RESEARCH ARTICLE

A Rough Set Bounded Spatially Constrained Asymmetric Gaussian Mixture Model for Image Segmentation

Zexuan Ji[1]*, Yubo Huang[1], Quansen Sun[1], Guo Cao[1], Yuhui Zheng[2]

1 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, China, 2 School of Computer and Software, Nanjing University of Information Science and technology, Nanjing, Jiangsu, China

* jizexuan@njust.edu.cn

Abstract

Accurate image segmentation is an important issue in image processing, where Gaussian mixture models play an important part and have been proven effective. However, most Gaussian mixture model (GMM) based methods suffer from one or more limitations, such as limited noise robustness, over-smoothness for segmentations, and lack of flexibility to fit data. In order to address these issues, in this paper, we propose a rough set bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation. First, based on our previous work where each cluster is characterized by three automatically determined rough-fuzzy regions, we partition the target image into three rough regions with two adaptively computed thresholds. Second, a new bounded indicator function is proposed to determine the bounded support regions of the observed data. The bounded indicator and posterior probability of a pixel that belongs to each sub-region is estimated with respect to the rough region where the pixel lies. Third, to further reduce over-smoothness for segmentations, two novel prior factors are proposed that incorporate the spatial information among neighborhood pixels, which are constructed based on the prior and posterior probabilities of the within- and between-clusters, and considers the spatial direction. We compare our algorithm to state-of-the-art segmentation approaches in both synthetic and real images to demonstrate the superior performance of the proposed algorithm.

Introduction

As one of the classical problems in image processing, image segmentation has been extensively studied, which can be treated as a classification problem [1–5] for the target image. Various image segmentation algorithms have been developed such as active contour models [6, 7], graph based methods [8, 9] and clustering techniques [10–12]. Over the last decades, model-based techniques [13, 14] have been widely used in image segmentation, where the standard Gaussian mixture model (GMM) [15, 16] is a well-known method because of its simplicity and ease of implementation [17]. The parameters involved in GMM can be efficiently estimated by

expectation maximization (EM) algorithm [18]. However, the standard GMM still suffers from the following limitations: sensitivity to noise, less flexibility to fit the shape of the data and unbounded distributions [19].

In order to reduce noise sensitivity for segmentation, (hidden) Markov random fields ((H)MRF) based mixture models have been widely utilized for pixel labels [20–22], where (H)MRF is acted on neighboring labels, and the clustering result for each pixel depends on the neighboring pixels [23]. On the other hand, in order to impose spatial constraints among neighboring pixels, another group of mixture models with MRF has been proposed by modeling the joint distribution of the priors for each pixel [24–27]. For example, in [26], Displaros et al. proposed a generative GMM model by introducing a pseudo-likelihood quantity to incorporate the spatial smoothness constraints based on Kullback-Leibler (KL) divergence. In [27], Nikou et al. proposed a novel spatial constraint that can adaptively select spatial directions. In order to directly apply the EM algorithm to estimate the involved parameters, Nguyen and Wu [23] proposed a robust spatially constrained GMM by introducing a spatial factor into the prior distribution. Although the forementioned algorithms can reduce the impact of noise in the image, most (H)MRF based algorithms are still not sufficiently robust with respect to different noise types and levels.

Because of the utilization of Gaussian distribution in GMM, the distribution tail is often shorter for many applied problems [17], which means that the Gaussian distribution is not sufficiently flexible to fit data [19]. In order to improve the flexibility for the data fitness, the Student's-t distribution, Laplace distribution and generalized Gaussian distribution are used to replace the Gaussian distribution in mixture model. Therefore, the Student's-t mixture model (SMM) [28, 29], Laplace mixture model (LMM) [30, 31], and generalized Gaussian mixture model (GGMM) [32, 33] have been proposed. On the other hand, using only one distribution for each component in the mixture model is not sufficiently satisfactory for many practical applications. Therefore, another solution for fitting data with different shapes is using multiple distributions for each component. For example, Zhang et al. [34] proposed a modified GMM by incorporating local spatial and intensity information. The conditional probability for each pixel is constructed based on the probabilities of neighboring pixels [34]. Browne et al. [35] proposed a mixture of mixture model by combining a multivariate Gaussian distribution and a multivariate uniform distribution together to model the component density [35]. Nguyen et al. [19] proposed an asymmetric mixture model by modeling the component with multivariate Gaussian distributions [19].

Moreover, the distributions in most mixture models are unbounded with a supporting range of $(-\infty, +\infty)$, which is not consistent with the practical application where the practical data generally fall in a bounded region [19]. In [36], a bounded GMM (BGMM) was proposed for speech processing. In [37], a bounded generalized GMM was proposed that included GMM, LMM, GGMM, and BGMM as special cases. Nguyen et al. [17, 19, 38] proposed various bounded mixture models to fit different data shapes. However, the above mentioned approaches still suffer from the following limitations: (1) Without considering any spatial information, the mixture of mixture model [35] and the bounded asymmetric mixture model (BAMM) [19] are still sensitive to noise, although both of these types of models are more flexible. (2) For the bounded mixture models [19, 36–38], the bounded support regions of observed data should be predefined. Moreover, the indicator function of the bounded support region is a binary function that cannot easily manage uncertainty, vagueness, and incompleteness in data.

Motivated by the aforementioned observations, in this paper, we propose a rough set bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation. First, in our previous work [39], based on the rough set theory [40], we proposed a generalized rough fuzzy c-means (GRFCM) algorithm, where, for each cluster, an image is automatically

partitioned into three rough regions with two adaptively computed thresholds. In this paper, we utilize these two thresholds to partition the target image into three rough regions, i.e., the positive, boundary and negative regions [41]. Second, a new bounded indicator function is proposed to determine the bounded support regions of the observed data. The bounded indicator of a pixel that belongs to each sub-region is estimated with respect to the rough region where the pixel lies. Only those pixels in the positive and boundary regions have non-zero indicators. Therefore, because of the benefits of rough set theory, the proposed bounded indicator function can further manage uncertainty in data. Third, to further overcome the impact of noise and reduce over-smoothness for segmentations, two novel prior factors are proposed to introduce the spatial information. The proposed prior factors can be treated as the within- and between-cluster spatial constraints with spatial direction. Finally, to further improve the robustness of the model, for each component, the posterior probabilities of within- and between-cluster for each pixel are estimated with respect to the rough regions. The proposed algorithm is compared to several state-of-the-art segmentation algorithms on simulated and real images to demonstrate its superior performance.

Finite Mixture Model

The notations used throughout this paper are as follows. The target image is denoted as $X = \{x_i, i = 1, 2, \ldots, N\}$, where x_i with dimension D is the intensity values for the ith pixel. The neighborhood of the ith pixel is denoted as ∂_i, and the labels are denoted as $(\Omega_1, \Omega_2, \ldots, \Omega_K)$. In order to segment an image with N pixels into K labels, the density function of the finite mixture model [42] is given by:

$$p(x_i|\Pi, \Theta) = \sum_{k=1}^{K} \pi_{ik} p(x_i|\Omega_k), \tag{1}$$

where $\Pi = \{\pi_{ik}\}$, $i = \{1, 2, \ldots, N\}$, $k = \{1, 2, \ldots, K\}$ are the prior probabilities, and satisfy the constraints $0 \leq \pi_{ik} \leq 1$ and $\sum_{k=1}^{K} \pi_{ik} = 1$.

In GMM [15, 16], the component $p(x_i|\Omega_k)$ is the Gaussian distribution $\Phi(x_i|\mu_k, \Sigma_k)$ that can be written in the form:

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

where μ_k is the mean vector with D dimension, Σ_k is the covariance matrix with $D \times D$ dimension, and $|\Sigma_k|$ is the determinant of Σ_k.

In order to address the issue that the observed data generally fall within the bounded support regions in practical applications, in [19, 36, 37], the bounded support region in $\Re^D$ is defined as ∂_{Ω_k} for each label Ω_k, and the indicator function can be written as

$$H(x_i|\Omega_k) = \begin{cases} 1 & \text{if} \quad x_i \in \partial_{\Omega_k} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

With the above indicator function $H(x_i|\Omega_k)$ and distribution $p(x_i|\Omega_k)$, a bounded distribution $\tilde{p}(x_i|\Omega_k)$ can be defined as

$$\tilde{p}(x_i|\Omega_k) = \frac{p(x_i|\Omega_k)H(x_i|\Omega_k)}{\int_{\partial_{\Omega_k}} p(x|\Omega_k)dx}. \tag{4}$$

For additional analysis details, please refer to [19, 36] and [37]. However, the major disadvantage of indicator function $H(x_i|\Omega_k)$ is that it is a binary function that cannot easily manage uncertainty in data.

To improve the noise robustness, the spatial information is generally incorporated through MRF distribution:

$$p(\Pi) = \frac{1}{Z} exp\left\{-\frac{1}{T}U(\Pi)\right\}, \tag{5}$$

where $U(\Pi)$ is the smoothing prior, and Z and T are two constants. Based on Bayes' rules, the probability density function can be written as:

$$p(\Pi, \Theta|X) \propto p(X|\Pi, \Theta)p(\Pi). \tag{6}$$

Most MRF-based mixture models have been successfully applied to image segmentation by adopting different energy functions $U(\Pi)$. Nguyen and Wu [23] pointed out that the M-step of EM cannot be applied directly to the prior distribution π_{ik} due to the complexity of the log-likelihood function. Thus, the resulting algorithms are computationally complex and have to utilize large amounts of computational power to solve the constrained optimization problem of the prior distribution π_{ik} [23]. To overcome these disadvantages, they introduced a novel factor G_{ik} by defining a multiplication of both posterior probability and prior distributions as follows.

$$G_{ik} = exp\left[\frac{\beta}{2N_i}\sum_{m\in\partial_i}(z_{mk} + \pi_{mk})\right], \tag{7}$$

where z_{mk} is the posterior probability and β is the balance parameter to control the smoothing prior. The main advantage of G_{ik} is the ease of implementation and incorporation of the spatial relationships amongst neighborhood pixels in a simpler metric. Then the smoothing prior $U(\Pi)$ is given by:

$$U(\Pi) = -\sum_{i=1}^{N}\sum_{k=1}^{K}G_{ik}log\pi_{ik}. \tag{8}$$

However, the energy $U(\Pi)$ can cause over-smoothing for segmentation and loss of details, especially for regions with abundant textures.

Proposed Model

In order to fit different data shapes, Nguyen et al [19] defined a new distribution $p(x_i|\Omega_k)$ to model the component density. Motivated by the bounded asymmetric distribution, in this paper we modify distribution $p(x_i|\Omega_k)$ to allow the model to easily incorporate the spatial information, which can be defined as:

$$p(x_i|\Omega_k) = \sum_{l=1}^{L}\eta_{ikl}\Psi(x_i|\mu_{kl}, \Sigma_{kl}), \tag{9}$$

where L is the number of bounded multivariate Gaussian distribution $\Psi(x_i|\mu_{kl}, \Sigma_{kl})$, and η_{ikl} is the weighting factor and satisfies the constraints $0 \leq \eta_{ikl} \leq 1$ and $\sum_{l=1}^{L}\eta_{ikl} = 1$. The bounded

Gaussian distribution $\Psi(x_i|\mu_{kl}, \Sigma_{kl})$ is defined as:

$$\Psi(x_i|\mu_{kl}, \Sigma_{kl}) = \frac{\Phi(x_i|\mu_{kl}, \Sigma_{kl})H(x_i|\Omega_k)}{\int_{\partial_{\Omega_k}} \Phi(x|\mu_{kl}, \Sigma_{kl})dx}, \tag{10}$$

where $\Phi(x_i|\mu_{kl}, \Sigma_{kl})$ is the Gaussian distribution defined in Eq (2) and $H(x_i|\Omega_k)$ is the indicator function for the bounded support region defined in Eq (3). $\int_{\partial_{\Omega_k}} \Phi(x|\mu_{kl}, \Sigma_{kl})dx$ is the normalization constant.

Therefore, in this paper, we propose a rough set bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation. First, based on the rough set theory, we utilize our previous work [39] to partition the target image into three rough regions with two adaptively computed thresholds. Second, a new bounded indicator function is proposed to determine the bounded support regions of the observed data based on the above rough regions. Third, to further overcome the impact of noise and reduce over-smoothness for segmentations, two novel prior factors with spatial direction are constructed based on the prior and posterior probabilities of the within- and between-clusters. Finally, to further improve model robustness, for each component, the posterior probability is re-estimated based on the adaptively determined rough regions.

Determination of rough set region

Rough set theory can manage the uncertainty with lower and upper approximations. Specifically, let $U \neq \emptyset$ be a universe of discourse, and R be an equivalence relationship that can lead to a U partition. By denoting $U/R = \{X_1, X_2, ..., X_n\}$, where X_c is an equivalence class for R, the lower and upper approximations of subset X are defined as:

$$\begin{cases} \bar{R}X = \cup\{Y \in U/R | Y \cap X \neq \emptyset\} \\ \underline{R}X = \cup\{Y \in U/R | Y \subseteq X\} \end{cases}. \tag{11}$$

The lower approximation is a set where all categories certainly belong to X, and the upper approximation is a set where all categories possibly belong to X. Based on the approximations, three rough regions of X, i.e., the R-positive Po region, R-negative Ne region, and R-boundary Bo region, can be defined as follows [41]:

$$Po = \underline{R}X, \quad Ne = U - \bar{R}X, \quad Bo = \bar{R}X - \underline{R}X. \tag{12}$$

Fig 1 illustrates the definition of the three rough regions. The figure shows that positive region Po equals the lower approximation of Ω_k, whereas negative region Ne equals the complement of the upper approximation of Ω_k.

In our previous work [39], based on the distance between each pixel value x_i and intensity level g_j, two thresholds are adaptively computed to determine the rough regions:

$$d_i(g_j) = \frac{\sqrt{\sum_{k \in \partial_i} (x_k - g_j)^2}}{n(J_{max} - J_{min})}, \quad j = 1, 2, ..., J, \tag{13}$$

where ∂_i is the neighborhood of pixel i with n pixels. In this paper, the size of neighborhood ∂_i is set as 3×3. J_{max} and J_{min} are the maximum and minimum intensity values of the image, respectively, and J is the number of intensity levels in the image. Thus, we construct distance vector $d_i = \{d_i(g_1), d_i(g_2), ..., d_i(g_J)\}$ for each pixel i. The two thresholds can be estimated based
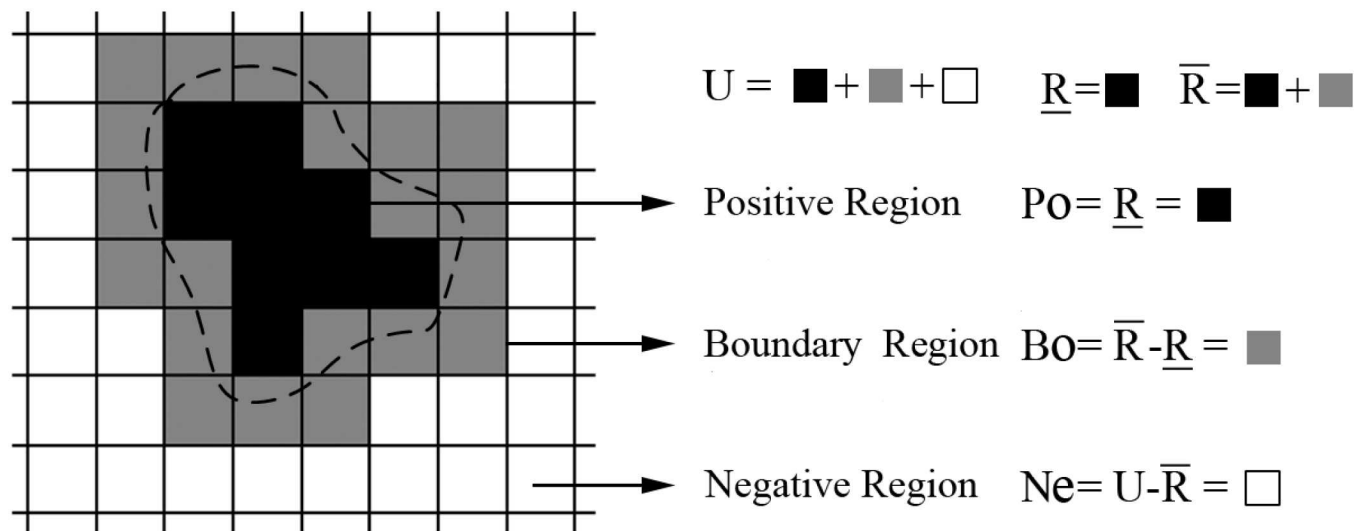
Fig 1. Illustration of three rough regions.

on mean value d_{i_mean} and minimum value d_{i_min} of vector d_i:

$$t_1 = \frac{1}{N} \sum_{i=1}^{N} d_{i\_min} \quad and \quad t_2 = \frac{1}{N} \sum_{i=1}^{N} d_{i\_mean}. \tag{14}$$

Based on the distance between each pixel i and the mean value of class k using Eq (13), we can determine the rough regions for each cluster Ω_k as follows:

$$x \in \begin{cases} Po_k & if \quad d_i(\mu_k) \le t_1 \\ Bo_k & if \quad t_1 < d_i(\mu_k) \le t_2 \\ Ne_k & otherwise \end{cases} \tag{15}$$

For more analysis and discussion details, please refer to [39].

Construction of bounded support region

As mentioned before, the bounded support regions of observed data were predefined in [19, 36–38], and the indicator function of the bounded support region is a binary function that cannot easily manage uncertainty in data. Motivated by the aforementioned observations, in this paper, we propose a new bounded indicator function based on the rough regions. For each label Ω_k, the bounded support region in $\Re^D$ is defined as ∂_{Ω_k}, and the new indicator function can be written as:

$$\tilde{H}(x_i|\Omega_k) = \begin{cases} 1 & x_i \in Po_k \\ \frac{t_2 - d_i(\mu_k)}{t_2 - t_1} & x_i \in Bo_k \\ 0 & x_i \in Ne_k \end{cases} \tag{16}$$

With the new indicator function $\tilde{H}(x_i|\Omega_k)$ and distribution $p(x_i|\Omega_k)$ in Eq (2), a bounded multivariate Gaussian distribution $\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})$ can be defined as:

$$\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl}) = \frac{\Phi(x_i|\mu_{kl}, \Sigma_{kl})\tilde{H}(x_i|\Omega_k)}{\int_{\Omega_k}\Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx},$$ (17)

where $\int_{\Omega_k}\Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx$ is the normalization constant, and it is identified as the share of $\Phi(x_i|\mu_{kl}, \Sigma_{kl})$ that belongs to support region ∂_{Ω_k} [19].

Similar to BAMM [19], each component density in our model is constructed with multiple bounded asymmetric distribution. The corresponding distribution $p(x_i|\Omega_k)$ is defined as:

$$p(x_i|\Omega_k) = \sum_{l=1}^{L}\eta_{ikl}\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl}),$$ (18)

where L is the number of bounded multivariate Gaussian distribution $\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})$, and η_{ikl} is the weighting factor that satisfies the constraints $0 \leq \eta_{ikl} \leq 1$ and $\sum_{l=1}^{L}\eta_{ikl} = 1$.

It should be noted that the above distribution always satisfies the conditions of the probability density [14]:

$$\begin{cases} p(x_i|\Omega_k) \geq 0 \\ \int_{-\infty}^{+\infty} p(x_i|\Omega_k)dx = \int_{-\infty}^{+\infty}\sum_{l=1}^{L}\eta_{ikl}\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl}) = 1 \end{cases}.$$ (19)

Therefore, the log-likelihood function of the proposed model can be written as

$$\begin{aligned} L(\Pi, \Theta|X) &= \sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_{ik}p(x_i|\Omega_k)\right) \\ &= \sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_{ik}\sum_{l=1}^{L}\eta_{ikl}\frac{\Phi(x_i|\mu_{kl}, \Sigma_{kl})\tilde{H}(x_i|\Omega_k)}{\int_{\Omega_k}\Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx}\right). \end{aligned}$$ (20)

In order to maximize the above likelihood function, two variables z_{ik} and y_{ikl} are defined as follows:

$$z_{ik} = \frac{\pi_{ik}\displaystyle\sum_{l=1}^{L}\eta_{ikl}\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})}{\displaystyle\sum_{m=1}^{K}\left[\pi_{im}\sum_{l=1}^{L}\eta_{ikl}\tilde{\Psi}(x_i|\mu_{ml}, \Sigma_{ml})\right]},$$ (21)

$$y_{ikl} = \frac{\eta_{ikl}\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})}{\displaystyle\sum_{m=1}^{L}\eta_{ikm}\tilde{\Psi}(x_i|\mu_{km}, \Sigma_{km})}.$$ (22)

The values z_{ik} and y_{ikl} always satisfy the conditions $\sum_{k=1}^{K} z_{ik} = 1$ and $\sum_{l=1}^{L} y_{ikl} = 1$, respectively. It is worth mentioning that, based on Bayes'rules, we can treat both variables z_{ik} and y_{ikl} as the posterior probability. Variable z_{ik} indicates the relationship between pixels and clusters that can be treated as the between-cluster relationship. Meanwhile, variable y_{ikl} indicates the relationship between pixels and distribution components that can be treated as the within-cluster relationship because each cluster is modeled with multiple distributions. Consequently,

variable z_{ik} can be treated as the posterior probabilities of the between-cluster, whereas variable y_{ikl} can be treated as the posterior probabilities of the within-cluster.

To further improve model robustness, for each component, the posterior probability of each pixel is re-estimated with respect to the rough region where the pixel lies. Thus, the new hidden variables $\tilde{z}_{ik}$ and $\tilde{y}_{ikl}$ can be rewritten as:

$$\tilde{z}_{ik} = \begin{cases} 1 & \text{if} \quad x_i \in Po_k \\ z_{ik} & \text{if} \quad x_i \in Bo_k \\ 0 & \text{if} \quad x_i \in Ne_k \end{cases}, \tag{23}$$

$$\tilde{y}_{ikl} = \begin{cases} 1 & \text{if} \quad x_i \in Po_k \\ y_{ikl} & \text{if} \quad x_i \in Bo_k \\ 0 & \text{if} \quad x_i \in Ne_k \end{cases}. \tag{24}$$

Construction of prior factor

As we mentioned before, the prior factor G_{ik} in [23] plays a role as an average filter on both posterior probability and prior distributions for smoothing noisy images, which may cause over smoothing for the segmentation and lose the details especially for the regions with abundant textures. The smoothing prior in [27] can adaptively select spatial directions, which introduces additional training complexity. To reduce the complexity of the smoothing prior and preserve more details for the segmentations, we propose two novel prior factors E_{ik} and F_{ikl} that consider spatial direction based on the prior and posterior probabilities of the between- and within-cluster in this paper. Both prior factors are defined as:

$$E_{ik} = \frac{\exp\left[\sum_{m\in\partial_i^{S_k^*}}(z_{mk} + \pi_{mk})/N_i^{S_k^*}\right]}{\sum_{h=1}^{K}\exp\left[\sum_{m\in\partial_i^{S_h^*}}(z_{mh} + \pi_{mh})/N_i^{S_h^*}\right]}, \tag{25}$$

$$F_{ikl} = \frac{\exp\left[\sum_{m\in\partial_i^{S_l^*}}(y_{mkl} + \eta_{mkl})/N_i^{S_l^*}\right]}{\sum_{h=1}^{L}\exp\left[\sum_{m\in\partial_i^{S_h^*}}(y_{mkh} + \eta_{mkh})/N_i^{S_h^*}\right]}, \tag{26}$$

where $\partial_i^{S_k^*}$ is the neighborhood of pixel i at direction S_k^* for cluster k, which contains $N_i^{S_k^*}$ pixels, and S_k^* is given by:

$$S_k^* = \arg\min_{s=1}^{S} \sum_{m\in\partial_i^s} dist(x_m, \mu_k), \tag{27}$$

where $dist(x_i, \mu_k)$ is the Euclidean distance between point i and cluster center μ_k, and ∂_i^s is the neighborhood of pixel i at direction s. In this paper, we set $S = 4$, i.e., four directions (horizontal, vertical and two diagonal) are considered.

More explicitly, taking 3×3 neighborhood pixels as example, the filters for each neighboring direction can be constructed as shown in Fig 2. During the algorithm procedure, we only need to operate the convolution over the prior and posterior probabilities with these four
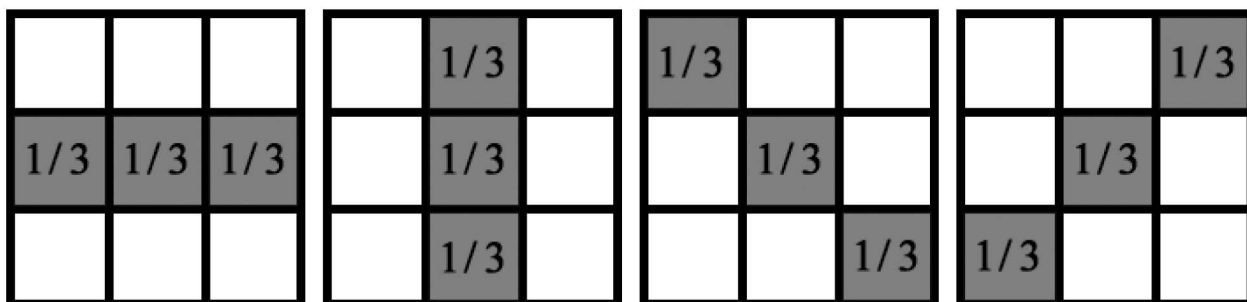
Fig 2. An example of spatial filters that considers four directions. From left to right: horizontal, vertical and two diagonal directions.

predefined filters, and then adaptively select the satisfactory direction based on the differences among the intensity values along each direction. Therefore, the proposed prior factors can efficiently preserve more details.

Then, we incorporate the proposed prior factors into the smoothing prior:

$$U(\Pi) = -\left(\sum_{i=1}^{N}\sum_{k=1}^{K} E_{ik} \log \pi_{ik} + \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L} F_{ikl} \log \eta_{ikl} \right) \tag{28}$$

MRF distribution $p(\Pi)$ can be given by:

$$p(\Pi) = Z^{-1} \exp\left\{ \frac{1}{T}\left(\sum_{i=1}^{N}\sum_{k=1}^{K} E_{ik} \log \pi_{ik} \right.\right.$$
$$\left.\left. + \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L} F_{ikl} \log \eta_{ikl} \right) \right\}. \tag{29}$$

where Z and T are set as 1. Therefore, the log-likelihood function of the proposed algorithm can be written in the form:

$$
\begin{aligned}
L(\Pi, \Theta|X) =& \sum_{i=1}^{N}\sum_{k=1}^{K} E_{ik} \log \pi_{ik} + \sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L} F_{ikl} \log \eta_{ikl} \\
& + \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_{ik} \sum_{l=1}^{L} \frac{\eta_{ikl}\Phi(x_i|\mu_{kl},\Sigma_{kl})\tilde{H}(x_i|\Omega_k)}{\int_{\Omega_k}\Phi(x|\mu_{kl},\Sigma_{kl})\tilde{H}(x|\Omega_k)dx} \right).
\end{aligned}
\tag{30}
$$

Finally, we can optimize the involved parameters by maximizing the above log-likelihood function.

Parameter estimation

In order to determine label Ω_k for each pixel x_i, we need to estimate parameters $\Pi = \{\pi_{ik}, \eta_{ikl}\}$ and $\Theta = \{\mu_{kl}, \Sigma_{kl}\}$ by maximizing log-likelihood function $L(\Pi, \Theta|X)$. By monotonically increasing the logarithm character, maximizing log-likelihood function $L(\Pi, \Theta|X)$ can lead to

minimizing objective function $J(\Pi, \Theta | X)$ as follows:

$$
\begin{aligned}
J(\Pi, \Theta | X) &= -L(\Pi, \Theta | X) \\
&= -\left\{ \sum_{i=1}^{N} \log \left(\sum_{k=1}^{K} \pi_{ik} \sum_{l=1}^{L} \frac{\eta_{ikl} \Phi(x_i | \mu_{kl}, \Sigma_{kl}) \tilde{H}(x_i | \Omega_k)}{\int_{\Omega_k} \Phi(x | \mu_{kl}, \Sigma_{kl}) \tilde{H}(x | \Omega_k) dx} \right) \right. \\
&\quad \left. + \sum_{i=1}^{N} \sum_{k=1}^{K} E_{ik} \log \pi_{ik} + \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} F_{ikl} \log \eta_{ikl} \right\}
\end{aligned}
\tag{31}
$$

By applying the complete data condition in [24], minimizing the negative log-likelihood function in Eq (31) can also lead to minimizing the objective function $E(\Pi, \Theta | X)$ as follows:

$$
\begin{aligned}
E(\Pi, \Theta | X) &= -\left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \tilde{z}_{ik} \left[\log \pi_{ik} + \sum_{l=1}^{L} \tilde{y}_{ikl} \times \right. \right. \\
&\quad \left(\log \eta_{ikl} + \log \tilde{H}(x_i | \Omega_k) + \log \Phi(x_i | \mu_{kl}, \Sigma kl) \right. \\
&\quad \left. \left. \left. - \log \int_{\Omega_k} \Phi(x | \mu_{kl}, \Sigma kl) \tilde{H}(x | \Omega_k) dx \right) \right] \right\} \\
&\quad -\left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} E_{ik} \log \pi_{ik} + \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{l=1}^{L} F_{ikl} \log \eta_{ikl} \right\}.
\end{aligned}
\tag{32}
$$

Then we can apply the EM algorithm to minimize Eq (32) by considering the derivation of function $E(\Pi, \Theta | X)$ with respect to each variable. Finally, we can obtain the following updating function for each variable. Please refer to the Appendix for a detailed derivation of the proposed algorithm.

Mean value estimation:

$$
\mu_{kl} = \frac{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} x_i}{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl}} - \frac{\sum_{m=1}^{M} (\mu_{kl} - s_{mkl}) \tilde{H}(s_{mkl} | \Omega_k)}{\sum_{m=1}^{M} \tilde{H}(s_{mkl} | \Omega_k)}.
\tag{33}
$$

Covariance matrix estimation:

$$
\begin{aligned}
\Sigma_{kl} &= \frac{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} (x_i - \mu_{kl})(x_i - \mu_{kl})^T}{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl}} \\
&\quad - \frac{\sum_{m=1}^{M} ((s_{mkl} - \mu_{kl})(s_{mkl} - \mu_{kl})^T - \Sigma_{kl}) \tilde{H}(s_{mkl} | \Omega_k)}{\sum_{m=1}^{M} \tilde{H}(s_{mkl} | \Omega_k)}.
\end{aligned}
\tag{34}
$$

Prior probability estimation:

$$
\pi_{ik} = \frac{\tilde{z}_{ik} + E_{ik}}{\sum_{h=1}^{K} (\tilde{z}_{ih} + E_{ih})}.
\tag{35}
$$

$$
\eta_{ikl} = \frac{\tilde{y}_{ikl} + F_{ikl}}{\sum_{h=1}^{L} (\tilde{y}_{ikh} + F_{ikh})}.
\tag{36}
$$

Consequently, the proposed algorithm for image segmentation is summarized as follows.

Step 1. Initialize parameters $\Pi = \{\pi_{ik}, \eta_{ikl}\}$ and $\Theta = \{\mu_{kl}, \Sigma_{kl}\}$ with the K-means algorithm.

Step 2. Determine the rough regions for each cluster with Eqs (13)–(15).

Step 3. **E-step**
Update posterior probabilities $\tilde{z}_{ik}$ and $\tilde{y}_{ikl}$ with Eqs (21)–(24).

Step 4. **M-step**

Step 4.1. Update prior factors E_{ik} and F_{ikl} with Eqs (25) and (26).

Step 4.2. Update means μ_{kl} with Eq (33).

Step 4.3. Update covariance values Σ_{kl} with Eq (34).

Step 4.4. Update prior probabilities π_{ik} and η_{ikl} with Eqs (35) and (36).

Step 5. Check the convergence. Stop the iteration if the convergence criterion is satisfied; otherwise, go to Step 2.

It should be noted that the convergence criterion is generally the distance between the values of objective functions or variables (i.e. means or covariance values) from two successive iterations. In this paper, we utilize the total distance between the mean values obtained from two successive iterations. When this distance becomes smaller than a user specified threshold, which was set to 10^{-5} in all algorithms for this study, we think the algorithm converges and stop the iteration.

Experimental Results

In this paper, we compare the proposed algorithm with four algorithms, i.e., a Bayesian bounded asymmetric mixture model (BAMM) [19], a bounded generalized Gaussian mixture model (BGGMM) [37], a spatially constrained generative model and EM algorithm (SCGM-EM) [26], and a fast and robust spatially constrained GMM (FRSCGMM) [23].

Unless otherwise specified, the parameters of the proposed algorithm are set as follows: The window size for prior factor construction is 3×3. The number of bounded multivariate Gaussian distribution is $L = 3$. A summary of the parameter settings for each comparison algorithm is listed in Table 1. Please see the corresponding references for more details. It should be noted that for fair comparison, all algorithms, including the proposed model, use the same initializations generated by the k-means algorithm for each testing image. All algorithms were implemented using MATLAB 7.8 platform and tested on a PC (Intel Core i7-4790 CPU, 3.60GHz, 16GB RAM, and 64-bit Windows 8).

The algorithms are compared using synthetic, synthetic and real brain MR, and color images. Fig 3 shows an example of each type of testing image.

Table 1. Summary of parameter setting for each comparison algorithm in experiment.

| Algorithms	Parameter setting
SCGM-EM	Temperature value $\beta = 0.5$
	Size of neighborhood 5×5
FRSCGMM	Temperature value $\beta = 12$
	Size of neighborhood 5×5
BAMM	The number of Gaussian distribution $K_j = 3$
	The number of random variables $M = 10^6$
	The bounded support regions $\partial_{\Omega_j} \in (0, 255)$
BGGMM	The number of random variables $M = 10^6$
	The bounded support regions $\partial_{\Omega_j} \in (0, 255)$

doi:10.1371/journal.pone.0168449.t001
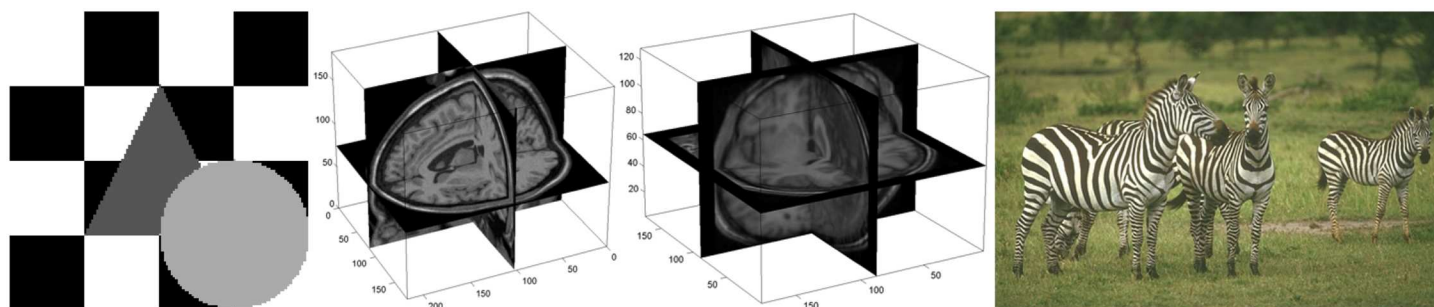
Fig 3. Examples of testing images. From left to right: synthetic image, simulated T1-weighted brain MR, real T1-weighted brain MR and natural images.

doi:10.1371/journal.pone.0168449.g003

(1) The synthetic images used in this paper, of size 128×128, contain different labels with luminance values. For example, the image shown at the left of Fig 3 contains four labels with luminance values [0, 1/3, 2/3, 1]. Then we add to the synthetic images different types of noise with different levels for each comparison experiment.

The segmentation results for synthetic images are quantitatively evaluated by correct classification ratio (CCR) [27], which is defined as:

$$CCR = \sum_{k=1}^{K} \frac{|gt_k \bigcap seg_k|}{|GT|},$$ (37)

where gt_k is the ground truth for cluster k, seg_k describes the pixels classified by the algorithm to cluster k and $GT = \bigcup_{k=1}^{K} gt_k$. CCR ranges from 0 to 1 with a higher value representing a better segmentation result.

(2) The brain MR images are selected from two open sources: the BrainWeb (http://www.bic.mni.mcgill.ca/brainweb) [43] and IBSRv2.0 (https://www.nitrc.org/projects/ibsr) [44] databases. The objective for brain MR image segmentation is to partition the image into three tissue labels: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The Dice coefficient (DC) [45] is utilized to quantitatively evaluate the performance for segmenting each type of brain tissue, which is the ratio between the intersection and union of segmented volume S_1 and ground truth volume S_2

$$DC(S_1, S_2) = \frac{2|S_1 \bigcap S_2|}{|S_1| + |S_2|}.$$ (38)

The DC value ranges from 0 to 1, with a higher value representing a more accurate segmentation result.

(3) The natural images are selected from the open source Berkeley Segmentation Dataset 500 (BSDS500, https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/) [46]. The probabilistic rand index (PRI) [47] is used to assess the segmentation performances on natural images. PRI between segmentation map S_{seg} to be evaluated and a set of M ground truth images $S_{gt} = \{S_1, \ldots, S_M\}$ is given by:

$$PRI(S_{seg}, S_{gt}) = \frac{2}{N(N-1)} \sum_{\substack{i,j \\ i<j}} \left[c_{ij} p_{ij} + (1 - c_{ij})(1 - p_{ij}) \right],$$ (39)
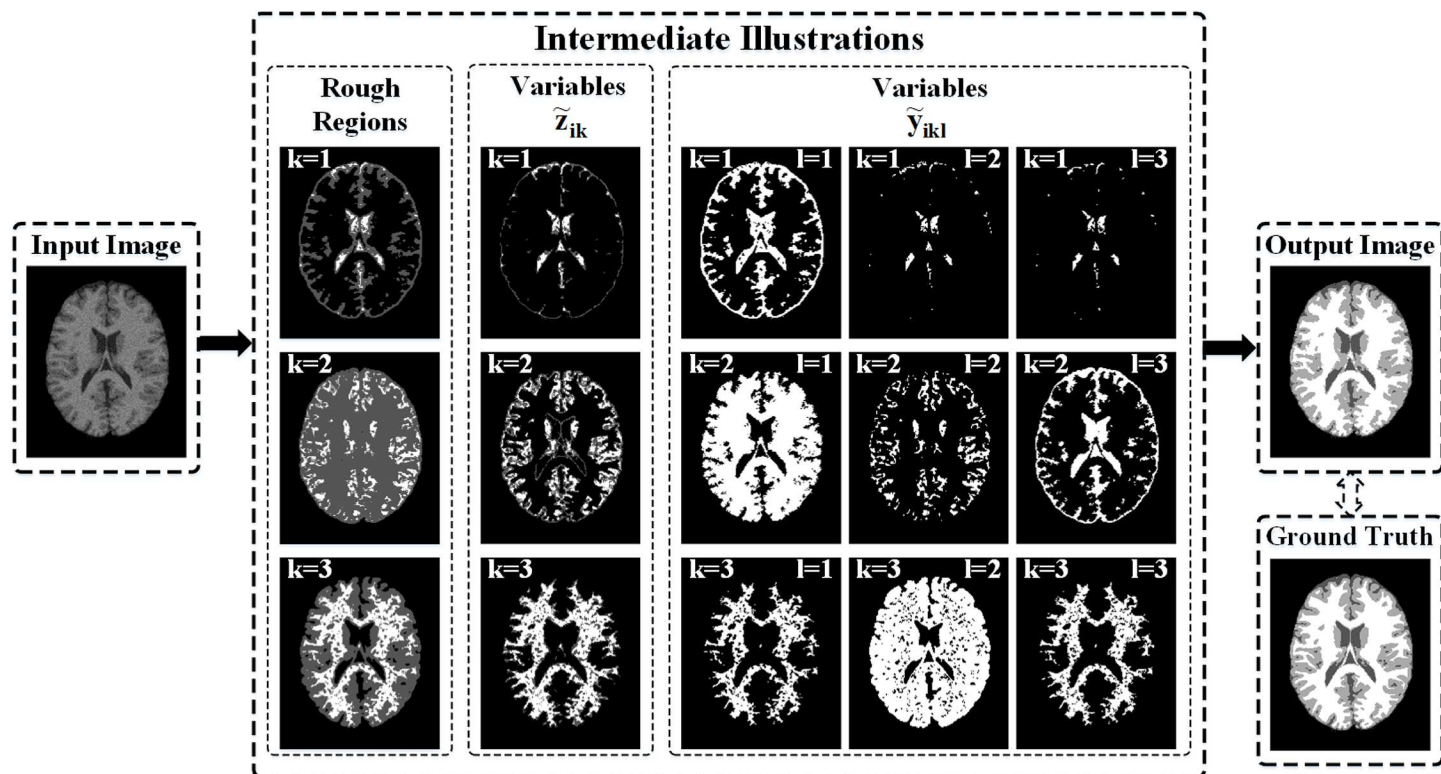
Fig 4. Illustrations of proposed algorithm.

where $c_{ij} = 1$ if pixels i and j belong to the same cluster, and $c_{ij} = 0$ if pixels i and j belong to different clusters. PRI takes values between 0 and 1, with a higher value representing a more accurate segmentation result.

Illustration of proposed algorithm

In order to show the advantages and performance of the proposed algorithm in detail, we first segmented a brain MR image that is the 90-th slice of the 3D brain MR image with 9% noise selected from BrainWeb. The component number for the mixture model is $K = 3$. Meanwhile, we set the number of bounded multivariate Gaussian distribution as $L = 3$. The original noisy brain MR image, intermediate illustrations of the proposed algorithm, and final output segmentation are shown in Fig 4. Three rough regions for each cluster, i.e., CSF ($k = 1$), GM ($k = 2$), and WM ($k = 3$) are shown in the first part of the intermediate illustrations, where the positive, boundary, and negative regions are illustrated in bright, gray, and dark respectively. This reveals that the proposed algorithm can construct rough regions appropriately. The corresponding posterior probability of the between and within-cluster, i.e., $\tilde{z}_{ik}$ and $\tilde{y}_{ikl}$, are depicted in the second and third parts of the intermediate illustrations, respectively. From the nine images of variables $\tilde{y}_{ikl}$, we find that variables $\tilde{y}_{ikl}$ provide different types of segmentation results for each cluster/tissue, including over and insufficient segmentations, which can further improve the accuracy of the posterior probability estimation by combining with the automatically computed prior probabilities (weights for each distribution). Comparing the last two images on the left of Fig 4, we find that the proposed algorithm can effectively overcome the
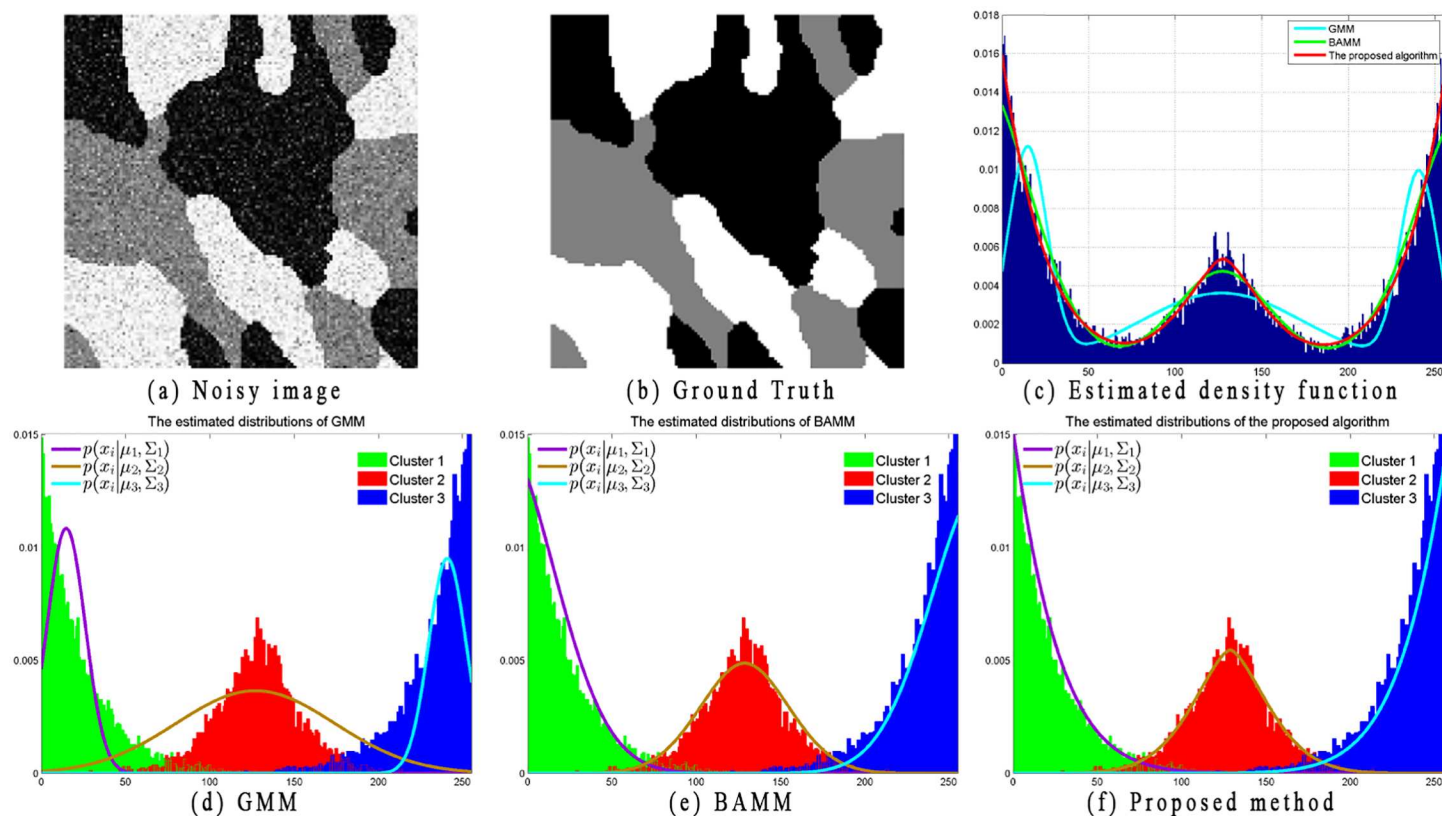
Fig 5. Illustrations of estimated distributions on synthetic image.

impact of noise, and the corresponding segmentation result is consistent with the ground truth.

Moreover, to further explain the limitations of GMM and the advantages of the proposed distributions, two examples with image histogram and corresponding estimated distributions are illustrated in Figs 5 and 6. Because BAMM [19] has already proven its superior performance over GMM, SMM, and GGMM, in this section, we only present the estimated distributions obtained by employing $\phi(x_i|\mu_k, \Sigma_k)$ (i.e., the Gaussian distribution in the GMM model, Eq (2)), $\Psi(x_i|\mu_{kl}, \Sigma_{kl})$ (i.e., the bounded Gaussian distribution in BAMM model, Eq (10)), and $\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})$ (i.e. the proposed distribution, Eq (17)).

In Fig 5, the noisy image with size 128×128 shown in Fig 5(a) is used to compare the performance of the proposed algorithm with GMM and BAMM. This image contains three labels ($K = 3$). The number of bounded multivariate Gaussian distributions for BAMM and the proposed algorithm is set as $L = 3$. The bounded region for BAMM is set as [0, 255] for each cluster. The ground truth of this image is shown in Fig 5(b), which presents the estimated density function obtained by GMM, BAMM, and the proposed algorithm based on the histogram of the observed data. Fig 5(d)–5(f) show the ground truth distributions for each class along with the estimated distributions of each corresponding cluster. From the histogram of observed data and corresponding ground truth distributions, we find that the image data are non-Gaussian, non-symmetric and bounded support data. Without the bounded constraint, GMM performance is not sufficiently satisfactory because the estimated density function cannot fit
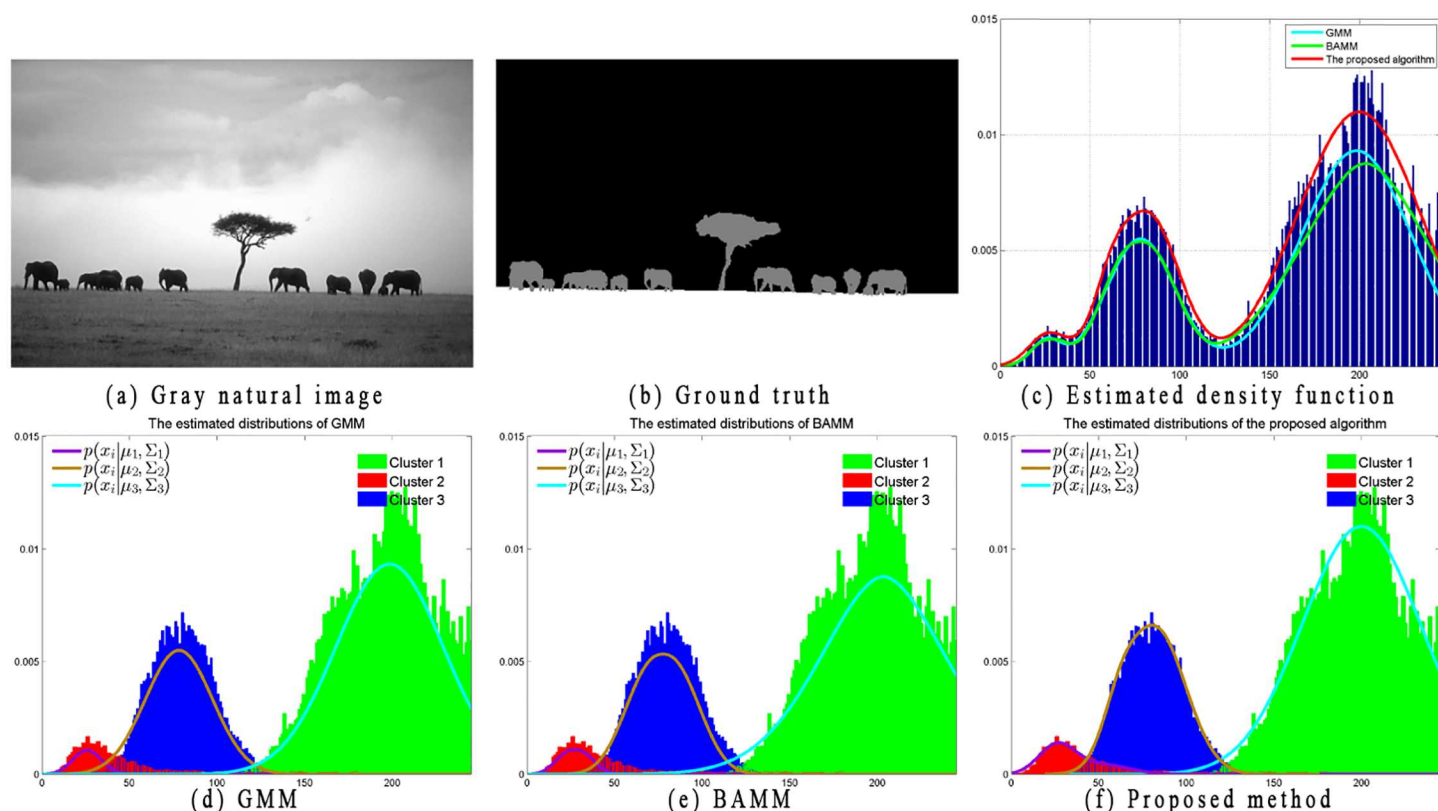
Fig 6. Illustrations of estimated distributions on natural image.

doi:10.1371/journal.pone.0168449.g006

well with the histogram of observed data. Compared with GMM, both BAMM and the proposed algorithm better fit the observed data because of the introduction of the bounded multivariate Gaussian distribution. It can also be visualized that distribution $\tilde{\Psi}(x_i|\mu_{kl}, \Sigma_{kl})$ in Eq (17) allows flexibility for a better fit of the observed data compared with distribution $\Psi(x_i|\mu_{kl}, \Sigma_{kl})$ of BAMM in Eq (10).

Similar to Figs 5 and 6 shows a gray natural image of size 481 × 321 that is used to compare the performance of the proposed algorithm with GMM and BAMM. This image contains three labels (K = 3). In comparison, the estimated distribution of the proposed algorithm is much better than that of the others.

Segmentation of synthetic images

In the second experiment, a synthetic image (image size: 128 × 128) as shown in Fig 7(a), is used to compare the performances among different algorithms. The image contains four labels with luminance values [0, 1/3, 2/3, 1]. The noisy image with Gaussian noise (0 means, 0.07 variance) is shown in Fig 7(b). From Fig 7(c)–7(g), we present the segmentation results obtained by employing the proposed algorithm, SCGM-EM, FRSCGMM, BAMM and BGGMM, respectively. Without considering any spatial information, the segmentation accuracies of BAMM and BGGMM are quite poor. The anti-noise ability for FRSCGMM is limited. Both SCGM-EM and the proposed algorithm obtain better performances. Nevertheless, the proposed method obtains higher CCR, especially for the pixels around the boundaries.
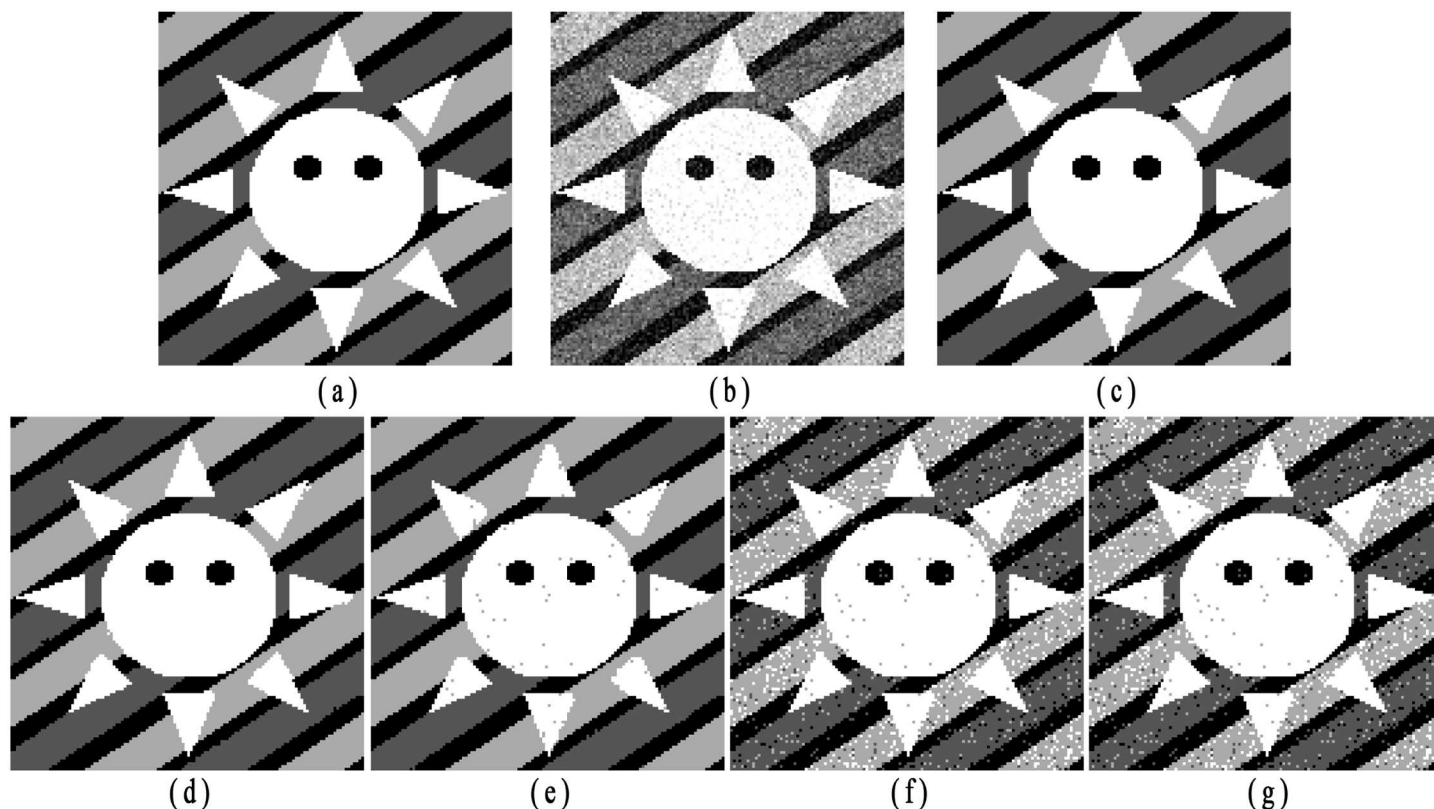
Fig 7. Experimental results on synthetic image with Gaussian noise (image size: 128 × 128). (a) Original image, (b) Noisy image with Gaussian noise (0 mean, 0.07 variance); segmentation results by applying (c) proposed algorithm (CCR = 0.9954), (d) SCGM-EM (CCR = 0.9942), (e) FRSCGMM (CCR = 0.9906), (f) BAMM (CCR = 0.9131), (g) BGGMM (CCR = 0.9152).

doi:10.1371/journal.pone.0168449.g007

Then, the synthetic image with luminance values [0, 1/4, 2/4, 3/4, 1] shown in Fig 8(a) is utilized to test the performance of different algorithms on different noise type. As shown in Fig 8 (b), the original image is corrupted with multiplicative noise (speckle noise with mean 0 and variance 0.04.). Fig 8(c)–8(g) present the segmentation results obtained by employing the proposed algorithm, SCGM-EM, FRSCGMM, BAMM and BGGMM, respectively. Among these methods, the proposed algorithm obtains better preservation of image details.

Segmentation of brain MR images

Because of the utilization of our previous work [39], in this section, we supplement the generalized rough fuzzy c-means (GRFCM) algorithm [39] as a comparison method. It should be noted that GRFCM can overcome the impact of intensity inhomogeneity in brain MR images, but all other comparison algorithms can only overcome the impact of noise. Therefore, in this experiment, we apply all algorithms to segment the synthetic T1-weighted 1 mm brain MR images selected from BrainWeb, which only contain different noise levels. For a fair comparison, the intensity inhomogeneity estimation part in GRFCM is removed.

Three sample brain MR images with 9% noise (80-th axial, sagittal and coronal slice), along with their segmentation results and ground truths, are shown in Fig 9. Similar to the proposed algorithm, both SCGM-EM and FRSCGMM construct spatial information based on the posterior and prior probabilities. However, without considering any directional differences, these
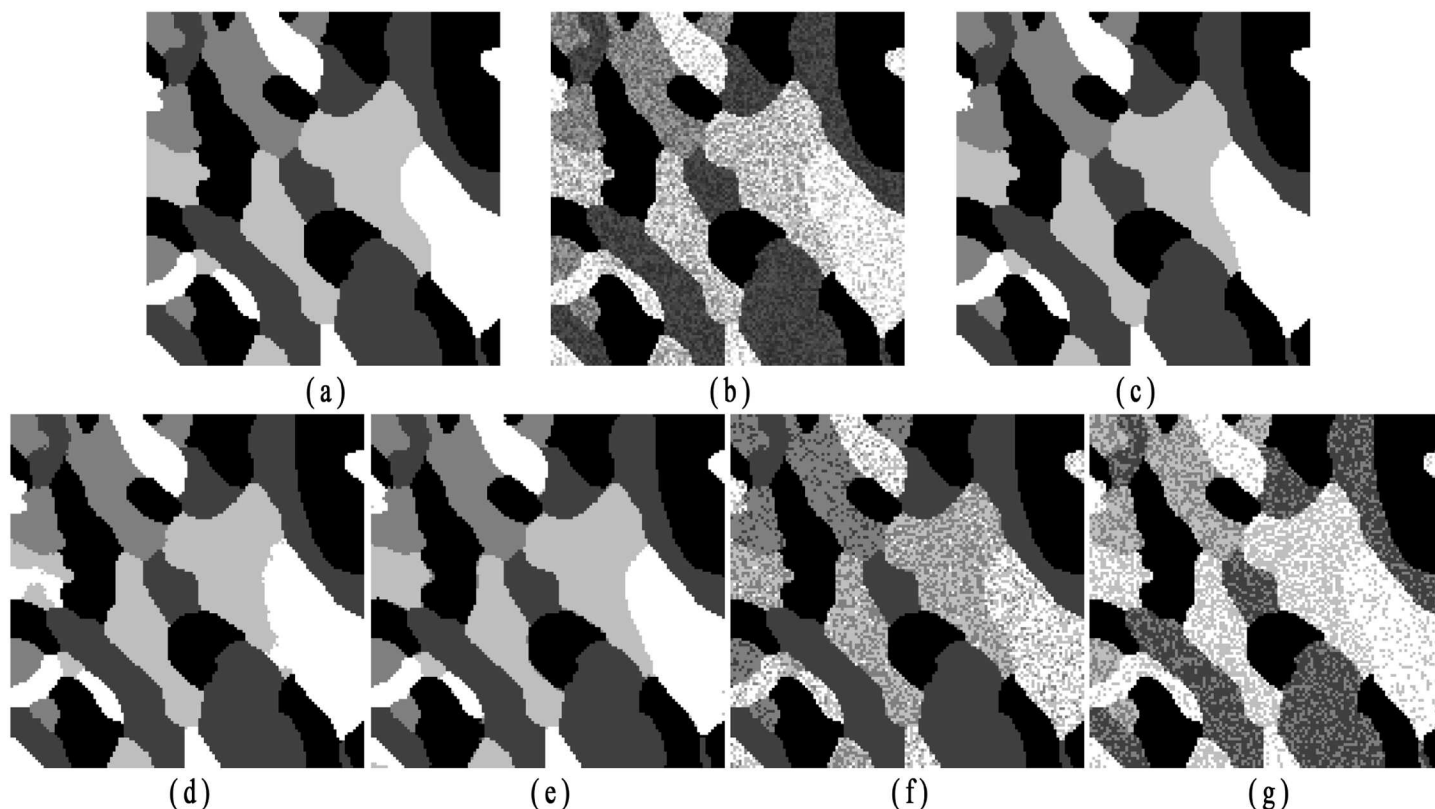
Fig 8. Experimental results on synthetic image with speckle noise (image size: 128 × 128). (a) Original image, (b) Noisy image with speckle noise (0 mean, 0.04 variance), segmentation results by applying (c) proposed algorithm (CCR = 0.9956), (d) SCGM-EM (CCR = 0.9802), (e) FRSCGMM (CCR = 0.9835), (f) BAMM (CCR = 0.8174), (g) BGGMM (CCR = 0.7456).

doi:10.1371/journal.pone.0168449.g008

two algorithms can obtain over-smoothness segmentations and lose details, especially for the CSF tissue. Meanwhile, both BAMM and BGGMM cannot well distinguish noisy pixels without considering any spatial information. GRFCM is not robust to noise, and the corresponding segmentations are not sufficiently smooth. By comparing the ground truth with the segmentations obtained with all algorithms, we see that the proposed algorithm visually obtains better results.

Segmentation accuracy for each tissue is measured in terms of DC values, and the results are listed in Table 2, which further demonstrates superior performance of the proposed algorithm. It is worth mentioning that the DC values of axial and sagittal CSF segmentation for BAMM and BGGMM are slightly higher than other algorithms because of the low percentage and abundant texture details of the CSF tissue. For any spatially constrained algorithm, the introduced spatial information can improve segmentation accuracy for images occupied by noise, but this can lead to smoothing the texture details in the image. Therefore, it is a dilemma to balance the trade-off between the anti-noise ability and over-smoothness for image textures. By comparing with two other spatially constrained algorithms (SCGM-EM and FRSCGMM), we see that the proposed algorithm achieves better trade-off and balance.

To statistically show the significance of the proposed algorithm, we apply the previous six algorithms to segment 10 axial, 10 sagittal and 10 coronal (from slice 86th to 95th) MR images for each noise level, where the level ranges from 3% to 9%. The statistical results (means and
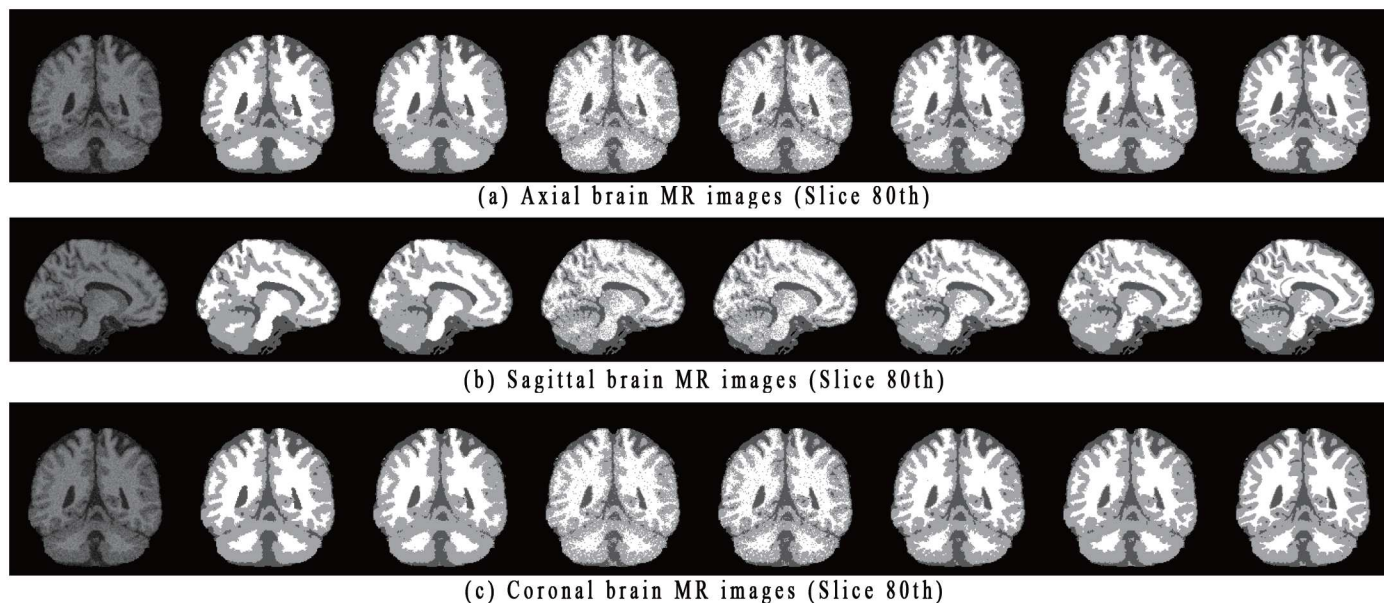
(a) Axial brain MR images (Slice 80th)

(b) Sagittal brain MR images (Slice 80th)

(c) Coronal brain MR images (Slice 80th)

Fig 9. Illustration of three simulated T1-weighted brain MR images with 9% noise and corresponding segmentation results obtained by each algorithm. In each subfigure, the images from left to right show: original image, segmentation results obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, proposed algorithm, and ground truth.

doi:10.1371/journal.pone.0168449.g009

Table 2. DC values of each tissue for the segmentations shown in Fig 9.

Algorithms	Tissue	Axial	Sagittal	Coronal	Average
SCGM-EM	GM	0.8854	0.8654	0.9162	0.8890
	WM	0.9090	0.9080	0.9308	0.9159
	CSF	0.8743	0.9032	0.9094	0.8956
FRSCGMM	GM	0.8692	0.8662	0.9075	0.8810
	WM	0.8810	0.9014	0.9137	0.8987
	CSF	0.8417	0.9033	0.9016	0.8822
BAMM	GM	0.8352	0.8268	0.8508	0.8376
	WM	0.8496	0.8522	0.8600	0.8539
	CSF	0.8814	0.9061	0.8707	0.8876
BGGMM	GM	0.8387	0.8241	0.8586	0.8405
	WM	0.8524	0.8554	0.8660	0.8579
	CSF	0.8814	0.9061	0.8921	0.8932
GRFCM	GM	0.8988	0.9135	0.9085	0.9069
	WM	0.8877	0.9140	0.9303	0.9107
	CSF	**0.8853**	**0.9101**	0.8964	0.8962
Proposed algorithm	GM	**0.9106**	**0.8934**	**0.9278**	**0.9106**
	WM	**0.9261**	**0.9235**	**0.9369**	**0.9288**
	CSF	0.8746	0.9040	**0.9126**	**0.8971**

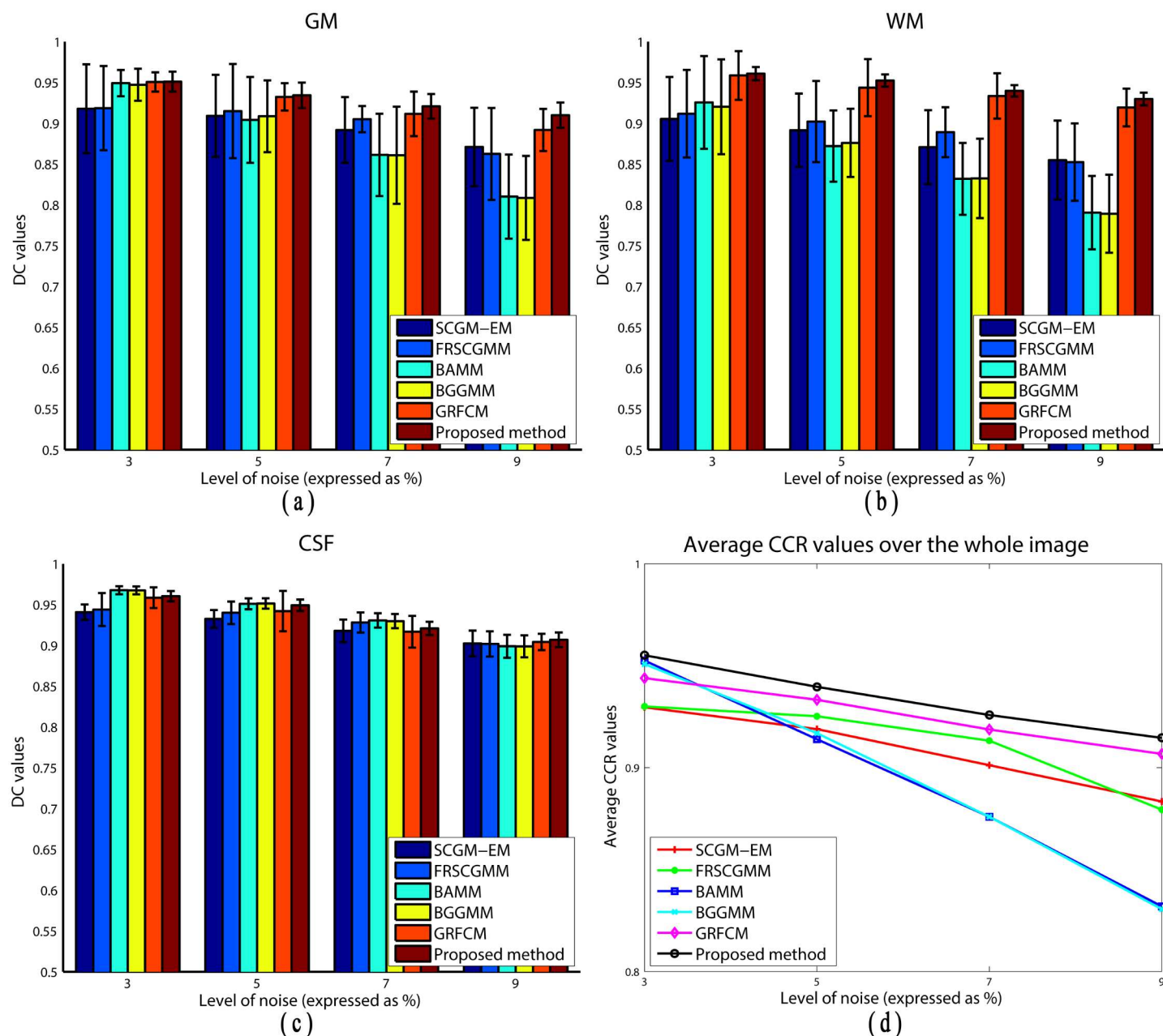doi:10.1371/journal.pone.0168449.t002

Fig 10. DC values for: (a) GM segmentation, (b) WM segmentation, (c) CSF segmentation, (d) CCR values over the entire images obtained by applying six segmentation algorithms to simulated brain MR images with increasing noise levels.

doi:10.1371/journal.pone.0168449.g010

standard deviations of DC values) are shown in Fig 10(a)–10(c). Moreover, the average CCR value over the entire segmentation results is show in Fig 10(d). From Fig 10, we can observe that the proposed algorithm produces more accurate segmentation (higher means) and has better robustness to noise (lower standard deviations). As indicated before, without considering any spatial information, BAMM and BGGMM produce satisfactory results when the noise level in the image is low (3%). SCGM-EM, FRSCGMM, and the proposed algorithm sacrifice segmentation accuracy for image details in orders to achieve better anti-noise ability.

Table 3. Paired-sample t-test results (p-value) of DC and CCR values between the five comparison methods and the proposed algorithm on all the testing images used in Fig 10.

Methods	SCGM-EM	FRSCGMM	BAMM	BGGMM	GRFCM
DC for GM	4.5e-05	2.1e-03	3.8e-20	2.7e-18	1.7e-02
DC for WM	1.9e-06	7.0e-05	2.8e-28	1.0e-25	2.9e-04
DC for CSF	4.1e-04	1.4e-02	1.7e-16	2.8e-15	6.3e-03
CCR	4.5e-06	1.8e-04	4.3e-31	3.9e-28	3.3e-03

doi:10.1371/journal.pone.0168449.t003

Therefore, the DC values of BAMM and BGGMM for the CSF tissue are slightly higher than the other three algorithms. By increasing the noise level, the performances of BAMM and BGGMM might decrease dramatically, especially for GM and WM tissues. Table 3 lists the statistical results (p-value of paired-sample t-test) between those five methods and the proposed algorithm on the testing images utilized in Fig 10. It is observed that, considering 0.05 as the level of significance, the proposed algorithm provides significantly better segmentation results with respect to both DC and CCR on the BrainWeb dataset.

In the next experiment, we apply all algorithms on the IBSR v2.0 data set [44], which contains 18 3D images. It is worth mentioning that BAMM [19] has been verified to outperform widely used algorithms, i.e., EMS [45] and SPM [48]. Therefore, in this comparison experiment, we only need to compare BAMM with the proposed algorithm in order to demonstrate the superior performance of the latter.

Fig 11 shows a 3D slice view of the real dataset (IBSR04). The image shown in Fig 11(b) is the ground truth of the original image. Fig 11(c)–11(h) show the results obtained by implementing the proposed method, GRFCM, SCGM-EM, FRSCGMM, BAMM, and BGGMM. It is obvious that case IBSR04 contains low contrast between the GM and CSF tissues. BAMM and BGGMM cannot well distinguish the GM and CSF tissues when low contrast occurs. SCGM-EM and FRSCGMM lead to over-smooth segmentations. Without estimation of the intensity inhomogeneity, GRFCM also fails to distinguish the tissues with low contrast. By compareing with these methods, we see that the effect of noise and low contrast on the final segmentation of our algorithm is small and has the most similarity with the ground truth.

Fig 12 shows the tissue surfaces of the segmentation results obtained by six algorithms (SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, and the proposed algorithm) on case IBSR12, and the corresponding ground truth. Fig 12(a) and 12(h) show the ground truth of the GM and WM surfaces, respectively. Fig 12(b)–12(g) show the GM surface obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, and the proposed method, respectively. Fig 12(i)–12(n) show the WM surface obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, and the proposed method, respectively. By comparing with the ground truth, it can be observed that our method obtains more accurate segmentation result.

Then we tested all those algorithms on 18 cases in the IBSR v2.0 dataset. The segmentation results were assessed in term of DC, and the variation of DC values was depicted in Fig 13. The statistical results, including mean, standard deviation (STD), and p-value of the t-test, of those methods on 18 3D brain images (from IBSR01 to IBSR18) were listed in Table 4. According to the obtained mean and STD, we can tell that the proposed algorithm steadily outperforms other five approaches. Based on the p-values, we find that, considering 0.05 as the level of significance, the proposed algorithm provides significantly more accurate segmentations on the IBSR v2.0 dataset than other five algorithm.
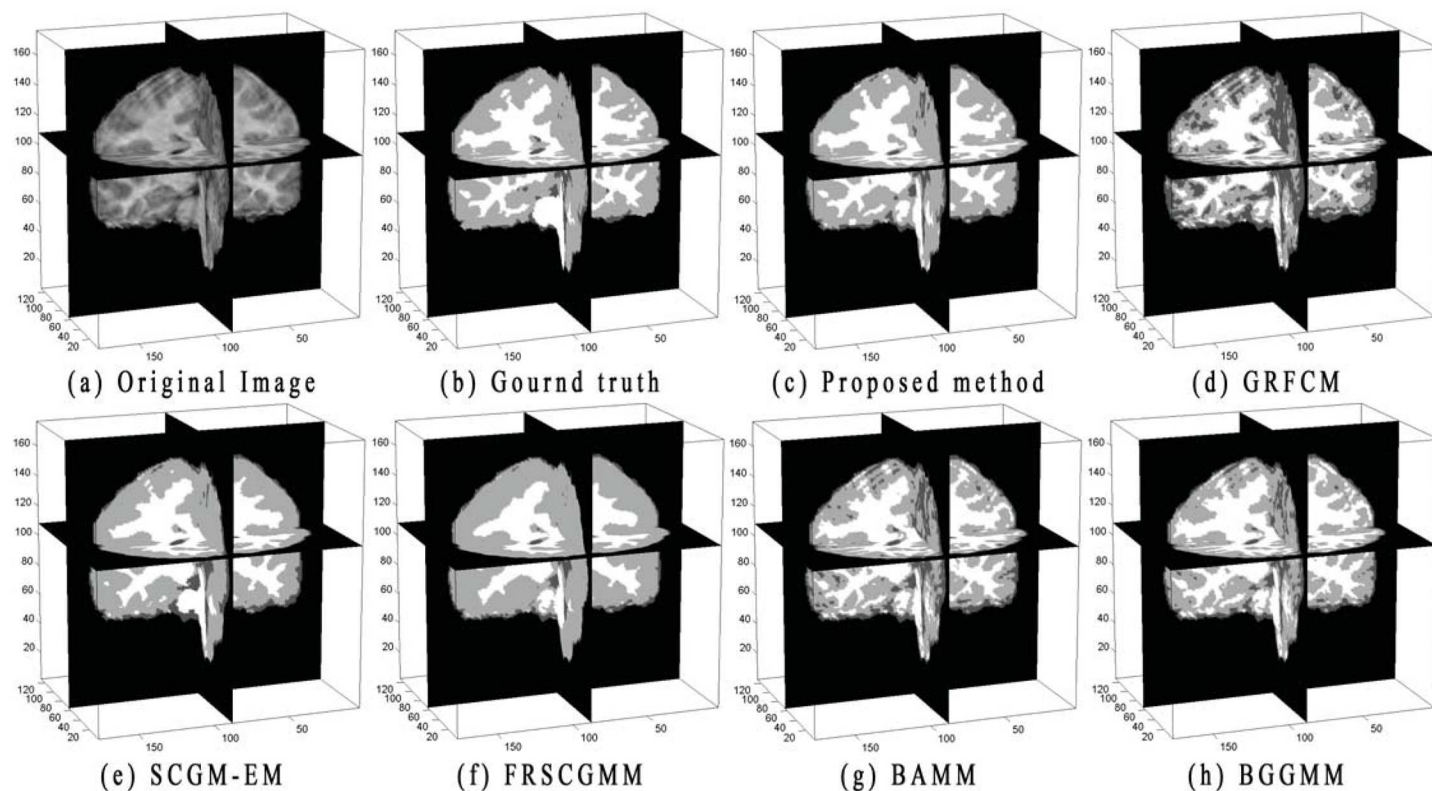
Fig 11. 3D slice view of the real dataset (IBSR04), corresponding ground truth and segmentations by applying the proposed method, GRFCM, SCGM-EM, FRSCGMM, BAMM, and BGGMM.

doi:10.1371/journal.pone.0168449.g011

Segmentation of color images

In this section, we test all comparison methods on color images in the Lab color space selected from Berkeley dataset. In Fig 14, we compare the segmentation results of three real-world color images. The first row shows the original images with image ID "105019,", "100007," and "28083," from left to right, and the corresponding number of clusters is 2, 3, and 4. The segmentation results obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM and the proposed SCGAGMM algorithm are shown from the second to sixth row.

The first image (ID: 105019) is segmented into two classes: "lions" and background. As shown in the first row of Fig 14, SCGM-EM, FRSCGMM, BAMM and BGGMM cannot extract the "lions" accurately from the background. Parts of the background region are misclassified as the target. In comparison, the proposed algorithm can successfully extract the target from the background. We attempted to segment the second image (ID: 100007) into three classes, and the proposed method obtains better classification results with more detail. The third image (ID: 374067) consists mainly of four color components: "grassland", "trees", "mountain"and "sky". All comparison methods cannot distinguish well the "trees"or "mountain"from "grassland". In comparison, our method can successfully segments all objects, especially for the "mountain"region, and does not result in obvious misclassifications.

Finally, a set of color images is tested to evaluate the performance of the proposed algorithm against the SCGM-EM, FRSCGMM, BAMM and BGGMM methods. Table 5 lists the PRI values obtained with all methods for 30 real world images. The statistical results (SR), including
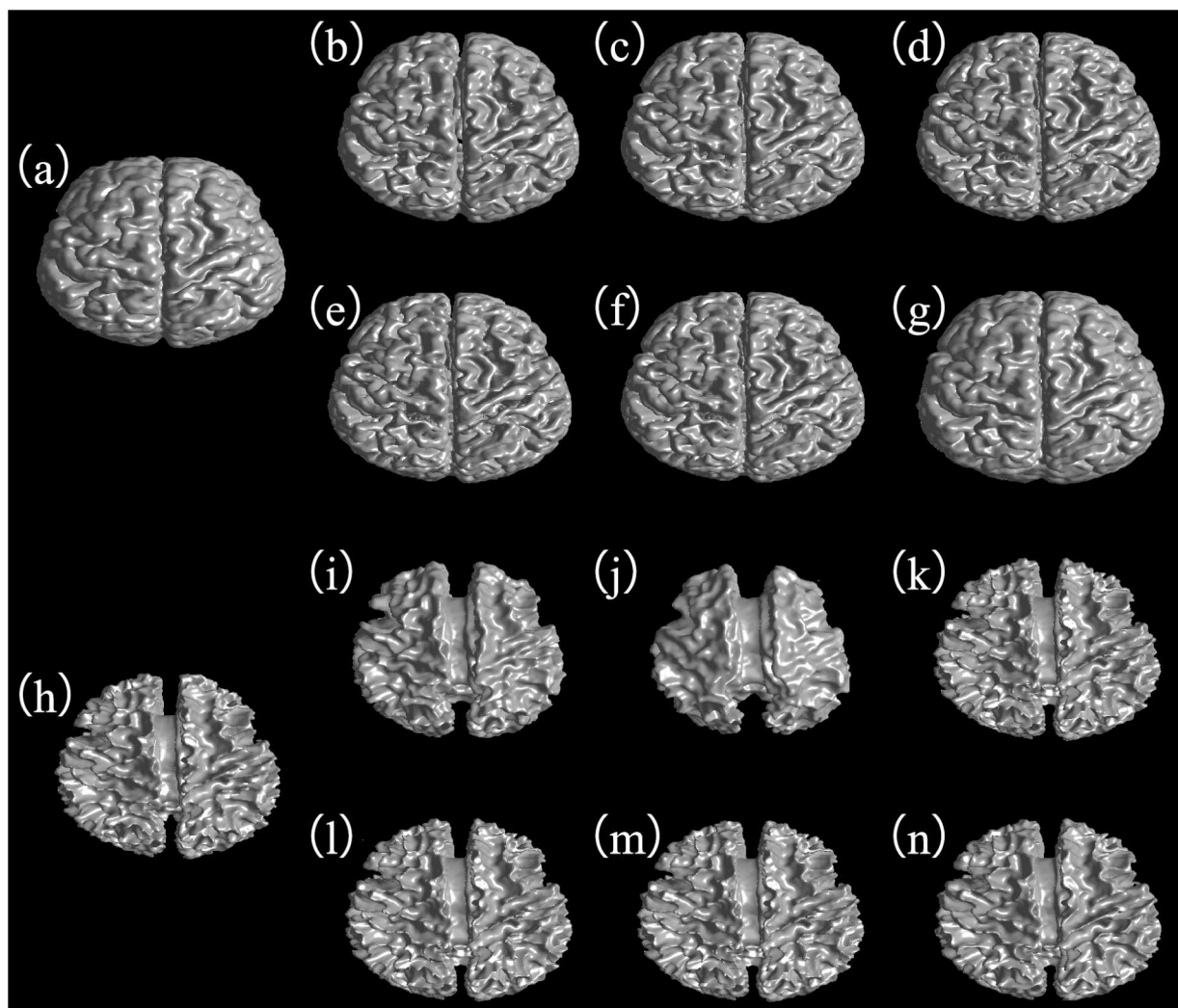
Fig 12. Example of tissue surfaces for case IBSR12. (a) and (h) show ground truth of GM and WM surfaces, respectively. (b) to (g) show GM surface obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, and the proposed method, respectively. (i) to (n) show the WM surface obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, GRFCM, and the proposed method, respectively.

doi:10.1371/journal.pone.0168449.g012

mean values, standard deviation (STD), and p-value for the t-test, are listed at the bottom of Table 5. From t-test results, we find that considering 0.05 as the level of significance, the proposed algorithm provides significantly better segmentations with respect to the PRI index. It is evident from the results that the proposed algorithm outperforms other methods with higher PRI values in most cases.

Comparison of computational complexity

The computational complexity of those five algorithms was compared in Table 6, where T is the number of iterations when the algorithm converges, N is number of pixels in an image, D is the dimension of each pixel, K is the number of clusters, L is the number of distributions and N_{∂_i} is the number of pixels in the neighborhood ∂_i. It should be noted that the computational complexity for the EM algorithm is of the order $O(NKD^2)$ for each iteration [49].
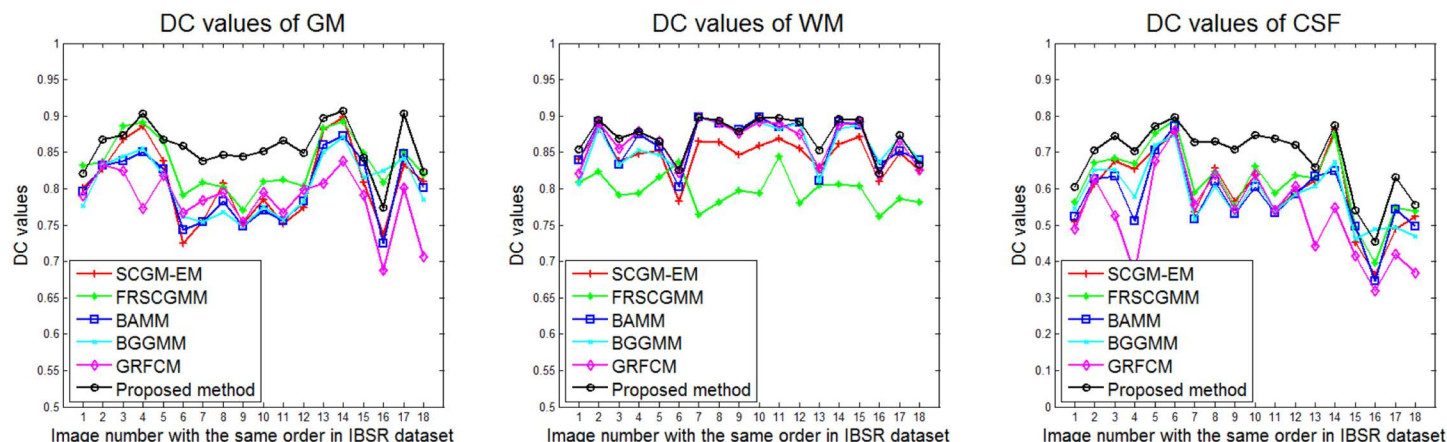
Fig 13. Performance of six segmentation algorithms on 18 benchmark data sets.

Table 4. Statistics of DC values (mean, standard deviation (STD) and p-value) obtained by applying six algorithms to 18 cases from the IBSR v2.0 dataset.

Algorithms	Statistics	GM	WM	CSF
SCGM-EM	Mean	0.8077	0.8470	0.5917
	STD	0.0531	0.0239	0.1051
	p-value	3.6e-05	1.0e-08	9.0e-07
FRSCGMM	Mean	0.8340	0.7986	0.6198
	STD	0.0373	0.0224	0.0975
	p-value	4.2e-03	3.6e-08	1.4e-05
BAMM	Mean	0.8014	0.8644	0.5738
	STD	0.0455	0.0313	0.0952
	p-value	6.6e-07	1.9e-02	3.4e-07
BGGMM	Mean	0.8037	0.8617	0.5826
	STD	0.0531	0.0239	0.1051
	p-value	8.3e-06	1.0e-02	1.8e-06
GRFCM	Mean	0.7849	0.8653	0.5270
	STD	0.0395	0.0292	0.1204
	p-value	6.8e-09	2.4e-03	3.7e-08
Proposed algorithm	Mean	0.8577	0.8662	0.6843
	STD	0.0336	0.0290	0.0925
	p-value	–	–	–

Table 6 also gives the converging time, number of iterations and time-cost per iteration of five algorithms obtained by applying each of them to 100 2D brain MR images, which have a size of 217×181 and were selected from the BrainWeb dataset. In this comparative experiment, we checked for the convergence of the parameter values, set the stopping criteria to $\varepsilon = 10^{-5}$ and executed each algorithm with 100 iterations (Intel Core i7-4790 CPU, 3.60GHz, 16GB RAM, 64-bit Windows 8, and Matlab Version 7.8). It is worth mentioning that the proposed algorithm was performed in the MATLAB environment without any particular code optimization.
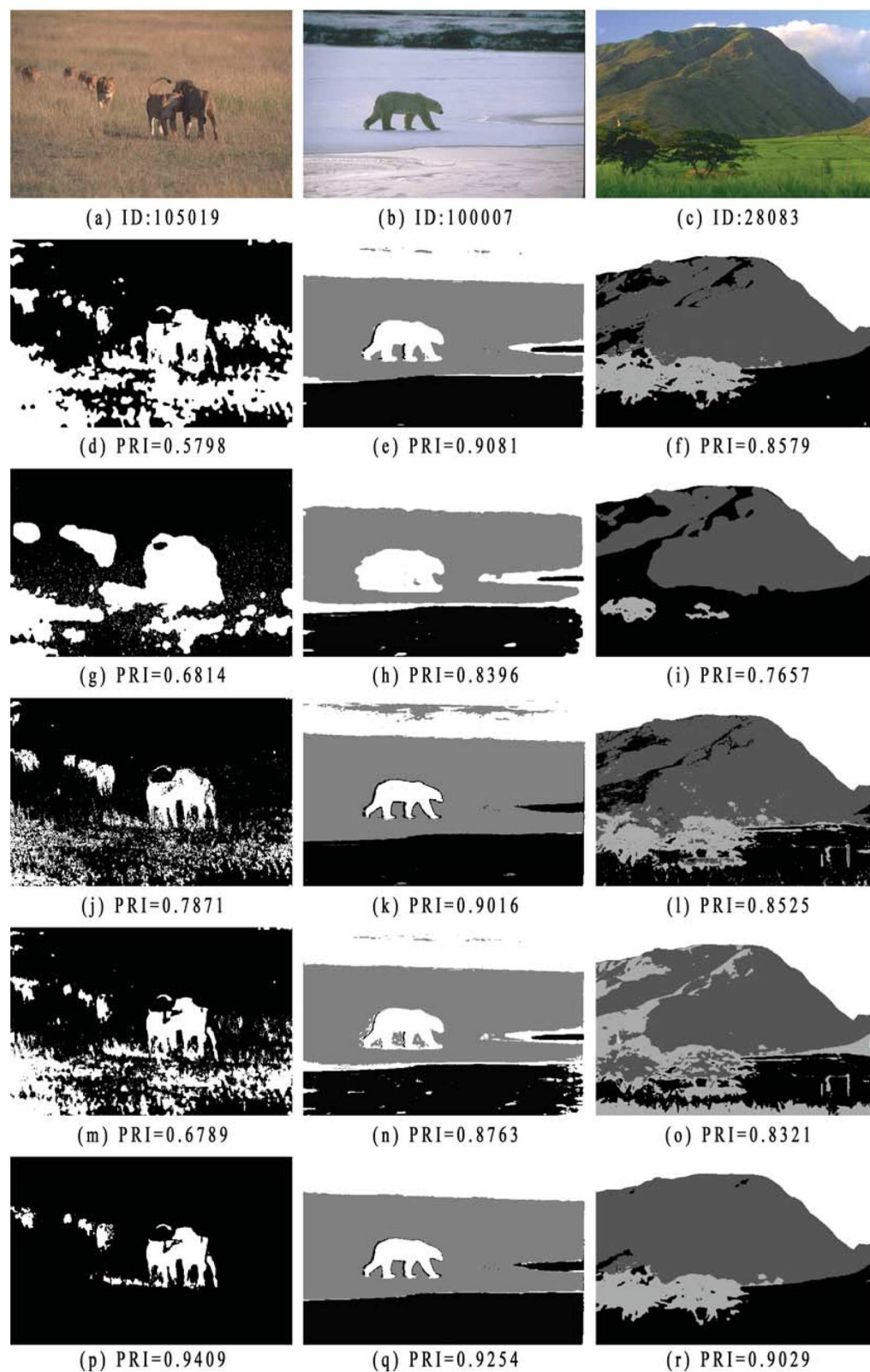
(a) ID:105019　　(b) ID:100007　　(c) ID:28083

(d) PRI=0.5798　　(e) PRI=0.9081　　(f) PRI=0.8579

(g) PRI=0.6814　　(h) PRI=0.8396　　(i) PRI=0.7657

(j) PRI=0.7871　　(k) PRI=0.9016　　(l) PRI=0.8525

(m) PRI=0.6789　　(n) PRI=0.8763　　(o) PRI=0.8321

(p) PRI=0.9409　　(q) PRI=0.9254　　(r) PRI=0.9029

Fig 14. Comparison of color image segmentations. The image IDs are: (a) 105019, (b) 100007, (c) 28083. Images from second to sixth row show segmentation results obtained by SCGM-EM, FRSCGMM, BAMM, BGGMM, and proposed algorithm.

doi:10.1371/journal.pone.0168449.g014

Table 5. PRI values of image segmentation results on Berkeley's color image dataset.

K	Image ID	SCGM-EM	FRSCGMM	BAMM	BGGMM	Proposed
2	3063	0.8584	0.8549	0.8372	0.8321	**0.8961**
	3096	0.8770	0.8747	0.7710	0.8720	**0.8779**
	80090	0.7445	0.6368	0.7182	0.7344	**0.7850**
	80099	0.6882	0.8810	0.8041	0.8314	**0.8912**
	105019	0.5798	0.6814	0.7871	0.6789	**0.8789**
	108073	0.6334	0.5603	0.5531	0.6451	**0.7408**
	124084	0.7781	0.7359	0.7467	0.6940	**0.7818**
	135069	0.9746	0.9727	0.9829	0.9829	**0.9841**
	147091	0.8501	0.8340	0.7896	0.7752	**0.8548**
	159091	0.7064	0.7129	0.6962	0.7414	**0.7793**
3	51084	0.6934	0.6834	0.6588	0.6846	**0.6950**
	76002	0.7527	0.7521	0.7667	0.7377	**0.7723**
	113009	0.6683	0.6686	0.6347	0.6108	**0.6725**
	134052	0.5819	0.5657	0.5636	0.5667	**0.7526**
	163014	0.7108	0.7158	0.7188	0.7014	**0.7280**
	176039	0.8345	0.7918	0.8155	0.8057	**0.8393**
	176051	0.7494	0.7431	0.7537	0.6026	**0.7744**
	183055	0.7806	0.7339	0.7843	0.7934	**0.8032**
	249061	0.8847	0.8794	**0.9151**	0.7839	0.9149
	253055	0.9641	0.9086	0.9619	0.7934	**0.9725**
4	14037	0.7634	0.8237	0.8058	0.8039	**0.8245**
	97010	0.8985	0.8940	0.8840	0.8135	**0.9014**
	106025	0.7941	0.7886	0.7943	**0.8340**	0.8223
	117025	0.7836	0.7720	0.7874	0.7852	**0.8112**
	163004	0.7538	0.7111	0.7319	0.7413	**0.7623**
	197017	0.8707	0.8694	0.8487	0.8636	**0.9086**
	198004	0.7587	0.6372	0.7408	0.7210	**0.7922**
	232038	0.8311	0.8125	**0.8382**	0.8030	0.8321
	241004	0.7954	0.7596	0.7923	0.7924	**0.8042**
	361084	0.8090	0.8067	0.8116	0.8002	**0.8273**
SR	Mean	0.7790	0.7687	0.7765	0.7609	**0.8227**
	STD	0.0976	0.1015	0.0970	0.0880	**0.0750**
	p-value	0.0014	2.0e-05	1.4e-05	2.7e-06	–

doi:10.1371/journal.pone.0168449.t005

Table 6. Computational complexity, converging time, number of iterations and per iteration time (average ± standard deviation, UNIT: Second) by applying five algorithms on BrainWeb dataset.

Algorithms	SCGM-EM	FRSCGMM	BAMM	BGGMM	Proposed
Computational complexity	$O(NKD^2TN_{\partial_i}^2)$	$O(NKD^2TN_{\partial_i}^2)$	$O(NKD^2TL)$	$O(NKD^2TL)$	$O(NKD^2TLN_{\partial_i}^2)$
Converging time	0.3804 ± 0.1090	0.9808 ± 0.3585	3.4747 ± 0.6914	3.4901 ± 1.4738	4.4978 ± 0.6397
Number of iterations	41.25 ± 11.90	109.70 ± 39.58	160.90 ± 25.06	38.40 ± 15.94	37.60 ± 19.29
Per iteration time	0.0092 ± 0.0001	0.0088 ± 0.0001	0.0207 ± 0.0005	0.0780 ± 0.0012	0.1391 ± 0.0519

doi:10.1371/journal.pone.0168449.t006

Conclusion

To overcome the limitations involved in most GMM-based algorithms, in this paper, we proposed a rough set bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation. Based on the rough set theory, a new bounded indicator function was proposed to determine the bounded support regions of the observed data. The bounded indicator and posterior probability of a pixel that belongs to each sub-region were estimated based on the rough regions. The within- and between-cluster spatial constraints were introduced by incorporating the spatial information with adaptively selected direction in order to reduce over-smoothness for segmentations. Experimental results demonstrated that the proposed algorithm is flexible to fit the data shapes, and robust to noise, which makes our method be capable of producing more accurate segmentation results comparing with several state-of-the-art algorithms. Future work will be devoted to reducing the complexity of the proposed algorithm.

Appendix

Mean value estimation

Considering the derivation of function $E(\Pi, \Theta|X)$ in Eq (32) with respect to μ_{kl}, we have

$$
\frac{\partial E(\Pi, \Theta|X)}{\partial \mu_{kl}} = -\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} \Bigg\{ -\Sigma_{kl}^{-1}(\mu_{kl} - x_i) \\
+ \Sigma_{kl}^{-1} \frac{\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)(\mu_{kl} - x)dx}{\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx} \Bigg\},
\tag{40}
$$

where the term $\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)(\mu_{kl} - x)dx$ is the expectation of function $\tilde{H}(x|\Omega_k) \times (\mu_{kl} - x)$ under distribution $\Phi(x|\mu_{kl}, \Sigma_{kl})$, which can be approximated as [14, 19, 36]:

$$
\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)(\mu_{kl} - x)dx \approx \frac{1}{M} \sum_{m=1}^{M} (\mu_{kl} - s_{mkl})\tilde{H}(s_{mkl}|\Omega_k),
\tag{41}
$$

where $s_{mkl} \sim \Phi(x|\mu_{kl}, \Sigma_{kl})$ is the random vector drawn from distribution $\Phi(x|\mu_{kl}, \Sigma_{kl})$, and M is the number of random vectors s_{mkl} [19]. In this paper, we set $M = 10^6$ for all experiments.

Similarly, the term $\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx$ can be approximated as [14, 19, 36]:

$$
\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx \approx \frac{1}{M} \sum_{m=1}^{M} \tilde{H}(s_{mkl}|\Omega_k).
\tag{42}
$$

Based on Eqs (41) and (42), $\partial E(\Pi, \Theta|X)/\partial \mu_{kl}$ from Eq (40) can be rewritten as:

$$
\frac{\partial E(\Pi, \Theta|X)}{\partial \mu_{kl}} = -\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} \Bigg\{ -\Sigma_{kl}^{-1}(\mu_{kl} - x_i) \\
+ \Sigma_{kl}^{-1} \frac{\sum_{m=1}^{M} (\mu_{kl} - s_{mkl})\tilde{H}(s_{mkl}|\Omega_k)}{\sum_{m=1}^{M} \tilde{H}(s_{mkl}|\Omega_k)} \Bigg\}.
\tag{43}
$$

The solution $\partial E(\Pi, \Theta|X)/\partial \mu_{kl} = 0$ yields the updating function for μ_{kl} during the iterations:

$$
\mu_{kl} = \frac{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} x_i}{\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl}} - \frac{\sum_{m=1}^{M} (\mu_{kl} - s_{mkl})\tilde{H}(s_{mkl}|\Omega_k)}{\sum_{m=1}^{M} \tilde{H}(s_{mkl}|\Omega_k)}.
\tag{44}
$$

Covariance matrix estimation

Considering the derivation of function $E(\Pi, \Theta|X)$ in Eq (32) with respect to Σ_{kl}, we have

$$
\begin{aligned}
\frac{\partial E(\Pi, \Theta|X)}{\partial \Sigma_{kl}} &= -\sum_{i=1}^{N} \tilde{z}_{ik} \tilde{y}_{ikl} \left\{ \frac{1}{2}\Sigma_{kl} - \frac{1}{2}(x_i - \mu_{kl})(x_i - \mu_{kl})^T \right. \\
&\left. - \frac{\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)\left(\frac{1}{2}\Sigma_{kl} - \frac{1}{2}(x_i - \mu_{kl})(x_i - \mu_{kl})^T\right)dx}{\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)dx} \right\}.
\end{aligned}
\tag{45}
$$

Similar to Eq (41), the term $\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)\left(\frac{1}{2}\Sigma_{kl} - \frac{1}{2}(x_i - \mu_{kl})(x_i - \mu_{kl})^T\right)dx$ in the Eq (45) can be approximated as:

$$
\begin{aligned}
&\int_{\Omega_k} \Phi(x|\mu_{kl}, \Sigma_{kl})\tilde{H}(x|\Omega_k)\left(\frac{1}{2}\Sigma_{kl} - \frac{1}{2}(x_i - \mu_{kl})(x_i - \mu_{kl})^T\right)dx \\
&\approx \frac{1}{2M}\sum_{m=1}^{M}\left(\Sigma_{kl} - (s_{mkl} - \mu_{kl})(s_{mkl} - \mu_{kl})^T\right)\tilde{H}(s_{mkl}|\Omega_k).
\end{aligned}
\tag{46}
$$

By setting $\partial E(\Pi, \Theta|X)/\partial \Sigma_{kl} = 0$, we obtain the updating function for Σ_{kl} during the iterations:

$$
\begin{aligned}
\Sigma_{kl} &= \frac{\sum_{i=1}^{N} \tilde{z}_{ik}\tilde{y}_{ikl}(x_i - \mu_{kl})(x_i - \mu_{kl})^T}{\sum_{i=1}^{N} \tilde{z}_{ik}\tilde{y}_{ikl}} \\
&- \frac{\sum_{m=1}^{M}\left((s_{mkl} - \mu_{kl})(s_{mkl} - \mu_{kl})^T - \Sigma_{kl}\right)\tilde{H}(s_{mkl}|\Omega_k)}{\sum_{m=1}^{M}\tilde{H}(s_{mkl}|\Omega_k)}.
\end{aligned}
\tag{47}
$$

Prior probability estimation

Because the prior distribution of between-cluster π_{ik} satisfies constraint $\sum_{k=1}^{K} \pi_{ik} = 1$, Lagrange's multiplier τ_i is used to enforce these constraints for each data point:

$$
\frac{\partial}{\partial \pi_{ik}}\left[E - \sum_{i=1}^{N}\tau_i\left(\sum_{k=1}^{K}\pi_{ik} - 1\right)\right] = 0.
\tag{48}
$$

Constraint $\sum_{k=1}^{K} \pi_{ik} = 1$ allows:

$$
\pi_{ik} = \frac{\tilde{z}_{ik} + E_{ik}}{\sum_{h=1}^{K}(\tilde{z}_{ih} + E_{ih})}.
\tag{49}
$$

Similarly, we obtain the updating function for the prior distribution of within-cluster η_{ikl} under constraint $\sum_{l=1}^{L} \eta_{ikl} = 1$:

$$
\eta_{ikl} = \frac{\tilde{y}_{ikl} + F_{ikl}}{\sum_{h=1}^{L}(\tilde{y}_{ikh} + F_{ikh})}.
\tag{50}
$$

Acknowledgments

Author Contributions

Conceptualization: ZJ.

Data curation: ZJ YH QS.

Formal analysis: ZJ YH GC YZ.

Funding acquisition: ZJ.

Investigation: ZJ YH.

Methodology: ZJ GC QS.

Project administration: ZJ QS.

Resources: ZJ YZ.

Software: ZJ YH.

Supervision: GC QS YZ.

Validation: GC QS YZ.

Visualization: ZJ YH.

Writing – original draft: ZJ YH.

Writing – review & editing: GC QS YZ.

References

1. Gu B, Sun X, Sheng VS. Structural Minimax Probability Machine. IEEE Transactions on Neural Networks and Learning Systems. 2016; doi: 10.1109/TNNLS.2016.2544779 PMID: 27101618

2. Wen XZ, Shao L, Xue Y, Fang W A rapid learning algorithm for vehicle classification. Information Sciences. 2015 Feb; 295(1):395–406. doi: 10.1016/j.ins.2014.10.040

3. Gu B, Sheng VS. A Robust Regularization Path Algorithm for v-Support Vector Classification. IEEE Transactions on Neural Networks and Learning Systems. 2016; doi: 10.1109/TNNLS.2016.2527796 PMID: 26929067

4. Gu B, Sheng VS, Li S. Bi-parameter space partition for cost-sensitive SVM. Proceedings of the 24th International Conference on Artificial Intelligence. AAAI Press. 2015:3532–3539.

5. Gu B, Sheng VS, Wang ZJ, Ho D, Osman S, Li S. Incremental learning for v-Support Vector Regression. Neural Networks. 2015 Jul; 67:140–150. doi: 10.1016/j.neunet.2015.03.013 PMID: 25933108

6. Cremers D, Rousson M, Deriche R. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International Journal of Computer Vision. 2007 Feb; 72 (2):195–215. doi: 10.1007/s11263-006-8711-1

7. Li C, Xu C, Gui C, Fox MD. Distance Regularized Level Set Evolution and Its Application to Image Segmentation. IEEE Transactions on Image Processing. 2010 Dec; 19(12):430–436. doi: 10.1109/TIP.2010.2069690 PMID: 20801742

8. Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000 Aug; 22(8):888–905. doi: 10.1109/34.868688

9. Grady L. Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006 Nov; 28(11)1768–1783. doi: 10.1109/TPAMI.2006.233 PMID: 17063682

10. Zheng Y, Jeon B, Xu D, Wu QMJ, Zhang H. Image segmentation by generalized hierarchical fuzzy C-means algorithm. Journal of Intelligent and Fuzzy Systems. 2015 Feb; 28 (2):961–973.

11. Li C, Gore JC, Davatzikos C. Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. Magnetic Resonance Imaging. 2014 Jul; 32(7):913–923. doi: 10.1016/j.mri.2014.03.010 PMID: 24928302

12. Li J, Li X, Yang B, Sun X. Segmentation-based Image Copy-move Forgery Detection Scheme. IEEE Transactions on Information Forensics and Security. 2015 Mar; 10(3)507–518. doi: 10.1109/TIFS.2014.2381872

13. McLachlan G, Peel D. Finite Mixture Models. New York: Wiley; 2000.

14. Bishop CM Pattern Recognition and Machine Learning New York: Springer; 2006.

15. Titterington DM, Smith AFM, Makov UE. Statistical Analysis of Finite Mixture Distributions. Hoboken, New York: Wiley; 1985.

16. Jain AK, Duin RPW, Mao JC. Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000 Jan; 22(1):4–37. doi: 10.1109/34.824819

17. Nguyen TM, Wu QMJ. A Nonsymmetric Mixture Model for Unsupervised Image Segmentation. IEEE Transactions on Cybernetics. 2013 Feb; 43(2):751–765. doi: 10.1109/TSMCB.2012.2215849 PMID: 22987532

18. McLachlan GJ, Krishnan T. The EM Algorithm and Extensions. Wiley Series in Probability and Statistics, New York: Wiley; 1997.

19. Nguyen TM, Wu QMJ, Mukherjee D, Zhang H. A Bayesian Bounded Asymmetric Mixture Model with Segmentation Application. IEEE Journal of Biomedical and Health Informatics. 2014 Jan; 18(1):109–119. doi: 10.1109/JBHI.2013.2264749 PMID: 24403408

20. Clifford P. Markov random fields in statistics. In Geoffrey Grimmett and Dominic Welsh, editors, Disorder in Physical Systems: A Volume in Honour of John M. Hammersley, pp. 19–32. Oxford University Press; 1990.

21. Forbes F, Peyrard N. Hidden Markov random field model selection criteria based on mean field-like approximations. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003 Sep; 25 (9):1089–1101. doi: 10.1109/TPAMI.2003.1227985

22. Celeux G, Forbes F, Peyrard N. EM procedures using mean field-like approximations for Markov model-based image segmentation. Pattern Recognition. 2003 Jan; 36(1):131–144. doi: 10.1016/S0031-3203(02)00027-4

23. Nguyen TM, Wu QMJ Fast and Robust Spatially Constrained Gaussian Mixture Model for Image Segmentation IEEE Transactions on Circuits and Systems for Video Technology. 2013 Apr; 23(4):621–635. doi: 10.1109/TCSVT.2012.2211176

24. Sanjay GS, Hebert TJ. Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. IEEE Transactions on Image Processing. 1998 Jul; 7(7):1014–1028. doi: 10.1109/83.701161 PMID: 18276317

25. Blekas K, Likas A, Galatsanos NP, Lagaris IE. A spatially constrained mixture model for image segmentation. IEEE Transactions on Neural Networks. 2005 Feb; 16(2):494–498. doi: 10.1109/TNN.2004.841773 PMID: 15787156

26. Diplaros A, Vlassis N, Gevers T. A Spatially Constrained Generative Model and an EM Algorithm for Image Segmentation. IEEE Transactions on Neural Networks. 2007 Mar; 18(3):798–808. doi: 10.1109/TNN.2007.891190 PMID: 17526345

27. Nikou C, Galatsanos N, Likas A. A class-adaptive spatially variant mixture model for image segmentation. IEEE Transactions on Image Processing. 2007 Apr; 16(4):1121–1130. doi: 10.1109/TIP.2007.891771 PMID: 17405442

28. Peel D, McLachlan G. Robust mixture modeling using the t distribution. Stat. Comput. 2000 Apr; 10 (4):339–348. doi: 10.1023/A:1008981510081

29. Xin W, Zhen Y. The infinite Students t-factor mixture analyzer for robust clustering and classification. Pattern Recognition. 2012 Dec; 45(12):4346–4357. doi: 10.1016/j.patcog.2012.05.003

30. Scallan AJ. Maximum likelihood estimation for a normal/Laplace mixture distribution. Journal of the Royal Statistical Society: Series D. 1992 Feb; 41(2):227–231.

31. Franczak BC, Browne RP, McNicholas P. Mixtures of Shifted Asymmetric Laplace Distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014 Jun; 36(6):1149–1157. doi: 10.1109/TPAMI.2013.216 PMID: 26353277

32. Allili MS, Ziou D, Bouguila N, Boutemedjet S. Image and video segmentation by combining unsupervised generalized Gaussian mixture modeling and feature selection. IEEE Transactions on Circuits and Systems for Video Technology. 2010 Oct; 20(10):1373–1377. doi: 10.1109/TCSVT.2010.2077483

33. Tarek E, Nizar B. Bayesian learning of finite generalized Gaussian mixture models on images. Signal Processing. 2011 Apr; 91(4):801–820. doi: 10.1016/j.sigpro.2010.08.014

34. Zhang H., Wu QMJ, Nguyen TM. Image Segmentation by a Robust Modified Gaussian Mixture Model. The 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP).2013;1478–1482.

35. Browne RP, McNicholas PD, Sparling MD. Model-Based Learning Using a Mixture of Mixture of Gaussian and Uniform Distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012 Apr; 34(4):814–817. doi: 10.1109/TPAMI.2011.199 PMID: 22383342

36. Lindblom J, Samuelsson J. Bounded support Gaussian mixture modeling of speech spectra. IEEE Trans. Speech Audio Processing. 2003 Jan; 11(1):88–99. doi: 10.1109/TSA.2002.805639

37. Nguyen TM, Wu QMJ, Zhang H. Bounded Generalized Gaussian Mixture Model. Pattern Recognition. 2014 Sep; 47(9):3132–3142. doi: 10.1016/j.patcog.2014.03.030

38. Nguyen TM, Wu QMJ. Bounded Asymmetrical Student's-t Mixture Model. IEEE transactions on cybernetics. 2014 Jun; 44(6):857–869. doi: 10.1109/TCYB.2013.2273714 PMID: 23893763

39. Ji ZX, Sun QS, Xia Y, Chen Q, Xia DS, Feng DG. Generalized rough fuzzy c-means algorithm for brain MR image segmentation. Computer Methods and Programs in Biomedicine. 2012 Feb; 108(2):644–655. doi: 10.1016/j.cmpb.2011.10.010 PMID: 22088865

40. Pawlak Z. Rough Sets, Theoretical Aspects of Reasoning about Data. Netherlands: Kluwer, Dordrecht; 1991.

41. Hirano S, Tsumoto S. Rough representation of a region of interest in medical images. International Journal of Approximate Reasoning. 2005 Jan; 40(1):23–34. doi: 10.1016/j.ijar.2004.11.008

42. Figueiredo M, Jain AK. Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002 Mar; 24(3):381–396. doi: 10.1109/34.990138

43. Kwan RKS, Evans AC, Pike GB. MRI simulation-based evaluation of image-processing and classification methods. IEEE Transactions on Medical Imaging. 1999 Nov; 18(11):1085–1097. doi: 10.1109/42.816072 PMID: 10661326

44. Rohlfing T. Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable. IEEE Transactions on Medical Imaging. 2012 Feb; 31(2):153–163. doi: 10.1109/TMI.2011.2163944 PMID: 21827972

45. Leemput KV, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging. 1999 Oct; 18(10):897–908. doi: 10.1109/42.811270 PMID: 10628949

46. Martin D, Fowlkes C, Tal D, Malik J. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. 2001;2:416–423.

47. Unnikrishnan R., Pantofaru C., Hebert M. A measure for objective evaluation of image segmentation algorithms. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops. 2005;3:34–41.

48. Ashburner J, Friston KJ. Unified segmentation NeuroImage. 2005 Mar; 26(3):839–851. doi: 10.1016/j.neuroimage.2005.02.018 PMID: 15955494

49. Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review. 1984 Feb; 26(2):195–239. doi: 10.1137/1026034