorithm and fast running speed. However, it also has many problems, for example, it only applies to symmetric data and has no good results for asymmetric data; Only the distribution information is considered, and the spatial information is not considered, so a good segmentation result cannot be obtained for the data with serious noise.

As the algorithm is sensitive to noise, many people propose to integrate spatial information, that is, Markov into the algorithm. Markov process is a kind of stochastic process, which was proposed by Russian mathematician A.A. Markov in 1907. We have selected the algorithm that uses Markov to improve spatial information in recent years, as follows: SCDMM6, FRSCGMM13, SCAGMM9, SCGAGMM10 and SCGAEM12. These five algorithms have achieved good segmentation accuracy, but in order to better reduce the noise interference, we model the noise on the model, which can estimate the variance of various noises, not limited to Gaussian white noise.

In addition to noise modeling, we also improve the spatial information. We propose anisotropic spatial information, which makes the algorithm show isotropy in homogeneous regions and anisotropy in edge and intersection regions during segmentation. Inspired by these articles, we propose a skew normal mixture model with noise estimation and spatial information constraints for image segmentation.

Our method is to first model the noise, then introduce the anisotropic spatial information, then integrate the three ideas into our overall model, and finally use the improved EM algorithm to estimate the parameters. The experimental results show that our method has the best segmentation accuracy and robustness.

The rest of this paper is organized as follows. In the second section, in view of our inspiration from some articles, we provided the theoretical basis for the feasibility of our method, introduced how to construct noise, and proposed and coupled anisotropic spatial information. Finally, we gave how to use the improved EM algorithm to estimate parameters. In the third section, we present the experimental results and compare them with other five advanced methods. In the fourth section, we summarize.

2. Proposed method

2.1. Finite mixture model

Let $x_i, i = (1,2, \dots, N)$ denote the target image, where x_i with dimension D is an observation at the i th pixel of the image. Let the neighborhood of the i th pixel be presented by ∂_i. Labels are denoted by $(\Omega_1, \Omega_2, \dots, \Omega_k)$. The finite mixture model (FMM) assumes that each observation x_i is independent and its probability density function(pdf) is defined as:

$$f(x_i|\Pi,\Theta)=\sum_{i=1}^{K}\pi_k p(x_i|\Theta_k) \tag{2.1}$$

where $\Pi=\{\pi_k\}, k=\{1,2,\dots,K\}$ is the prior probabilities for the pixel x_i, which satisfies $0\le\pi_k\le 1$, and $\sum_{k=1}^{K}\pi_k=1$. $p(x_i|\Theta_k)$ can be written any kind of distributions11. At present, the most widely used distribution is Gaussian distribution. Gaussian distribution $\phi(x_i|\Theta_k)$ can be written in the form

$$\phi(x_i|\Theta_k)=\frac{1}{(2\pi)^{D/2}|\Sigma_k|^{1/2}}\exp\{-\frac{1}{2}(x_i-\mu_k)^T\Sigma_k^{-1}(x_i-\mu_k)\} \tag{2.2}$$

where μ_k is the mean, Σ_k is the covariance matrix. Because Gaussian mixture model can not fit asymmetric data, we adopt skew normal distribution. The pdf of the skew normal distribution $\phi(x_i|\Theta_k)$ can be written by

$$\phi(x_i|\Theta_k)=\frac{1}{(2\pi)^{D/2}|\Gamma_k|^{1/2}}\exp\{-\frac{1}{2}\left(x_i-\mu_k-\Sigma_k^{1/2}\delta_k\tau_k\right)^T\Sigma_k^{-1}(x_i-\mu_k-\Sigma_k^{1/2}\delta_k\tau_k)\} \tag{2.3}$$

where $\delta_k=\frac{\lambda_k}{\sqrt{1+{\lambda_k}^T\lambda_k}}$ and λ is the skewness parameter. $\Gamma_k=\Sigma_k^{\frac{1}{2}}(I-\delta_k{\delta_k}^T)\Sigma_k^{\frac{1}{2}}$. When $\lambda=0$ it is the Gaussian distribution.

2.2. Spatial information on MRF

In many cases, there will inevitably be noise in the image, and the traditional finite mixture model is sensitive to noise. In recent years, many scholars have devoted themselves to the research of denoising in image segmentation. Some scholars have found that the correlation between adjacent pixels can be characterized by Markov random field (MRF), so that the information of adjacent pixels can be used to remove image noise. In order to improve the robustness to the noise, the MRF distribution incorporates the spatial relationship of x_i amongst its neighborhood region ∂_i. Based on Hamersley-Clifford theory, MRF can be expressed by Gibbs random field:

$$P(\mathbf{\Pi})=\frac{1}{Z}\exp\frac{1}{T}\{-U(\mathbf{\Pi})\} \tag{2.4}$$

where Z is a normalized constant, T is a temperature constant and $U(\mathbf{\Pi})$ is the smoothing prior. The posterior probability density function given by Bayes rules can be written as

$$P(\mathbf{\Pi},\mathbf{\Theta}\mid\mathbf{X})\propto P(\mathbf{X},\mathbf{\Pi}\mid\mathbf{\Theta})P(\mathbf{\Pi}) \tag{2.5}$$

Ji et al. 10 employ the Besag approximation for modeling the joint density over pixel priors:

$$P(\mathbf{\Pi})\approx\prod_{i=1}^{N}p\left(\pi_i\mid\pi_{\partial_i}\right) \tag{2.6}$$

where π_{∂_i} are defined as mixture distributions over the between-cluster priors of neighboring pixels of pixel i, i.e.,

$$\pi_{\partial_i} = \sum_{j\in\partial_i, i\neq j}\ \omega_{ij}\pi_j \tag{2.7}$$

where ω_{ij} is the positive weight of each $j(j \in \partial_i, j \neq i)$, which satisfies $\sum_j \omega_{ij} = 1$ and can be written as

$$\omega_{ij} = \frac{\exp\left(-\sum_{m\in\partial_i, m\neq i,j} |\mathbf{x}_j - \mathbf{x}_m| / \left(N_{\partial_i} \times \left(N_{\partial_i} - 1\right)\right)\right)}{\sum_{h\in\partial_i} \exp\left(-\sum_{m\in\partial_i, m\neq i,j} |\mathbf{x}_h - \mathbf{x}_m| / \left(N_{\partial_i} \times \left(N_{\partial_i} - 1\right)\right)\right)} \tag{2.8}$$

where N_{∂_i} is the number of pixels in ∂_i. For the expression of $p\left(\pi_i \mid \pi_{\partial_i}\right)$, the $\log\ p\left(\pi_i \mid \pi_{\partial_i}\right)$ is defined as:

$$\log\ p\left(\pi_i \mid \pi_{\partial_i}\right) = -\beta\left[K\left(\pi_i \mid \pi_{\partial_i}\right) + H(\pi_i)\right] \tag{2.9}$$

where $K\left(\pi_i \mid \pi_{\partial_i}\right) = \sum_{k=1}^{K} \pi_{ik} \log\ \pi_{ik} - \sum_{k=1}^{K} \pi_{ik} \log\ \pi_{\partial_i k}$ is the KL divergence between π_i and π_{∂_i}. $H(\pi_i) = -\sum_{k=1}^{K} \pi_{ik} \log\ \pi_{ik}$ is the entropy of the distribution π_i.

When the pixel j is quite differ from other pixels in ∂_i, the value of ω_{ij} becomes quite small and can effectively remove noise. However, when the object has slim structure, the factor is hard to preserve detail information.

Depending on the type of energy $U(\Pi)$ selected, we can have different kinds of models. The main motivation for using this model as opposed to the traditional MRF model on pixel labels is its flexibility with respect to the initial conditions, in which the spatial constraints are directly enforced over the neighboring priors to obtain a smoother energy function and make the algorithm less dependent on the initializations.

2.3. Skew normal mixture model with noise estimation

Since the general skewed normal model can only deal with Gaussian white noise sensitive to noise, we model noise. The model is defined as

$$y = x_i + \mu + \varepsilon_i\ ,\ i = 1,2, \ldots, N \tag{2.10}$$

where x_i follows $SN(0, \Sigma, \lambda)$ with mean μ, ε_i follows $N(0, \Lambda)$. The hierarchical representation can be written by

$$y_i | x_i \sim N(\mu + x_i, \Lambda) \tag{2.11}$$

$$x_i | \tau \sim N(\Delta t, \Gamma) \tag{2.12}$$

$$\tau \sim HN(0,1) \tag{2.13}$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^T\lambda}}$, $\Delta = \Sigma^{\frac{1}{2}}\delta\ and\ \Gamma = \Sigma^{\frac{1}{2}}(I - \delta\delta^T)\Sigma^{\frac{1}{2}}$. Then the likelihood function

of skew normal mixture model with noise estimation is expressed as:

$$L(\Theta|Y) = \sum_{i=1}^{N}\sum_{k=1}^{K}\left\{-\frac{1}{2}\log|\Lambda_k| - \frac{1}{2}(y_i - x_i - \mu_k)^T\Lambda_k^{-1}(y_i - x_i - \mu_k) - \frac{1}{2}\log|\Gamma_k| - \frac{1}{2}(x_i - \Delta_k t_i)^T\Gamma_k^{-1}(x_i - \Delta_k t_i) - \frac{t_i^2}{2}\right\} + C \tag{2.14}$$

where C is a constant, b is the bias field and others are consistent with the above.

2.4. Anisotropic spatially information

Since the model only considers the distribution information and does not consider the location information, we also introduce anisotropic spatial information into the model to reduce noise interference.

The structure tensor14 is proposed to describe the structure of objects. The traditional definition of structural tensor is $\mathbf{T}_X = \nabla I_X \nabla I_X^T$. Because the eigenvector of traditional structure tensor is based on gradient information, it is sensitive to noise. To solve this problem, we improve the traditional structure tensor and propose a structure tensor based on nonlocal information to minimize noise interference. Then it is described as follows:

$$D = f_1\left(\hat{\lambda}^1, \hat{\lambda}^2\right)\hat{\xi}^1\left(\hat{\xi}^1\right)^T + f_2\left(\hat{\lambda}^1, \hat{\lambda}^2\right)\hat{\xi}^2\left(\hat{\xi}^2\right)^T \tag{2.15}$$

with

$$f_1\left(\hat{\lambda}^1, \hat{\lambda}^2\right) = 1/\left(\gamma + \hat{\lambda}^1 + \hat{\lambda}^2\right)^2, f_2\left(\hat{\lambda}^1, \hat{\lambda}^2\right) = 1/\exp(\gamma)$$

$$\hat{\xi}^1 = \sum_{j=1}^{N_i} w_{ij}\xi_j^1, \hat{\xi}^2 \perp \hat{\xi}^1$$

$$\hat{\lambda}^1 = \sum_{j=1}^{N_i} w_{ij}\lambda_j^1, \hat{\lambda}^2 = \sum_{j=1}^{N_i} w_{ij}\lambda_j^2$$

$$w_{ij} = \exp\left(-\frac{\left\|N(\mathbf{x}_i) - N\left(\mathbf{x}_j\right)\right\|_F^2}{h}\right) \Big/ \sum_{k=1}^{|N_i|}\exp\left(-\frac{\|N(\mathbf{x}_i) - N(\mathbf{x}_k)\|_F^2}{h}\right)$$

where $\hat{\xi}^1$and $\hat{\xi}^2$ are the local maximum eigenvector variations and minimum eigenvector variations, $\hat{\lambda}^1$ and $\hat{\lambda}^2$ are the eigenvalues along $\hat{\xi}^1$and $\hat{\xi}^2$. $w_{ij} = \frac{G(x_i,x_j)}{\sum_{k=1}^{K}G(x_i,x_j)}$ and $\sum w_{ij} = 1$. $G(x_i, x_j)$ can be written as $G\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\left(X_i - X_j\right)^T D_i\left(X_i - X_j\right)\right) \times w_{ij}$ and D is the new structure tensor of X. γ and h are constants.

2.5. Anisotropic Skew normal mixture model with noise estimation

In order to coupling spatial information into the model, we introduce an approximation that makes use of a auxiliary set of s_i distribution:

$$\log\left(p\left(\pi_i \mid \pi_{\partial_i}, s_i\right)\right) = -\beta\left[K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i}\right) + H(s_i)\right] \tag{2.16}$$

So, the likelihood function is written as:

$$\begin{aligned} L(\mathbf{\Pi}, \mathbf{\Theta} \mid \mathbf{Y}) &= \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_{ik} p(\mathbf{y}_i \mid \boldsymbol{\theta}_k)\right) \\ &-\beta\left[K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i}\right) + H(s_i)\right] - \frac{1}{2}\left[K(q_i \mid z_i) + K\left(q_i \mid z_{\partial_i}\right) + H(q_i)\right] \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K}\left\{-\frac{1}{2}\log|\mathbf{\Lambda}_k| - \frac{1}{2}(\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k)^T \mathbf{\Lambda}_k^{-1}(\mathbf{y}_i - \mathbf{x}_i - \boldsymbol{\mu}_k)\right. \\ &\left. -\frac{1}{2}\log|\mathbf{\Gamma}_k| - \frac{1}{2}(\mathbf{x}_i - \mathbf{\Delta}_k t_i)^T \mathbf{\Gamma}_k^{-1}(\mathbf{x}_i - \mathbf{\Delta}_k t_i) - \frac{t_i^2}{2}\right\} \\ &-\beta\left[K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i}\right) + H(s_i)\right] - \frac{1}{2}\left[K(q_i \mid z_i) + K\left(q_i \mid z_{\partial_i}\right) + H(q_i)\right] \end{aligned}$$

where $K(a \mid \mathrm{b}) = \sum_{k=1}^{K} a \log \mathrm{a} - \sum_{k=1}^{K} a \log \mathrm{b}$, $H(\mathrm{c}) = -\sum_{k=1}^{K} c \log \mathrm{c}$. letting $\hat{x}_{i,k} = E\left[x_i \mid \mathbf{y}_i, \boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k\right]$, $\widehat{\Omega}_{i,k} = \mathrm{cov}\left[x_i \mid \mathbf{y}_i, \boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k\right]$, $\hat{t}_{i,k} = E[t_i \mid \mathbf{y}_i, \boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k]$, ${\hat{t}_{ik}}^2 = E[t_i \mid \mathbf{y}_i, \boldsymbol{\theta}_k = \widehat{\boldsymbol{\theta}}_k]$. According to the truncated normal moment, we can get:

$$\begin{aligned} \hat{t}_{i,k} &= \hat{\mu}_{t_{ik}} + W_{\Phi_1}\left(\frac{\hat{\mu}_{t_{ik}}}{\widehat{M}_{t_k}}\right)\widehat{M}_{t_k} \\ \hat{t}_{i,k}^2 &= \hat{\mu}_{t_{ik}}^2 + \widehat{M}_{t_k}^2 W_{\Phi_1}\left(\frac{\hat{\mu}_{t_{ik}}}{\widehat{M}_{t_k}}\right)\widehat{M}_{t_k}\hat{\mu}_{t_{ik}} \\ \hat{x}_{i,k} &= \hat{r}_{i,k} + \hat{s}_{i,k}\hat{t}_{i,k} \\ \widehat{\Omega}_{i,k} &= \hat{T}_{x_{ik}}^2 + \hat{s}_{i,k}\hat{s}_{i,k}^T\left(\hat{t}_{i,k}^2 - (\hat{t}_{ik})^2\right) \\ \hat{t}_{i,k}x_{i,k} &= \hat{r}_{i,k}\hat{t}_{i,k} + \hat{s}_{i,k}\hat{t}_{i,k}^2 \end{aligned}$$

where $\hat{M}_{T_i}^{\ 2} = \frac{1}{1+\hat{\Delta}^T\left(\hat{\Sigma}_i+\hat{\Gamma}\right)^{-1}\hat{\Delta}}$, $\hat{\mu}_{T_i} = \hat{M}_{T_i}^{\ 2}\hat{\Delta}^2\left(\hat{\Sigma}_i + \hat{\Gamma}\right)^{-1}(y_i - \mu_i)$, $\hat{T}_{b_i}^{\ 2} = \left(\hat{\Gamma}^{-1} + \Sigma_i^{-1}\right)^{-1}$, $\hat{r}_i = \hat{T}_{b_i}^{\ 2}\Sigma_i^{-1}(y_i - \mu_i)$, $\hat{s}_i = \left(I_n - \hat{T}_{b_i}^{\ 2}\hat{\Sigma}_i^{-1}\right)\hat{\Delta}$.

So, the Q function is written by

$$Q\left(\theta|\hat{\theta}\right) = E[L(\mathbf{\Pi}, \mathbf{\Theta} \mid \mathbf{Y})]$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K}\pi_{ik}\left\{-\frac{1}{2}log|\Lambda_k| - \frac{1}{2}\left(y_i-\mu_k-\hat{x}_i\right)^T\Lambda_k^{-1}\left(y_i-\mu_k-\hat{x}_i\right) - \frac{1}{2}tr\left(\Lambda_k^{-1}\hat{\Omega}_{ik}\right) - \frac{1}{2}log|\Gamma_k|\right.$$

$$\left. - \frac{1}{2}tr\left(\Gamma_k^{-1}\left(\hat{\Omega}_{ik} + \hat{x}_{ik}\hat{x}_{ik}^T - 2t_{ik}\hat{x}_{ik}\Delta_k^T + \hat{t}_i^{\,2}\Delta_k\Delta_k^T\right)\right)\right\}$$

$$- \beta\left[K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i}\right) + H(s_i)\right] - \frac{1}{2}\left[K(q_i \mid z_i) + K\left(q_i \mid z_{\partial_i}\right) + H(q_i)\right]$$

E-step: $Z_{ik} = \frac{\pi_{ik}p(y|\theta)}{\sum_{k=1}^{K}\pi_{ik}p(y|\theta)}$

M-step: we can maximize Q as follows:

$$\hat{\pi}_{ik} = \frac{1}{1+2\beta}\left(\frac{1}{2}\left(q_{ik}+q_{\partial_i k}\right) + \beta\left(s_{ik}+s_{\partial_{ik}}\right)\right)$$

$$\widehat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)\left(y_i-\hat{x}_{i,k}\right)}{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)}$$

$$\widehat{\boldsymbol{\Delta}}_k = \frac{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)\hat{t}_{i,k}x_{i,k}}{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)\hat{t}_{i,k}^2}$$

$$\widehat{\boldsymbol{\Lambda}}_k = \frac{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)\left(\left(y_i-\mu_k-\hat{x}_{i,k}\right)\left(y_i-\mu_k-\hat{x}_{i,k}\right)^T + \widehat{\Omega}_{i,k}\right.}{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)}$$

$$\widehat{\boldsymbol{\Gamma}_k} = \frac{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)\left(\widehat{\Omega}_{i,k} + \hat{x}_{i,k}\hat{x}_{i,k}^T - 2\widehat{t_{\iota,k}}x_{i,k}\Delta_k + \hat{t}_{i,k}^2\Delta_k\Delta_k^T\right)}{\sum_{i=1}^{N}\left(q_{ik}+q_{\partial_i k}\right)}$$

Where $\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k + \boldsymbol{\Delta}_k\boldsymbol{\Delta}_k^T$, $\boldsymbol{\lambda}_k = \frac{\boldsymbol{\Sigma}_k^{-1/2}\boldsymbol{\Delta}_k}{\sqrt{1-\boldsymbol{\Delta}_k^T\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Delta}_k}}$.

Algorithm: Anisotropic Skew normal mixture model with noise estimation

Step1: Initialize the parameter $\boldsymbol{\Theta}$ using k-means.

Step2: Calculate the weight factor $\boldsymbol{w_{ij}}$.

Step3: E-step by using

Step4: M-step by using

Step5: Check for the convergence of either the objective function, or the parameter values. If the convergence criterion is satisfied, stop the iteration; otherwise, go to Step 3.

3. Experiment Results

In this paper, we compared our method with the other five most advanced methods, such as SCDMM, FRSCGMM, SCAGMM, SCGAGMM and SCGAEM. It is worth mentioning that in the corresponding papers of these five comparison methods, the authors have provided a wealth of comparative experimental results to prove that their methods are superior to most classical algorithms, including GMM15, SMM16, GGMM17 and SPM18. Therefore, in this paper, we only compare with SCDMM, FRSCGMM, SCAGMM, SCGAGMM and SCGAEM to prove the superior performance of the proposed algorithm.

In this section, our first experiment is to test the composite image. The experimental results are shown in Figure 1. From the experiment, we can see that our method has obtained the best performance results. Our method has obvious denoising effect and better preserves corner information and boundary information. Therefore, our method can obtain good performance in synthetic images.

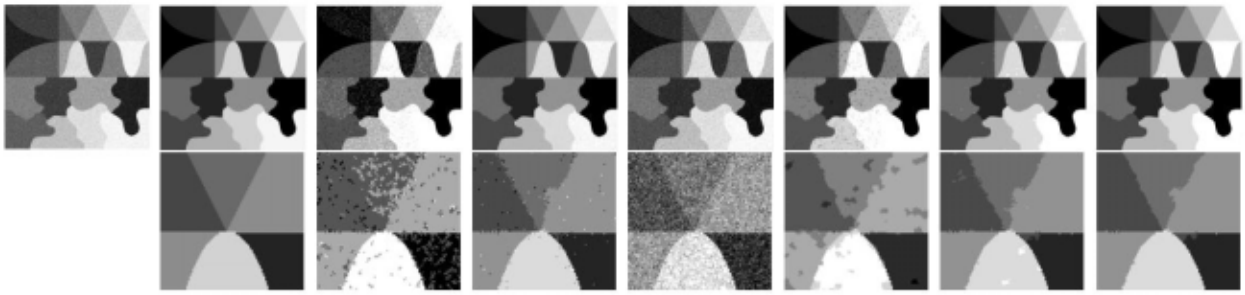

Figure 1: The segmentation results on synthetic images. The first column shows the initial images. The second to last column show the ground truth, the segmentation results of SCDMM, FRSCGMM, SCAGMM, SCGAGMM and SCGAEM and our method, respectively.

In order to prove the robustness of our method, we will conduct experiments on the virtual brain map. The data used in this paper are as follows: the virtual brain images that we used are from Brain Web (http://www.bic.mni.mcgill.ca/brainweb/) and Internet Brain Segmentation Repository(IBSR) . The parameters we set are as follows: h = 0.3, β = 0.4, γ = 1.3. The JS19 is used to quantify the accuracy of brain MR image, which is defined as:

$$JS(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{3.1}$$

Where S_1 and S_2 are segmentation results and labels, respectively. The JS value is [0,1]. Therefore, the higher the JS value is, the better the segmentation result we will obtain. We conducted experiments on 200 brain images, and the experimental results of the JS values are shown in Table 1.

As shown in Table 1, we conducted experiments without three different noises in the offset field. From the experimental results, we can see that our method has achieved good segmentation accuracy. Although the CSF accuracy of SCAGMM is higher than ours at N3F0, with the increase of noise level, the stability of our method gradually shows up, and SCAGMM is greatly affected by noise. Therefore, considering all the conditions, our method has achieved the best results.

In order to prove the superiority of our model, we also carried out experiments on the image with offset field. Because our model can also get a good fitting effect on the data of asymmetric distribution, our method can also improve the brain map with offset field unnecessarily. The experimental results are shown in Table 2:

Table 1: Average JS values of 200 virtual brain images with different noise

		SCDMM	FRSCGMM	SCAGMM	SCGAGMM	SCGAEM	Our proposed
N3F0	WM	81.93	88.46	89.20	84.82	87.83	**89.79**
	GM	76.97	87.13	87.84	78.25	86.05	**87.92**
	CSF	74.40	84.68	**86.10**	59.51	82.85	84.96
N5F0	WM	81.72	82.97	80.87	83.09	83.14	**85.69**
	GM	78.86	81.42	80.11	76.92	81.57	**83.88**
	CSF	73.75	78.74	80.68	58.71	79.24	**81.77**
N7F0	WM	80.68	77.71	71.95	78.98	78.68	**82.50**
	GM	78.22	76.32	71.51	76.26	76.77	**80.36**
	CSF	71.65	72.03	74.10	57.52	74.28	**78.42**

Table 2: Average JS values of 200 virtual brain images

		SCDMM	FRSCGMM	SCAGMM	SCGAGMM	SCGAEM	Our proposed
N3F40	WM	80.41090606	80.41090606	85.53863584	83.96339294	84.22046011	**87.37266891**
	GM	75.16552033	75.16552033	83.20568389	76.20050661	80.9577525	**84.43328313**
	CSF	73.64107112	73.64107112	**84.84778505**	58.37108612	81.16930317	83.14093842
N3F80	WM	72.13752077	71.92120214	74.64613541	72.68516795	73.60929157	**86.5802139**
	GM	69.24323022	72.38326982	74.45114369	70.83142757	70.63662773	**84.16833065**
	CSF	71.20878421	77.98925189	80.36417727	62.27985571	73.72360875	**83.53579382**
N3F100	WM	68.89410164	68.87784488	71.0531316	69.10978737	70.01211727	**86.63171284**
	GM	65.31373827	68.2656514	70.44113454	69.06320698	65.14550438	**84.17903553**
	CSF	66.35255802	72.58937826	75.72129046	63.15685939	66.3126772	**83.45200705**

As shown by the experimental results in Table 2, although the average accuracy of the CSF of SCAGMM in N3F40 is higher than our experimental results, the robustness of SCAGMM becomes worse with the increase of the bias field, and the experimental accuracy decreases greatly. Our method has the best robustness, and the bias field enhancement even has a tendency to increase CSF. Therefore, our method has the best robustness, and in general, our method obtains the best segmentation accuracy.

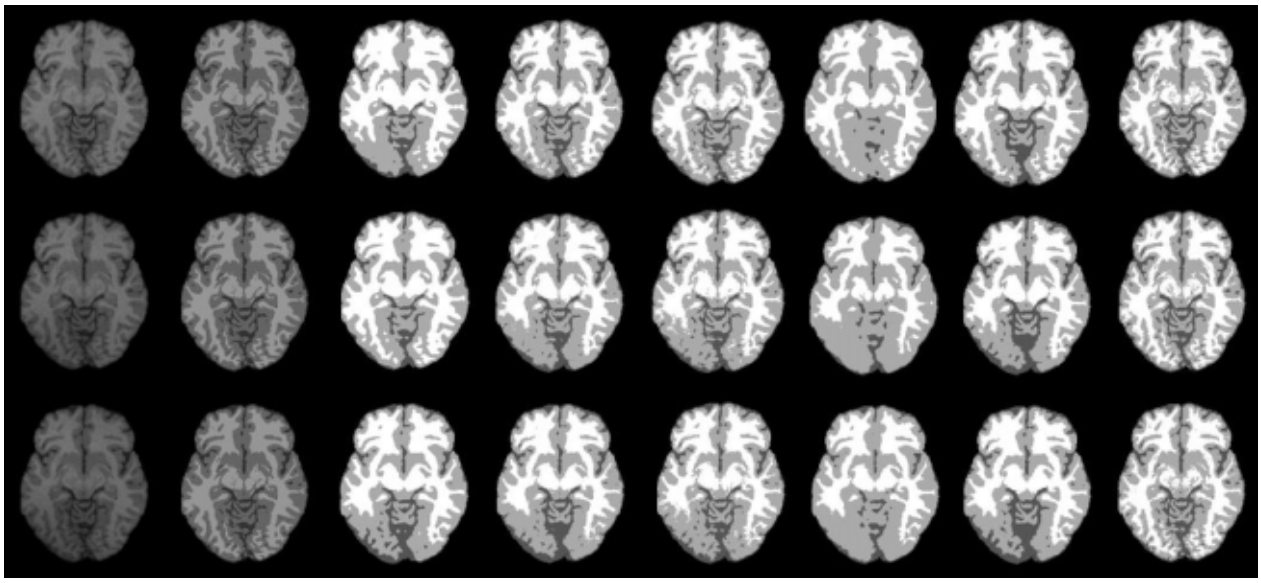

Figure 2: The segmentation results on simulated MR brain images. The first column shows the initial images with noise level 3% and bias filed level 40, 80 and 100, respectively. The second to last column show the ground truth, the segmentation results of SCDMM, FRSCGMM, SCAGMM, SCGAGMM and SCGAEM and our method, respectively.

As shown in Figure 2, we show the segmentation results of three different bias fields in one of the virtual brain images. From the visual effect, we can clearly see that our method has the best segmentation results. Our method retains more details, and with the increase of bias field strength, our method does not have too much deviation, which indicates that our method is more robust, and we can also draw this conclusion from the JS value in Table 1.

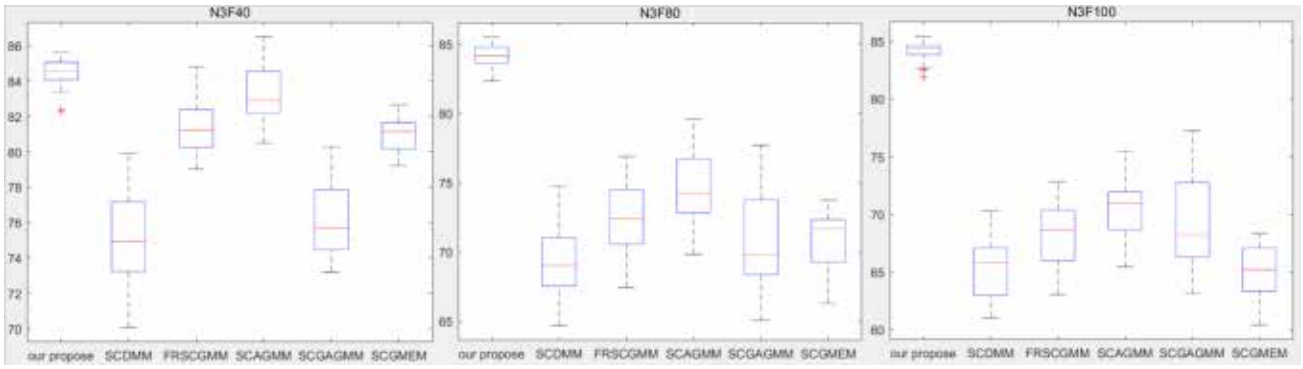

Figure 3: The boxplots of the GM segmentation results on simulated MR brain images. The picture shows boxplots with noise level 3% and bias filed level 40, 80 and 100, respectively.

It can be seen from the box graph in Figure 3 that our method obtains the best segmentation result, and our method has the lowest variance, so it has the highest stability. We not only conducted experiments on virtual maps, but also on real brain maps. The segmentation result of the real brain map is shown in Figure 3.

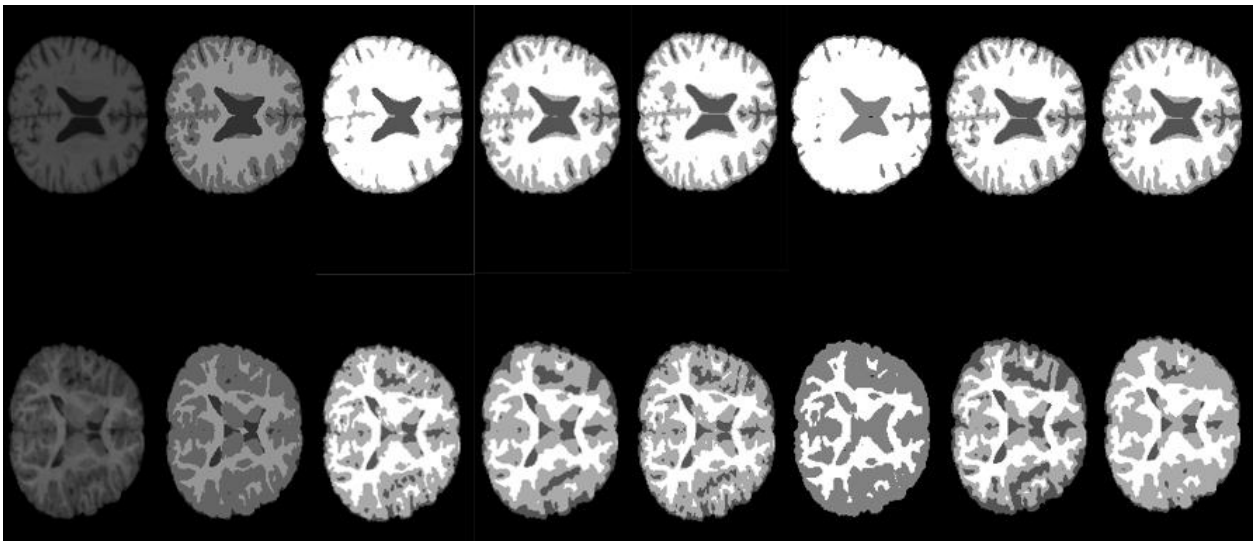

Figure 4: The segmentation results on real MR brain images. The first column shows the initial images. The second to last column show the ground truth, the segmentation results of SCDMM, FRSCGMM, SCAGMM, SCGAGMM and SCGAEM and our method, respectively.

In order to prove the robustness of our method, we will conduct experiments on real brain maps. From Figure 4, we can clearly see that our method retains more details, and we can also get good segmentation results for low contrast images. A large number of experimental data show that our method has better segmentation accuracy and higher stability in both virtual and real brain images.

Because there are complex classes in natural images, natural image segmentation is a big problem in image segmentation. As mentioned in 20, we introduce MMIC to determine the number of classes of natural images. The definition of MMIC is as follows

$$MMIC(K) = 2\log p_{MLE}(Y \mid K) - d_K \log(N) \tag{3.2}$$

where K is the number of classes in an image, $\log p_{MLE}(Y \mid K)$ is estimated by $\sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} P_k \, \Phi(y_i \mid \hat{\theta}_k)\right)$, y_i is the i th pixels in image, $\Phi(\cdot)$ is the pdf of Gaussian distribution. P_k denotes the mixture proportion of the k th component. $\hat{\theta}_k$ is the op-timal parameter by maximize the loglikelihood function $\sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} P_k \, \Phi(y_i \mid \theta_k)\right)$, d_k is number of parameters in the mixture model with K components $(\dim(\theta_K))$, in Gaussian mixture model, the value is $2 * K$. The larger values of MMIC (K) represent the better segmentation.

As shown in Table 3, we listed the MMIC values of the corresponding categories of the following test images. We can clearly see that the maximum values of MMIC of images test1, test2, test3, test4, test5, and test6 appear at K=3, 3, 6, 3, 2, 2, respectively. When K is equal to 7 or 8, test4 gets a larger value, but in view of the small difference and K is too large, we take the K of test4 as 3.

Table3: The values of MMIC of the different K

	K=2	K=3	K=4	K=5	K=6	K=7	K=8
test1	-210265.69	**-160766.25**	-184234.59	-206678.33	-199802.34	-210556.47	-217702.63
test2	-200173.28	**-145404.35**	-154875.32	-419046.69	-365669.91	-390747.92	-395082.81
test3	-437013.35	-495592.07	-470183.16	-414504.74	**-355404.05**	-474000.92	-414836.90
test4	-802320.32	**-235796.21**	-291589.19	-291458.25	-311475.41	-226339.18	-227921.21
test5	**-70317**	-167835	-292377	-233141	-92179.5	-142873	-207247
test6	**-610675.13**	-708297.44	-703198.18	-703039.70	-703340.85	-704084.31	-701885.02

Natural images are different from the evaluation criteria of brain images. As mentioned in 21, natural images use PRI to evaluate the segmentation results. The definition of PRI is as follows:

$$PRI(S,S') = \frac{2}{N(N-1)}\sum_{i,j,i<j}\left[p_{ij}I(l_i = l_j) + (1-p_{ij})I(l_i \neq l_j)\right] \tag{3.3}$$

where $I(\cdot)$ is an indicate function, the value of it is 0 or 1. $p_{ij} = \frac{1}{K}\sum_{m=1}^{K} I\left(l_i^{S_{2K}} = l_j^{S_{2K}}\right)$, S_1 is the experimental result, S_2 is the m-th ground truth in $S_2 = \{S_{21}, S_{22}, \dots, S_{2K}\}$. $l_i^{S_1}$ and $l_i^{S_2}$ denote the class label of pixel i for segmentation result and ground truth. The value of PRI is between 0 to 1.

Figure 5: Image segmentation results on natural images (test1). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SCDMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

From Figure 5, we can see that our method achieves the best segmentation results. Test1 is a picture with a normal distribution of data and a close background to the target. It mainly tests whether the algorithm can accurately segment the target and the background. SCAGMM does not completely distinguish the target from the background. Although SCDMM, FRSCGMM and SCGM-EM separate the target from the background, the target texture is not separated. SAGAGMM not only does not separate the background and

objectives, but also confuses the objectives with others. Our method can clearly separate the background and target and retain more details of the leopard.

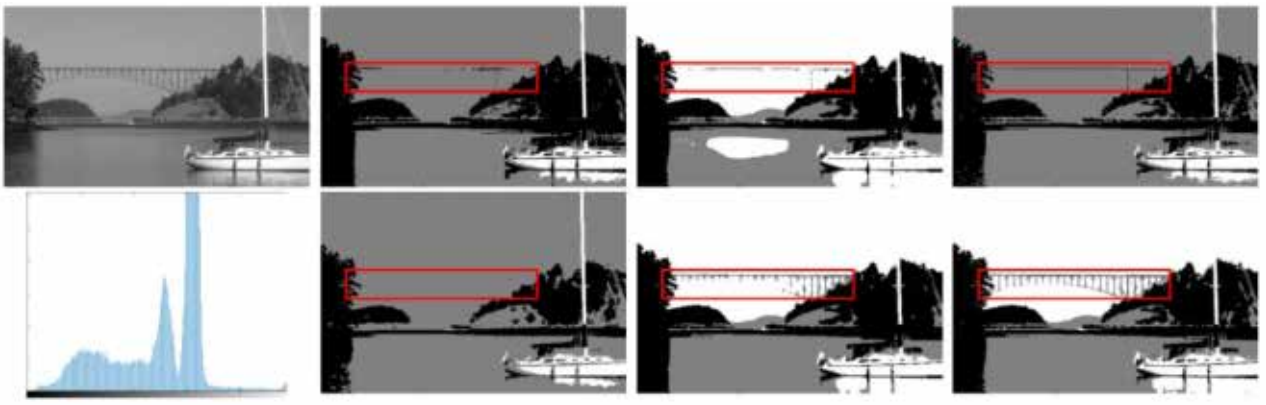

Figure 6: Image segmentation results on natural images (test2). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SC-DMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

From Figure 6, we can see that our method achieves the best segmentation results. Test2 is also a picture of data showing normal distribution. All the methods can accurately segment the ship and the mountain, but FRSCGMM classifies the reflective part of the water as a separate category, and does not accurately identify it. SCDMM, SCAGMM, and SCGAGMM did not accurately segment the bridge. Although SCGM-EM segmented the bridge, it did not completely retain the details of the bridge, and planned some parts of the bridge as the sky. Obviously, our method retains the information of the bridge completely. However, all methods do not retain the reflection of the bridge in the water, which will be an urgent problem to be solved in the future.

Figure 7: Image segmentation results on natural images (test3). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SC-DMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

From Figure 7, we can see that our method achieves the best segmentation results. From the histogram of test3, we can clearly see that this is a 6 classification problem with mixed distribution of data. When both symmetric and asymmetric distributions appear in the data, our method can still maintain stability. In addition to our method, the other five methods do not have a good segmentation of the mountain level. Our

method perfectly separates the tree from the mountain, and the mountain trend remains the most complete, so our method has the best segmentation performance.

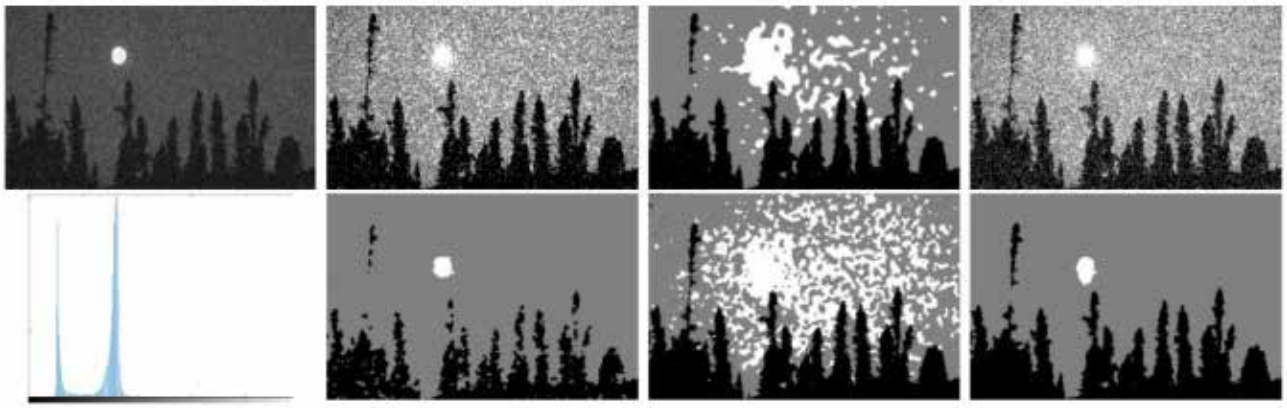

Figure 8: Image segmentation results on natural images (test4). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SC-DMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

From Figure 8, we can see that our method achieves the best segmentation results. Although the histogram of test4 shows two categories, our experience and the results of MMIC show that we should divide it into three categories: sky, tree and moon. We added noise to highlight the advantages of our method. From the figure, we can clearly see that SCDMM and SCAGMM are significantly affected by noise, while FRSCGMM and SCGM-EM do not separate the moon. Although SCGAGMM can clearly divide the three categories, it divides part of the tree information into the sky. Our method not only removes noise, but also separates the tree, sky and moon, and retains the most detailed information. Therefore, our method still shows high stability in the case of high noise in the image.

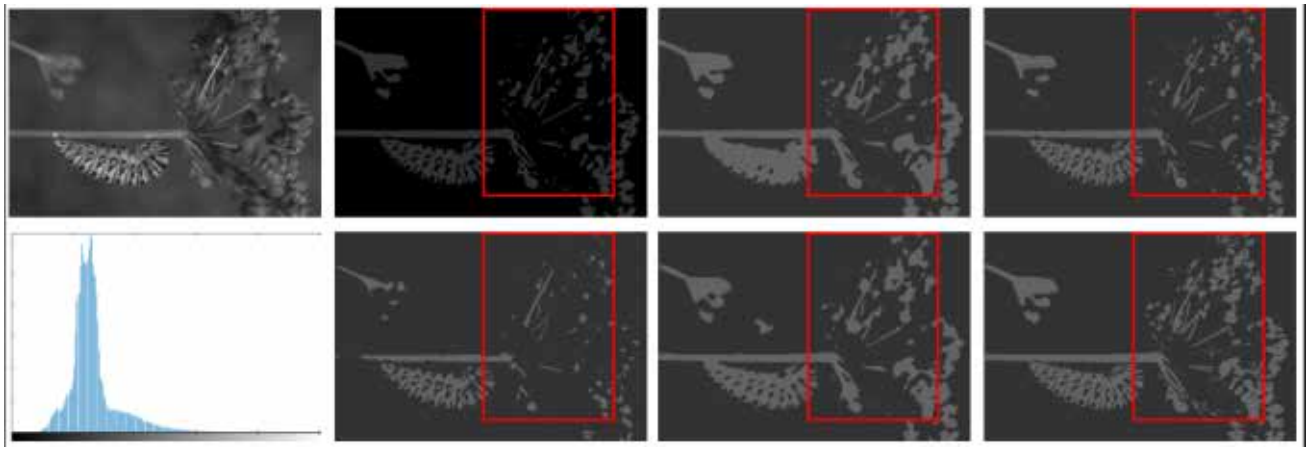

Figure 9: Image segmentation results on natural images (test5). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SC-DMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

From Figure 9, we can see that our method retains the most detailed information. It is difficult to determine the number of K from the histogram of test5, but MMIC can know that K is 2. All methods can roughly distinguish the target from the background, but the retention of details is quite different. SAGAGMM has not completely reserved the stamen part in the upper left corner, and many information is

planned as the background part. In addition, the other five methods have no good retention for the target stamen stem. Most of them divide the flower stem as the background, while our method retains the most details. From the visual point of view, our method has obtained the best segmentation results.

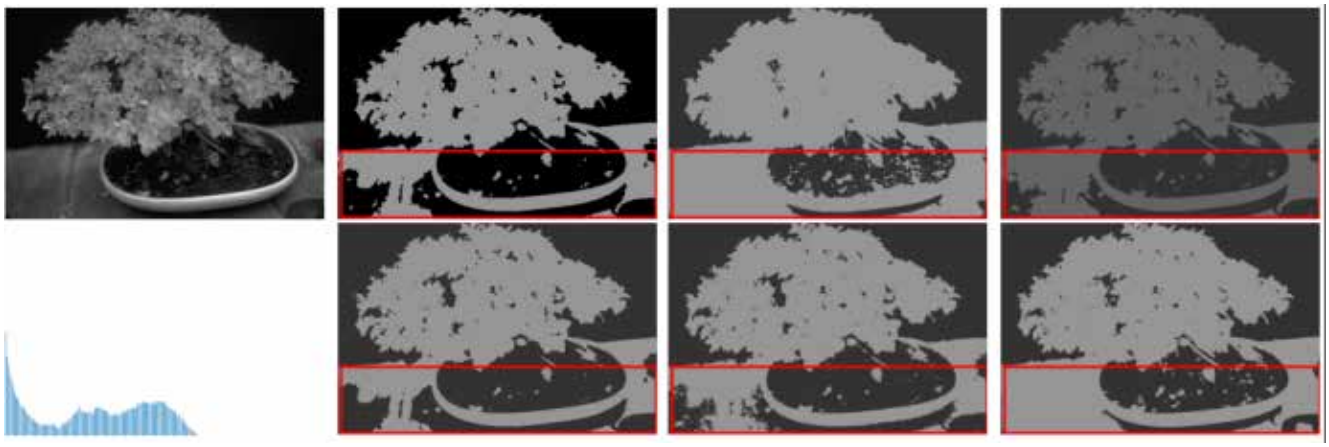

Figure 10: Image segmentation results on natural images (test6). The first column shows the initial image and its histogram. The second to last column of the first row show the segmentation results of SC-DMM, FRSCGMM and SCAGMM, respectively. The second to last column of the second row show the segmentation results of SCGAGMM and SCGA-EM and our method, respectively.

According to the MMIC value, we set the number of K as 2 in Figure 10. It can be seen from the figure that the other five methods do not clearly distinguish the flowerpot from the table. SCDMM, SCAGMM, SCGAGMM and SCGM-EM all divide the highlighted part of the table into another category, while our method completely preserves the table part. FRSCGMM divides a part of the flowerpot into tables, and does not completely retain the information of the flowerpot, so our method obtains the best segmentation results.

We conducted experiments on 70 natural images, and the PRI values of the experiments are shown in Table 4:

Table 4 : The values of PRI with 70 natural images

	SCDMM	FRSCGMM	SCAGMM	SCGAGMM	SCGM-EM	Our proposed
PRI	0.714±0.043	0.740±0.028	0.721±0.027	0.677±0.039	0.736±0.019	0.759±0.011

From Table 4, we can see that our method also has the best segmentation accuracy on natural images, and our method has the smallest variance, so our method has the highest stability.

On the other hand, the objective function of our method is non convex. Therefore, our method may fall into local optimization. The global optimization method can be used to find the global optimum. In addition, the method is based on light tail distribution, which makes it cannot obtain accurate results on images with heavy tail form. Our future work will focus on improving the robustness of initialization and selecting more robust statistical distributions

4. Conclusion

In recent years, GMM has been widely used in image segmentation. Aiming at some shortcomings of GMM, we propose a skewed normal model with anisotropic spatial information and noise estimation. Our method can fit the distributed data rather than the symmetric data. Moreover, our method has stronger anti-

interference ability to noise and can obtain more accurate segmentation results. Experimental results show that our method achieves the best segmentation accuracy.

However, our method still has some problems. For example, we can't fit the heavy tailed data very well. In the future research work, we will introduce the student T distribution into our model. In future work, we will also try to combine the model with deep learning to get better segmentation results.

Appendix

Derivation of the proposed method , the complete log-likelihood function Q has the form:

$$\begin{aligned} & Q(\boldsymbol{\Theta} \mid \widehat{\boldsymbol{\Theta}}) = E[L(\boldsymbol{\Pi}, \boldsymbol{\Theta} \mid \mathbf{X}) \mid \mathbf{X}, \widehat{\boldsymbol{\Theta}}] \\ = \quad & \sum_{i=1}^{N} \left\{ \log \sum_{k=1}^{K} \left\{ \pi_{ik} \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Gamma}_k|}} \exp \left\{ -\frac{1}{2} \left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k} \right)^T \boldsymbol{\Gamma}_k^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k} \right) \right. \right. \right. \\ & \left. \left. \left. -\frac{1}{2} \left(\hat{s}_{i,k}^2 - \left(\hat{s}_{i,k} \right)^2 \right) \boldsymbol{\Delta}_k^T \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\Delta}_k - \frac{1}{2} \hat{s}_{i,k}^2 \right\} \right\} \right\} \\ & -\beta \left[K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i} \right) + H(s_i) \right] \\ & -\frac{1}{2} \left[K(q_i \mid z_i) + K\left(q_i \mid z_{\partial_i} \right) + H(q_i) \right] \end{aligned} \tag{4.1}$$

Then we utilize EM algorithm to maximize the energy Q, In the E-step, we fix Θ and Π to maximize Q over s and q. In the M-step, we fix s and q to maximize Q over Θ and Π.

E-step: By fixing $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$, we can optimize over s_i. The terms involving s_i in Q are:

$$\begin{aligned} & K(s_i \mid \pi_i) + K\left(s_i \mid \pi_{\partial_i} \right) + H(s_i) \\ = \quad & -\sum_{k=1}^{K} s_{ik} \log s_{ik} + \sum_{k=1}^{K} s_{ik} \log \pi_{ik} - \sum_{k=1}^{K} s_{ik} \log s_{ik} + \sum_{k=1}^{K} s_{ik} \log \pi_{\partial_i k} + \sum_{k=1}^{K} s_{ik} \log s_{ik} \\ = \quad & -\sum_{k=1}^{K} s_{ik} \log s_{ik} + \sum_{k=1}^{K} s_{ik} \log \left(\pi_{ik} \pi_{\partial_i k} \right) \end{aligned} \tag{4.2}$$

The above formula is a negative KL-divergence which becomes zero when:

$$s_i \propto \pi_{ik} \pi_{\partial_i k} \tag{4.3}$$

By applying the similar derivation holds for q_i, we can get

$$q_i \propto z_{ik} z_{\partial_i k} \tag{4.4}$$

Therefore, we can get the updating functions for s_i and q_i in our method.

M-step: By fixing s and q, we can maximize Q over $\boldsymbol{\Theta}$ and Π. The terms involving Π and Θ are:

$$\begin{aligned} & \sum_{i=1}^{N} \left\{ \log \sum_{k=1}^{K} \left\{ \pi_{ik} p(\mathbf{x}_i \mid \boldsymbol{\Theta}) \right\} \right\} \\ & -\beta \left[K(s_i \mid \pi_i) + \sum_{j \in \partial_i} K\left(s_j \mid \pi_{\partial_j} \right) \right] - \frac{1}{2} \left[K(q_i \mid z_i) + \sum_{j \in \partial_i} K\left(q_j \mid z_{\partial_j} \right) \right] \end{aligned} \tag{4.5}$$

First of all, let us consider the derivation over z_i, then the terms involving only z_i

are:

$$\begin{aligned} & -\frac{1}{2}\left[K(q_i \mid z_i) + \sum_{j\in\partial_i} K\left(q_j \mid z_{\partial_j}\right)\right] \\ = \quad & -\frac{1}{2}\left[\sum_{k=1}^{K} q_{ik}\log q_{ik} - \sum_{k=1}^{K} q_{ik}\log z_{ik} + \sum_{j\in\partial_i, j\neq i}\left(\sum_{k=1}^{K} q_{ik}\log q_{jk} - \sum_{k=1}^{K} q_{ik}\log z_{\partial_{jk}}\right)\right] \end{aligned} \tag{4.6}$$

By ignoring the terms independent of z_{ik}, we get

$$-\frac{1}{2}\left[-\sum_{k=1}^{K} q_{ik}\log z_{ik} - \sum_{j\in\partial_i, j\neq i}\sum_{k=1}^{K} q_{jk}\log z_{\partial_{jk}}\right]$$

Where

$$z_{\partial_j} = \sum_{m\in\partial_j, m\neq j} \alpha_{jm} z_i = \alpha_{jm} z_i + \sum_{m\in\partial_j, m\neq i,j} \alpha_{jm} z_m \tag{4.7}$$

To make the M-step tractable, we using Jensens inequality to bound terms:

$$\log z_{\partial_{jk}} = \log \sum_{m\in\partial_j, m\neq j} \alpha_{jm} z_{mk} \geq \alpha_{jm}\log z_{ik} + \log \sum_{m\in\partial_j, m\neq i,j} \alpha_{jm} z_m \tag{4.8}$$

Since $\alpha_{ji} = \alpha_{ij}$, we obtain:

$$\begin{aligned} & \frac{1}{2}\left[-\sum_{k=1}^{K} q_{ik}\log z_{ik} - \sum_{j\in\partial_i, j\neq i}\sum_{k=1}^{K} q_{jk}\log z_{\partial_{jk}}\right] \\ = \quad & \frac{1}{2}\left[-\sum_{k=1}^{K} q_{ik}\log z_{ik} - \sum_{j\in\partial_i, j\neq i}\sum_{k=1}^{K} q_{jk}\log\left(\sum_{m\in\partial_j, m\neq j} \alpha_{jm} z_{mk}\right)\right] \\ \geq \quad & \frac{1}{2}\left[\sum_{k=1}^{K} q_{ik}\log z_{ik} + \sum_{j\in\partial_i, j\neq i}\left(\sum_{k=1}^{K} q_{jk}\left(\alpha_{ji}\log z_{ik}\right) + \sum_{m\in\partial_j, m\neq j} \alpha_{jm}\log z_{mk}\right)\right] \\ = \quad & \frac{1}{2}\left[\sum_{k=1}^{K} q_{ik}\log z_{ik} + \sum_{k=1}^{K}\sum_{j\in\partial_i, j\neq i} q_{jk}\left(\alpha_{ji}\log z_{ik}\right) + \sum_{j\in\partial_i, j\neq i}\sum_{m\in\partial_j, m\neq j} \alpha_{jm}\log z_{mk}\right] \end{aligned} \tag{4.9}$$

By only preserving the terms involving q_i, then the remaining terms are:

$$\begin{aligned} & \frac{1}{2}\left[-\sum_{k=1}^{K} q_{ik}\log z_{ik} - \sum_{k=1}^{K}\sum_{j\in\partial_i, j\neq i} q_{jk}\log q_{jk}\left(\alpha_{ji}\log z_{ik}\right)\right] \\ = \quad & \frac{1}{2}\left[-\sum_{k=1}^{K} q_{ik}\log z_{ik} - \sum_{k=1}^{K}\sum_{j\in\partial_i, j\neq i} \alpha_{ji} q_{jk}\log z_{ik}\right] \\ = \quad & \frac{1}{2}\left[\sum_{k=1}^{K}\left(q_{ik} + \sum_{j\in\partial_i, j\neq i} \alpha_{ji} q_{jk}\right)\log z_{ik}\right] \\ \Longrightarrow \quad & \frac{1}{2}\sum_{k=1}^{K}\left(q_{ik} + q_{\partial_i k}\right)\log z_{ik} \end{aligned} \tag{4.10}$$

Where the distribution q_{∂_i} is

$$q_{\partial_i} = \sum_{j\in\partial_i, j\neq i} \alpha_{ij} q_j \tag{4.11}$$

By applying the similar derivation holds for π_i, we can get:

$$\beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik}$$

Where the distribution s_{∂_i} is:

$$s_{\partial_i} = \sum_{j \in \partial_i, j \neq i} \alpha_{ij} s_j \tag{4.12}$$

Consequently, the lower bound of complete log-likelihood function Q involving the posterior z_i and prior π_i becomes:

$$\log \sum_{k=1}^{K} \{\pi_{ik} p(\mathbf{x}_i \mid \Theta)\} + \beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik} + \frac{1}{2} \sum_{k=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log\, z_{ik}$$

$\frac{1}{2}\sum_{k=1}^{K}\left(q_{ik} + q_{\partial_i k}\right) = \frac{1}{2}\sum_{k=1}^{K} q_{ik} + \frac{1}{2}\sum_{k=1}^{K} q_{\partial_i k} = 1$. By expanding the posterior z_{ik} , we find that maximizing

$$\begin{aligned}
& \log \textstyle\sum_{k=1}^{K} \{\pi_{ik} p(\mathbf{x}_i \mid \Theta)\} + \beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik} + \frac{1}{2} \sum_{k=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log\, z_{ik} \\
= \; & \log \textstyle\sum_{k=1}^{K} \{\pi_{ik} p(\mathbf{x}_i \mid \Theta)\} + \frac{1}{2} \sum_{k=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log\, z_{ik} + \beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik} \\
= \; & \log \textstyle\sum_{k=1}^{K} \{\pi_{ik} p(\mathbf{x}_i \mid \Theta)\} + \frac{1}{2} \sum_{k=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log \left\{\frac{\pi_{ik} p(\mathbf{x}_i|\Theta)}{\sum_{k=1}^{K} \pi_{ik} p(\mathbf{x}_i|\Theta)}\right\} + \beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik} \\
= \; & \frac{1}{2} \textstyle\sum_{k=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log \left(\pi_{ik} p(\mathbf{x}_i \mid \Theta)\right) + \beta \sum_{k=1}^{K} \left(s_{ik} + s_{\partial_i k}\right) \log\, \pi_{ik}
\end{aligned} \tag{4.13}$$

is equivalent to maximizing:

$$\frac{1}{2} \sum_{K=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log\, p(\mathbf{x}_i \mid \boldsymbol{\Theta}) + \sum_{K=1}^{K} \left(\frac{1}{2}\left(q_{ik} + q_{\partial_i k}\right) + \beta\left(s_{ik} + s_{\partial_i k}\right)\right) \log\, \pi_{ik}$$

Then, the Lagranges multiplier is used to enforce the constraint $\sum_{k=1}^{K} \pi_{ik} = 1$ for each data point, and we can easily get the updating function for the prior π_{ik} :

$$\hat{\pi}_{ik} = \frac{1}{1+2\beta}\left(\frac{1}{2}\left(q_{ik} + q_{\partial_i k}\right) + \beta\left(s_{ik} + s_{\partial_{ik}}\right)\right) \tag{4.14}$$

Similarly, we obtain the following update equations for $\boldsymbol{\mu}, \boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$, and . The energy function can be rewritten as:

$$\begin{aligned}
Q^* \quad & = \tfrac{1}{2}\textstyle\sum_{K=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log\, p(\mathbf{x}_i \mid \boldsymbol{\Theta}) + \sum_{K=1}^{K} \left(\frac{1}{2}\left(q_{ik} + q_{\partial_i k}\right) + \beta\left(s_{ik} + s_{\partial_i k}\right)\right) \log\, \pi_{ik} \\
& = \tfrac{1}{2}\textstyle\sum_{K=1}^{K} \left(q_{ik} + q_{\partial_i k}\right) \log \left\{\frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Gamma}_k|}} \mathrm{e\,x\,p} \left\{-\frac{1}{2}\left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k}\right)^T \boldsymbol{\Gamma}_k^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k}\right)\right.\right. \\
& \left.\left. -\frac{1}{2}\left(\hat{s}_{i,k}^2 - \left(\hat{s}_{i,k}\right)^2\right) \boldsymbol{\Delta}_k^T \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\Delta}_k - \frac{1}{2} \hat{s}_{i,k}^2 \right\}\right\} + \textstyle\sum_{k=1}^{K} \left(\frac{1}{2}\left(q_{ik} + q_{\partial_i k}\right) + \beta\left(s_{ik} + s_{\partial_i k}\right)\right)\Bigg) \log\, \pi_{ik}
\end{aligned} \tag{4.15}$$

Let $\frac{\partial Q^*}{\partial \boldsymbol{\mu}_k} = 0, \frac{\partial Q^*}{\partial \Gamma_k^{-1}} = 0$ and $\frac{\partial Q^*}{\partial \Delta_k} = 0$, we can obtain:

$$\widehat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i k}\right)\left(\mathbf{x}_i - \boldsymbol{\Delta}_k \hat{s}_{i,k}\right)}{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i k}\right)}$$

$$\widehat{\boldsymbol{\Delta}}_k = \frac{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i k}\right)\hat{s}_{i,k}(\mathbf{x}_i - \boldsymbol{\mu}_k)}{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i}\right)\hat{s}_{i,k}^2}$$

$$\hat{\boldsymbol{\Gamma}}_k = \frac{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i k}\right)\left(\left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k}\right)\left(\mathbf{x}_i - \boldsymbol{\mu}_k - \boldsymbol{\Delta}_k \hat{s}_{i,k}\right)^T + \left(\hat{s}_{i,k}^2 - (\hat{s}_{ik})^2\right)\boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^T\right)}{\sum_{i=1}^{N} \left(q_{ik} + q_{\partial_i k}\right)}$$

Where

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Gamma}_k + \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^T$$

$$\boldsymbol{\lambda}_k = \frac{\boldsymbol{\Sigma}_k^{-1} \boldsymbol{\Delta}_k}{\sqrt{1 - \boldsymbol{\Delta}_k^T \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\Delta}_k}}$$

References:

1. Maciel Zortea, Eliezer Flores, and Jacob Scharcanski. Fuzzy Multilevel Image Thresholding Based on Improved Coyote Optimization Algorithm IEEE Access, 99:1 – 1, 2021.
2. Kass, M. and Witkin, A. and Terzopoulos, D.. Snakes: Active contour models. IJCV, 1:321–331, 1988.
3. Fang, J. and Liu, H. and Liu, J. and Zhou, H. and Liu, H. Fuzzy region-based active contour driven by global and local fitting energy for image segmentation. Applied Soft Computing, 100:106982, 2021.
4. Siu Kai Choy, Shu Yan Lam, Kwok Wai Yu, Wing Yan Lee, and King Tai Leung. Fuzzy model-based clustering and its application in image segmentation. Pattern Recognition, 68:141 – 157, 2017.
5. Nabanita Mahata, Sayan Kahali, Sudip Kumar Adhikari, and Jamuna Kanta Sing. Local contextual information and gaussian function induced fuzzy clustering algorithm for brain mr image segmentation and intensity inhomogeneity estimation. Applied Soft Computing, 68:586 – 596, 2018.
6. Can Hu, Wentao Fan, Jixiang Du, and Yuchen Zeng. Model-based segmentation of image data using spatially constrained mixture models. Neurocomputing, 283:214 – 227, 2018.
7. Dusan Stosic, Darko Stosic, Teresa Bernarda Ludermir, and Tsang Ing Ren. Natural image segmentation with non-extensive mixture models. Journal of Visual Communication and Image Representation, 63:102598, 2019.
8. Sotirios Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation International Conference on Medical Image Computing and Computer-Assisted Intervention, 234– 241,2015
9. Zexuan Ji, Jinyao Liu, Hengdong Yuan, Yubo Huang, and Quansen Sun. A spatially constrained asymmetric gaussian mixture model for image segmentation. In Pacific-Rim Symposium on Image and Video Technology, 2015.
10. Zexuan Ji, Yubo Huang, Quansen Sun, and Guo Cao. A spatially constrained generative asymmetric gaussian mixture model for image segmentation. Journal of Visual Communication and Image Representation, 40:611 – 626, 2016.

11. Chi Liu, Hengchao Li, Kun Fu, Fan Zhang, Mihai Datcu, and William J Emery. Bayesian estimation of generalized gamma mixture model based on variational em algorithm. Pattern Recognition, 87:269–284, 2019.
12. A Diplaros, Nikos Vlassis, and Theo Gevers. A spatially constrained generative model and an em algorithm for image segmentation. IEEE Transactions on Neural Networks, 18(3):798–808, 2007.
13. Thanh Minh Nguyen and Q M Jonathan Wu. Fast and robust spatially constrained gaussian mixture model for image segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 23(4):621–635, 2013.
14. J. Weickert, B. M. T. H. Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Transactions on Image Processing, 7(3):398–410, 1998.
15. D.M. Titterington, A.F.M. Smith, U.E. Makov, Statistical Analysis of Finite Mixture Distributions, Wiley, Hoboken, NJ, 1985.
16. D. Peel, G. McLachlan, Robust mixture modeling using the t distribution, Stat. Comput. 10 (2000) 339–348.
17. M.S. Allili, D. Ziou, N. Bouguila, S. Boutemedjet, Image and video segmentation by combining unsupervised generalized Gaussian mixture modeling and feature selection, IEEE Trans. Circuits Syst. Video Technol. 20 (2010) 1373–1377.
18. J. Ashburner, K.J. Friston, Unified segmentation, Neuroimage 26 (2005) 839–851.
19. Yunjie Chen, Hui Zhang, Yuhui Zheng, Byeungwoo Jeon, and Q M Jonathan Wu. An improved anisotropic hierarchical fuzzy c-means method based on multivariate student t-distribution for brain mri segmentation. Pattern Recognition, 60:778–792, 2016.
20. Sotirios Stanford, D. C. and Raftery, A. E. Approximate Bayes Factors for Image Segmentation: The Pseudolikelihood Information Criterion (PLIC) IEEE Transactions on Pattern Analysis and Machine Intelligence., 24(11):1517-1520, 2002.
21. Sotirios Monteiro, Fernando C and Campilho, Aur´elio C. Performance Evaluation of Image Segmentation. Springer on International Conference Image Analysis and Recognition, 248–259, 2006.