A Robust Spatial Information-Theoretic GMM Algorithm for Bias Field Estimation and Brain MRI Segmentation

YUNJIE CHEN, MAO CAI, XINZE ZHOU, CHENG NING, CHUNZHENG CAO, AND JIANWEI YANG

School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding author: Yunjie Chen (priestcyj@nuist.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant 61672291 and Grant 61976241, in part by the Six Talent Climb Foundation of Jiangsu Province of China under Grant SWYY-034, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20191394, and in part by the Key Research Project of Jiangsu Province of China under Grant 2019A005.

ABSTRACT Due to their simplicity and flexibility, the unsupervised statistical models such as Gaussian mixture model (GMM) are powerful tools to address the brain magnetic resonance (MR) images segmentation problems. However, the GMM is based only on the intensity information, which makes it sensitive to noise. Lots of GMM-based segmentation algorithms have the low segmentation accuracy because of the influence caused by noise and intensity inhomogeneity. To further improve the segmentation accuracy, a robust spatial information-theoretic GMM algorithm is proposed in this paper to simultaneously estimate the intensity inhomogeneity and segment brain MR images. First, a novel spatial factor containing the non-local spatial information is incorporated to reduce the impact of noise. The proposed spatial factor not only takes into account the local neighboring information, but also considers the spatial structure information of pixels. Thus, our algorithm can retain more image details while reducing the influence of noise. Second, the mutual information (MI) maximization method is used to identify and eliminate outliers. Finally, to overcome the impact of intensity inhomogeneity, we use a linear combination of a set of orthogonal polynomials to approximate the bias field. The objective function is integrated with the bias field estimation model to segment the images and estimate the bias field simultaneously. The experimental results on both synthetic and clinical brain MR images show that the proposed algorithm can overcome the influence of noise and intensity inhomogeneity, and achieve more accurate segmentation results.

INDEX TERMS Brain MRI segmentation, Gaussian mixture model, Markov random fields, non-local spatial information, information-theoretic approach, intensity inhomogeneity.

I. INTRODUCTION

Brain magnetic resonance (MR) images have been widely used for diagnosis of brain diseases because of high contrast, high spatial resolution and multi-spectral characteristics. Accurate segmentation of major brain tissues into the gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) is a key technology and important link in the study of many brain disorders. Hence brain MRI segmentation has attracted extensive interests. However, MR images generally contain the noise and a smoothly varying bias field, which is also referred to as the intensity inhomogeneity [1]. Most of existing intensity-based segmentation methods cannot achieve accurate segmentation results because of the noise and the intensity inhomogeneity. In recent years, several methods were proposed and developed for accurate segmentation [2]–[6], [30]–[33]. These methods involve algorithms with respect to thresholding process, region based, clustering methods, soft segmentation, the matting.

The statistical model-based techniques are successfully applied in image segmentation because of the simplicity and efficiency. In the model-based algorithms, standard Gaussian mixture model (GMM) [8] is well known and has been widely utilized in image segmentation. The GMM is also generally...
utilized and modified to correct intensity inhomogeneity [9]–[11] and its parameters can be estimated by the expectation maximization (EM) algorithm [7]. Wells et al. [9] proposed an adaptive segmentation algorithm for MRI data (AS-EM), which uses the EM algorithm to simultaneously correct bias field and segment brain MR images. Van Leemput et al. [10] further developed an automated bias field correction method (AM-EM) based on EM algorithm and utilized a digital brain atlas to provide the prior probability maps of brain tissues. However, the traditional GMM supposes that the pixels of an image are independent and less sensitive to noise. Blekas et al. [12] proposed a spatially constrained mixture model for image segmentation (SVFMM), which imposes spatial smoothness constraints between neighboring pixels. Diplaros et al. [13] proposed a novel generative GMM and an EM algorithm (SCGM-EM) for image segmentation. The SCGM-EM assumes that the hidden class labels of the pixels are generated by prior distributions, which share similar parameters within neighbor pixels and can reduce the effect of noise. To retain more image details, Nikou et al. [14] proposed a class-adaptive spatially variant mixture model (CA-SVFMM) which can adaptively determine the spatial directions. CA-SVFMM uses MRF to model the joint prior distribution of each pixel label. However, the M-step of EM algorithm cannot be applied directly to estimate the model parameters from the observations. In order to deal with this problem, Nguyen and Wu [15] proposed a fast and robust spatially constrained GMM (FRSCGMM). It introduced a novel spatial factor into the prior distribution and can directly apply the EM algorithm to estimate parameters. Although these models work well for noisy image, most MRF-based algorithms still have some drawbacks that high computational complexity and difficulty to preserve detail information.

Retaining image details and reducing the influence of noise simultaneously are contradictory. On one hand, the smoothing parameter should be chosen small enough to prevent the image from losing much of its sharpness and details. On the other hand, the value of the smoothing parameter has to be chosen large enough to restore images corrupted by noise and outliers. So, if we want to overcome the effect of the noise and preserve more detail information simultaneously, we can choose the small smoothing parameter to retain more image details and use some methods to identify/eliminate outliers or high-level noises which are not smoothed out. Song [16] proposed a robust information clustering algorithm, which used the mutual information (MI) maximization method to evaluate the reliability of input data and identify outliers. However, it only gave an intuitive discussion without a detailed theoretical demonstration. Wang et al. [17], [18] developed the MI maximization method into a complete theorem, called the information-theoretic approach for high level outlier identification. The experimental results have shown that the MI maximization is indeed a powerful tool to identify outliers.

Motivated by the aforementioned observations, we proposed a robust spatial information-theoretic GMM algorithm for bias field estimation and brain MRI segmentation. To reduce the effect of noise and preserve more details, the non-local spatial information is introduced. The smoothing parameter can be chosen small enough to retain more image spatial structure information. In the end of experiment part, we analyze the effect of the smoothing parameter. The selection of the smoothing parameter is usually dictated by the considered application of the model. To identify and eliminate outliers, we extend the MI maximization method to the mixture model and achieve more smooth and accurate results. By combining these two approaches, the model can retain more image details while reducing the influence of noise. In order to reduce the influence of intensity inhomogeneity, we use a linear combination of a set of orthogonal polynomials to estimate the bias field. The objective function is then integrated with the bias field estimation to make the model estimate the bias field, meanwhile segmenting images. The experiment results on synthetic and clinical brain MR images illustrate the superior performance of the proposed model compared with other state-of-the-art segmentation methods.

II. PRELIMINARIES

A. BIAS FIELD ESTIMATION

The observed MRI signal $I$ can be modeled as

$$I = BJ + n$$

where $I$ is the observed MR image, $B$ is an unknown bias field, $J$ is the true image to be restored, and $n$ is an additive noise [22]. Within the truth signal $J$, the intensity of pixels in the same tissue have a specific value, reflecting the physical property being measured. The aim of bias field estimation is to eliminate the bias field $B$ from the observed image and correct the intensity inhomogeneity. Classically, the bias field $B$ is assumed to be smooth and slowly varying in small local region. We usually use a set of orthogonal polynomials to approximate the bias field. In this paper, we use two-dimensional Legendre polynomials to approximate the bias field $B$ and the number of polynomials is set to 3.

B. THE CLASSICAL GAUSSIAN MIXTURE MODEL

We set image $I = \{x_i\} i = 1, 2, 3 \ldots N$, where $x_i$ with dimension $D$ is the $i$th pixel of image $I$ and $N$ is the number of total pixels. Let the neighborhood of the $i$th pixel be presented by $\delta_i$. Labels are denoted by $(\Omega_1, \Omega_2, \ldots, \Omega_K)$. GMM [19] assumes that each observation $x_i$ is considered independent of the label $\Omega_k, k = 1, 2, \ldots, K$. So we can segment the whole image consisting of $N$ pixels into $K$ labels. The density
function of GMM is given by
\[ f(x_i | \Pi, \theta) = \sum_{k=1}^{K} \pi_k \Phi(x_i | \theta_k), \]
where \( \Pi = \{ \pi_i \}, i = (1, 2, \ldots, N), k = (1, 2, \ldots, K) \) is the set of prior probabilities for the pixel \( x_i \) in label \( \Omega_k \), which satisfies the constraint \( 0 \leq \pi_{ik} \leq 1 \) and \( \sum_{k=1}^{K} \pi_{ik} = 1 \). \( \Phi(x_i | \theta_k) \) is the Gaussian distribution, where \( \theta_k = (\mu_k, \Sigma_k) \).

Each Gaussian distribution called a component of the mixture can be written in the form:
\[ \Phi(x_i | \theta_k) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|^2}} \exp\left(-\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)\right), \]
where \( \mu_k \) with \( D \) is the mean vector, the \( D \times D \) matrix \( \Sigma_k \) is the covariance, and \( |\Sigma_k| \) denotes the determinant of \( \Sigma_k \). Since each data point \( x_i \) is modeled as statistically independent, the joint probability density of the image data \( I = \{ x_i \} i = (1, 2, 3 \ldots N) \) can be modeled as
\[ p(X | \Pi, \theta) = \prod_{i=1}^{N} f(x_i | \Pi, \theta) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_{ik} \Phi(x_i | \theta_k). \]

However, the classical Gaussian mixture model only considers the intensity distribution, which makes it sensitive to noise without any spatial information.

C. GAUSSIAN MIXTURE MODELS BASED ON MRF
To reduce the effect of noise, the Markov random field (MRF) has been widely utilized into the GMM. The spatial information is generally incorporated among prior values by utilizing the MRF distribution:
\[ p(\Pi) = Z^{-1} \exp(-\frac{1}{T} U(\Pi)), \]
where \( Z \) is a normalizing constant, \( T \) is a temperature constant, and \( U(\Pi) \) is the smoothing prior. Based on Bayes’s rules, the posterior probability density function can be written as
\[ p(\Pi, \theta | X) \propto p(X | \Pi, \theta) p(\Pi). \]
So the log-likelihood function can be written as
\[ L(\Pi, \theta | X) = \log(p(\Pi, \theta | X)) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_{ik} \Phi(x_i | \theta_k)\right) + \log(p(\Pi)) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} \pi_{ik} \Phi(x_i | \theta_k)\right) - \log Z - \frac{1}{T} U(\Pi). \]

We can obtain various models by choosing different energy \( U(\Pi) \). By adopt different kinds of \( U(\Pi) \), many MRF-based mixture models have been successfully applied to image segmentation. In SVFMM [17], \( Z \) and \( T \) in (7) are set to one, and the spatial information is taken into account by setting \( U(\Pi) \) as
\[ U(\Pi) = \beta \sum_{i=1}^{N} \sum_{m \in \partial_i} \sum_{k=1}^{K} (\pi_{ik} - \pi_{mk})^2, \]
where \( \beta \) is a constant value. In CA-SVFMM [14], \( U(\Pi) \) is given by
\[ U(\Pi) = \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{s} \left[ \frac{1}{2} \log \beta_{ks}^2 - \frac{1}{2} \left( \sum_{m \in \partial_i} (\pi_{ik} - \pi_{mk})^2 \right) \right], \]
where \( S \) is the total number of the considered directions and \( \beta_{ks} \) is a variable parameter. In general, \( S \) is equal to four (4: horizontal, vertical, and two diagonal directions).

However, introducing spatial information in the above way adds computational complexity. Due to the complexity of the log-likelihood function, the M-step of EM algorithm cannot be applied directly to the prior distribution \( \pi_{ik} \). Thus, the resulting algorithms are computationally complex and utilize large amounts of computational power to solve the constrained optimization problem of the prior distribution \( \pi_{ik} \).

In order to overcome the disadvantage of these algorithms, Nguyen and Wu [15] introduced a novel factor \( G_{ik} \) defined as
\[ G_{ik}^{(t)} = \exp\left(\frac{\beta}{2\gamma_{ik}} \sum_{m \in \partial_i} (\pi_{mk}^{(t)} + \pi_{mk}^{(t)})\right), \]
where \( \beta \) is the temperature value that controls the smoothing prior and \( z_{ik} \) is the posterior probability. \( t \) indicates the iteration step. Then the new smoothing prior \( U(\Pi) \) is defined as
\[ U(\Pi) = -\sum_{i=1}^{N} \sum_{k=1}^{K} G_{ik}^{(t)} \log(\pi_{ik}^{(t+1)}). \]

The main advantage of \( G_{ik} \) is the ease of implementation, and incorporation of the spatial relationships among neighboring pixels in a simpler metric. Actually, the factor \( G_{ik} \) plays a role as a linear filter for smoothing and restoring images corrupted by noise. However, the drawback of this energy \( U(\Pi) \) is that it may cause over smoothing for the segmentation and lose the image details especially for the CSF segmentation of brain MR images.

D. MI MAXIMIZATION FOR OUTLIER FURTHER ELIMINATION
In order to eliminate the high-level outlier, Wang et al. [17], [18] proposed a novel adaptive spatial information-theoretic fuzzy clustering algorithm (ASIFC). The ASIFC algorithm is a two-step image segmentation algorithm and its second step is the high level outlier elimination by MI maximization. The MI maximization method in information theory provides a way to reassess the reliability of input data. The objective function of ASIFC algorithm is
\[ J_{ASIFC} = \sum_{i=1}^{N} \sum_{k=1}^{K} e_{ik}^q d_i(x_i, w_k) \quad 1 \leq q < \infty, \]
where $e_i$ is the estimated input data distribution $p(x_i)$ (with constraint $\sum_{i=1}^{N} e_i = 1$). $u_{k|j} = p(w_k|x_i)$ is the membership of assigning observation $x_i$ to cluster center $w_k$ and $d_j(x_i, w_k)$ is an improved similarity measure. After generating the estimated cluster center $w_k$ and membership value $u_{k|j}$, we can evaluate the input data reliability. Highly unreliable data points need to be labeled as outliers and are going to be clustered again with a bigger neighboring window. We can obtain a set of new robust input data reliability values $e_i$ for every input data points by using the information-theoretic method. $e_i$ can be used to label/eliminate the remaining unreliable data points (high level outliers) and thus improve the clustering/segmentation results of the algorithm.

To derive the robust estimation of input data distribution $e_i$, ASIFC defines the mutual information between the input data distribution and the optimal membership distribution as follows:

$$I(X, W) = \sum_{i=1}^{N} \sum_{k=1}^{K} p(x_i, w_k) \log \left( \frac{p(x_i, w_k)}{p(x_i)p(w_k)} \right)$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} e_{i}N_{k}^{i} \log \left( \frac{u_{k|j}}{\sum_{i=1}^{N} e_{i}u_{k|j}} \right).$$

(13)

Considering $\sum_{i=1}^{N} e_i = 1$ limitation, Lagrange’s method is used. The robust estimation of $e_i$ is given by

$$e_i = \frac{e_i \exp \left( \sum_{k=1}^{K} u_{k|j} \log \left( \frac{u_{k|j}}{\sum_{i=1}^{N} e_{i}u_{k|j}} \right) \right)}{\sum_{i=1}^{N} \exp \left( \sum_{k=1}^{K} u_{k|j} \log \left( \frac{u_{k|j}}{\sum_{i=1}^{N} e_{i}u_{k|j}} \right) \right)}.$$  

(14)

where $u_{k|j} = p(w_k|x_i)$ is the membership function. $e_i$ will be iteratively updated by (14).

It has been proved that when $e_i$ is near to zero, the input data $x_i$ can be regard as an outlier point [18]. In this paper, we extend this method to the mixture model by replacing the membership $u_{k|j}$ with posterior probability $z_{ik}$. Both theory and experiments have proved that the MI maximization is a powerful tool to identify outliers as the highly unreliable data point (pixel).

III. PROPOSED MODEL

A. MODIFYING SPATIAL INFORMATION

Motivated by the constructions of factor $G_{ik}$ in FRSCGMM and the information-theoretic method in ASIFC, we proposed a robust spatial information-theoretic GMM algorithm for brain MR images segmentation. In order to circumvent these shortcomings in above Gaussian mixture models based on MRF, we introduce a modified spatial factor $F_{ik}$ considering the non-local information because the image patches contain more structure information. The novel spatial factor takes the non-local spatial information into account based on the posterior probabilities and prior probabilities, which can preserve more details for the final segmentation and overcome the over-smoothing disadvantage of factor $G_{ik}$. The modified spatial factor $F_{ik}$ based on the posterior probabilities and prior probabilities is defined as

$$F_{ik} = \exp(\beta \sum_{m \in N_j} \frac{w_{im}}{w_{im}} (z_{mk} + \pi_{nm})).$$

(15)

where $\beta$ is the temperature value that controls the smoothing prior, $N_j$ is the neighborhood of pixel $x_i$, $z_{nm}$ is the posterior probability and $\pi_{nm}$ is the prior probabilities for the pixel $x_m$ in label $k$. $w_{im}$ is the non-local weight, which is defined as

$$w_{im} = \exp\left( -\frac{||P_m - P_i||^2}{h} \right).$$

(16)

$h$ is the smoothing parameter, which controls the smoothness of noise. The smaller $h$ is, the more image details will be retained. In the end of experiment part, we analyze the effect of the smoothing parameter $h$. Defining the neighbor of pixel $x_i$ is a neighbor patch $P_i$, then the similarity of pixel $x_i$ and $x_m$ is the gray-similarity of image patch $P_i$ and $P_m$. The distance between patch $P_i$ and $P_m$ is calculated as following

$$||P_m - P_i||^2 = \sum_{n=1}^{||P_i||} (P_m^{(n)} - P_i^{(n)})^2.$$  

(17)

$P_i^{(n)}$ is the nth pixel value of the image patch $P_i$.

The non-local weight $w_{im}$ are used to characterize the similarity between image patch $P_j$ and $P_m$. If the similarity of $P_i$ and $P_m$ is high, the corresponding weight $w_{im}$ will be great and the neighboring pixel $x_m$ will have more influence on the current pixel $x_i$. Compare with (10), we find (15) can adjust the influence of the neighboring pixels automatically by the weight parameter $w_{im}$. If the neighbor of pixel $x_i$ and $x_m$ is similar, they may belong to a same tissue and the weight $w_{im}$ will increase. Otherwise, if two pixels are far different, the pixel $x_m$ will have small effect on current pixel $x_i$. Therefore, the introduced non-local spatial information can make the segmentation result retain more image details.

Different from the spatial factor $G_{ik}$ in FRSCGMM [15], the proposed factor $F_{ik}$ not only takes account the local neighborhood information, but also considers the spatial structure information of pixels. The proposed algorithm introduces the non-local information and integrates the non-local weight into model directly. Therefore, our algorithm can retain more image details while reducing the influence of noise.

Now, a novel factor $F_{ik}$ is proposed to incorporate the spatial information into the smoothing prior. The new smoothing prior $U(\Pi)$ is given by

$$U(\Pi) = -\sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log(\pi_{ik}^{(t+1)}),$$

(18)

where $t$ indicates the iteration step. The MRF distribution $p(\Pi)$ in (5) is given by

$$p(\Pi) = Z^{-1} \exp\left( \frac{1}{T} \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log(\pi_{ik}^{(t+1)}) \right).$$

(19)
Given the MRF distribution $p(\Pi)$, we can obtain the log-likelihood function:

$$L(\Pi, \theta | X) = \sum_{i=1}^{N} \log \left( \sum_{k=1}^{K} \pi_{ik}^{(t+1)} \Phi(x_i | \theta_k^{(t+1)}) \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log \left( \pi_{ik}^{(t+1)} \right).$$

(20)

Application of the complete data condition in [19], maximizing the log-likelihood function $L(\Pi, \theta | X)$ in (20) will be equivalent to maximizing the objective function $J(\Pi, \theta | X)$:

$$J(\Pi, \theta | X) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik}^{(t)} \left( \log \pi_{ik}^{(t+1)} + \log \Phi(x_i | \theta_k^{(t+1)}) \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log \left( \pi_{ik}^{(t+1)} \right).$$

(21)

The posterior probability $z_{ik}$ can be computed as follows:

$$z_{ik}^{(t)} = \frac{\pi_{ik}^{(t)} \Phi(x_i | \theta_k^{(t)})}{\sum_{j=1}^{K} \pi_{ij}^{(t)} \Phi(x_i | \theta_j^{(t)})}.$$  

(22)

In order to estimate the bias field $B$ in brain MR images, we assume that the bias field is modeled as a linear combination of a set of smooth basis functions

$$b_i = \sum_{l=1}^{S} w_l g_l(i) = w^T G(i),$$

(23)

where $g_l, l = (1, 2, \ldots, S)$ is a set of basis functions, $S$ is the number of $g_l$, and $w_l \in R, l = (1, 2, \ldots, S)$ are the real-valued combination coefficients. In this study, we use orthogonal polynomials to represent $g_l$ and $S = (P + 1)(P + 2)/2$, where $P$ is the degree of these polynomials. Theoretically, any function can be approximated by a linear combination of a set of basis functions up to arbitrary accuracy [20].

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

$$L(\Pi, \theta, B | X) = \sum_{i=1}^{N} \log \left( \sum_{k=1}^{K} \pi_{ik}^{(t+1)} \Phi(x_i | \theta_k^{(t+1)}) b_i^{(t)} \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log \left( \pi_{ik}^{(t+1)} \right),$$

(24)

where $\Phi(x_i | \theta_k, b_i)$ is the Gaussian distribution with parameter $\theta_k$ and the bias field $b_i$, which can be defined as

$$\Phi(x_i | \theta_k, b_i) = \frac{1}{(2\pi)^{\frac{P}{2}}} \frac{1}{\mu_k} \exp \left( -\frac{1}{2} (x_i - b_i \mu_k)^T \sum_k^{-1} (x_i - b_i \mu_k) \right).$$

(25)

Therefore, the objective function of the proposed model is given by

$$J(\Pi, \theta, B | X) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik}^{(t)} \left( \log \pi_{ik}^{(t+1)} \log \Phi(x_i | \theta_k^{(t+1)}) b_i^{(t)} \right) - \log Z + \frac{1}{T} \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log \left( \pi_{ik}^{(t+1)} \right).$$

(26)

where

$$z_{ik}^{(t)} = \frac{\pi_{ik}^{(t)} \Phi(x_i | \theta_k^{(t)}) b_i^{(t)}}{\sum_{j=1}^{K} \pi_{ij}^{(t)} \Phi(x_i | \theta_j^{(t)}) b_j^{(t)}}.$$  

(27)

The next step is to maximize the objective function $J(\Pi, \theta, B | X)$. Similar to the MRF-based methods [12], [14], $Z$ and $T$ in (26) are set to one. Then, we have the new objective function

$$J(\Pi, \theta, B | X) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik}^{(t)} \left( \log \pi_{ik}^{(t+1)} + \log \Phi(x_i | \theta_k^{(t+1)}) b_i^{(t)} \right) + \sum_{i=1}^{N} \sum_{k=1}^{K} F_{ik}^{(t)} \log \left( \pi_{ik}^{(t+1)} \right).$$

(28)

To obtain the estimations of $\mu_k^{(t+1)}$ and $\sum_k^{(t+1)}$, we need to maximize the objective function by solving $\frac{\partial J}{\mu_k^{(t+1)}} = 0$ and $\frac{\partial J}{\sum_k^{(t+1)}} = 0$. These lead to the iterative solution

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ik}^{(t)} \Phi(x_i | \theta_k^{(t+1)})}{\sum_{j=1}^{K} z_{ij}^{(t)} F_{ij}^{(t)}}.$$  

(29)

$$\sum_k^{(t+1)} = \frac{\sum_{i=1}^{N} z_{ik}^{(t)} (x_i - b_i^{(t+1)} \mu_k^{(t+1)}) (x_i - b_i^{(t+1)} \mu_k^{(t+1)})^T}{\sum_{i=1}^{N} z_{ik}^{(t)}}.$$  

(30)

Considering the prior distribution follows the constraint $\sum_k^{(t+1)} = 1$, we use Lagrange multiplier method to get the estimation of $\pi_{ik}^{(t+1)}$. We have

$$\pi_{ik}^{(t+1)} = \frac{z_{ik}^{(t)} + F_{ik}^{(t)}}{\sum_{j=1}^{K} (z_{ij}^{(t)} + F_{ij}^{(t)})}.$$  

(31)

Finally, we should obtain the combination coefficients $w = \{w_1, w_2, \ldots, w_S\}$ in order to estimate the bias field $B$. By solving $\frac{\partial J}{w(i^{(t+1)})}$ = 0, we can get

$$w(i^{(t+1)}) = A^{-1} Q.$$  

(32)

where

$$A = \sum_{i=1}^{N} G(i) G(i)^T \sum_{k=1}^{K} z_{ik} \mu_k^{-1} \sum_k$$

$$Q = \sum_{i=1}^{N} G(i) \sum_{k=1}^{K} z_{ik} \mu_k^{-1} x_i.$$  

(33)

Iterating through these necessary conditions (29)-(33) will lead to the maximum of objective function $J(\Pi, \theta, B | X)$.

### B. OUTLIER ELIMINATION BY MI MAXIMIZATION

To identify and eliminate outliers, we extend the MI maximization method to the mixture model by replacing membership value $u_{ki}$ in Eq. (14) with posterior probability $z_{ik}$ shown in Eq. (27). After generating the estimated posterior probability $z_{ik}$, we can evaluate the input data reliability by the information-theoretic method. Highly unreliable data points will be viewed as outliers and need to be eliminated. We can
obtain a set of robust input data reliability values \( p(x_i) \) for every input data points by using the mutual information (MI) maximization method. The reliability values \( p(x_i) \) of highly unreliable data points (outliers) are zero. So, we can use it to label/eliminate the outliers. Similar with [18], in order to get the robust estimation of input data distribution \( p(x_i) \), we define the mutual information between the optimal parameter distribution \( p(\theta) \) and the input data distribution \( p(X) \) as follows:

\[
I(X, \theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} p(x_i, \theta_k) \log \frac{p(x_i, \theta_k)}{p(x_i) p(\theta_k)}
\]

\[
= \sum_{i=1}^{N} \sum_{k=1}^{K} p(x_i) p(\theta_k | x_i) \log \frac{p(x_i) p(\theta_k | x_i)}{p(x_i) p(\theta_k)}
\]

\[
= \sum_{i=1}^{N} \sum_{k=1}^{K} p(x_i) z_{ik} \log \frac{z_{ik}}{\sum_{i=1}^{N} p(x_i) z_{ik}}. \quad (34)
\]

The next objective is to maximize the mutual information \( I(X, \theta) \). Considering \( \sum_{i=1}^{N} p(x_i) = 1 \) limitation, Lagrange’s method is used. Finally, we can get the robust estimate for \( p(x_i) \) by the following function

\[
p(x_i) \propto \frac{p(x_i) \exp \left( \sum_{k=1}^{K} z_{ik} \log \frac{z_{ik}}{\sum_{i=1}^{N} p(x_i) z_{ik}} \right)}{\sum_{i=1}^{N} \exp \left( \sum_{k=1}^{K} z_{ik} \log \frac{z_{ik}}{\sum_{i=1}^{N} p(x_i) z_{ik}} \right)}.
\]

\[
(35)
\]

In this paper, we introduce a novel spatial factor based on the non-local spatial information to reduce the effect of noise and preserve more image details. The spatial factor contains not only posterior information but also prior information, which can preserve more details for the final segmentation and overcome the over-smoothing disadvantage of factor \( G_{ik} \). It plays a role as linear filters for smoothing and restoring images corrupted by noise. The spatial factor is also easy to implement and incorporates the spatial relationships among neighboring pixels in a simpler metric. In addition, in order to reduce the influence of bias field on segmentation, we integrate the bias field estimation into the objective function, which makes the method can estimate the bias field meanwhile segmenting the image. Finally, we use the mutual information (MI) maximization method to identify outliers in order to further improve the accuracy of segmentation. The corresponding algorithm is a two-step segmentation algorithm. In the first step, it uses the improved GMM algorithm based on MRF to segment images and estimate the bias field at the same time. Afterwards, the information-theoretic method is used to eliminate outliers and further improve segmentation results. The flow step of the proposed algorithm can be summarized as follows:

Step 1) Initialize the parameters \( \{\Pi, \theta, B\} \).

Step 2) E step:

a) Evaluate the posterior probability \( z_{ik} \) in (27) by using the current parameter values.

b) Update the factor \( F_{ik} \) by using (15).

Step 3) M step:

a) Update \( \mu_k \) by using (29).

b) Update \( \sum_k \) by using (30).

c) Update \( \pi_k \) by using (31).

d) Update \( w \) and \( b_i \) by using (32) and (33);

Step 4) If the convergence criterion \( \left\| J^{(t+1)} - J^{(t)} \right\| \leq \varepsilon \) is satisfied, stop the iteration; otherwise, return to Step 2.

MI maximization for identifying the unreliable data points:

Step 5) Select \( p(x_i) = \frac{1}{N}, i = 1, \ldots, N, \) and \( \varepsilon > 0 \), then start the fixed-point iteration according to Steps 6-7.

Step 6) Update the input data distribution \( p(x_i) \) by using

\[
c_i = \exp \left( \sum_{k=1}^{K} z_{ik} \log \frac{z_{ik}}{\sum_{i=1}^{N} p(x_i) z_{ik}} \right)
\]

and

\[
p(x_i) = \frac{p(x_i) c_i}{\sum_{i=1}^{N} c_i}.
\]

Step 7) If

\[
\ln \sum_{i=1}^{N} p(x_i) c_i - \ln \max c_i < \varepsilon (\varepsilon > 0)
\]

then go to Step 8; otherwise, return to Step 6.

Step 8) Label unreliable data points. Create a \( 3 \times 3 \) window centered at the outlier, get the clustering labels for all neighboring pixels within this window.

Step 9) Store all label counts for each cluster into \( \xi_k \) (\( k = 1, 2, \ldots, K \)). Find the maximum value of \( \xi_k \) and its corresponding cluster index \( m = \max (\xi_k) \).

Step 10) If \( \xi_1 = \xi_2 = \ldots = \xi_k \), increase the current window edge by two more pixels, go to Step 8. Otherwise, update the outlier label with \( m \).

Step 11) Iterate for all outliers until there is no further change, then stop.

In the next section, experimental results will be given to demonstrate the robustness and accuracy of the proposed algorithm, as compared with other GMM-based models.

IV. EXPERIMENTAL RESULTS

In this section, we conduct the proposed algorithm on a set of synthetic and clinical 3T brain MR images to illustrate the performance of our model. These images are segmented to grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). In the comparative study, some models are chosen to compare with our proposed algorithm. These models are standard GMM [8], DCA-SVFMM [14], FRSCGMM [15], BMMSI [24], ASIFC [18], and PFCM [34].

Without special instructions, each algorithm in our experiments uses the following parameters. In DCA-SVFMM, the total number of considered directions is set to be 4 and the size of neighborhood window is \( 5 \times 5 \). For FRSCGMM, temperature value is obtained by searching the interval \([0, 10]\) in terms of the optimal JS value and the size of neighborhood window is \( 5 \times 5 \). The initial value of the smoothing parameter \( \beta \) in BMMSI is set to 65 and the learning rate \( l \) is set
The size of neighborhood window in BMMSI is also $3 \times 3$. The maximum iteration number for each algorithm is 100. For PFCM, the values of $a$, $b$, $m$, $\eta$ are set to 1, 5, 7, and 1.5, respectively. Finally, the temperature value $\beta$ and the smoothing parameter $h$ in the proposed algorithm are obtained by searching the interval $[0,10]$ and $[0,200]$ respectively in terms of the optimal JS value. All methods are initialized by using K-means method.

To quantify the segmentation results and verify the performance of the proposed model, Jaccard similarity (JS) [1] is used as a criterion. It is defined as the ratio between intersection and union of the segmentation result $S_1$ and the ground truth $S_2$:

$$J(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}. \quad (36)$$

The JS value is between 0 and 1 and a more accurate segmentation result should correspond to a higher JS value. In order to show the robustness to noise, we test our algorithm on a synthetic image. Fig. 1 (a) is the original image and (b) is the original image corrupted by Gaussian noise (parameters: mean 0 and variance 0.007). Fig. 1 (c)-(g) are the segmentation results of the standard GMM, BMMSI, the proposed algorithm, ASIFC and PFCM, respectively. Table 1 shows the JS values comparison of segmentation results on the synthetic image with different noise level. We can see from the results that our algorithm is much more robust to the presence of noise as compared to other algorithms because of the incorporation of the non-local spatial information and the outliers elimination by MI maximization.

### Table 1. Comparison of JS values for segmentation results on simple synthetic image.

| Noise | JS     | Proposed Algorithm | GMM | BMMSI | ASIFC | PFCM |
|-------|--------|--------------------|-----|-------|-------|------|
| 3%    | 0.9785 | 0.4471             | 0.9664 | 0.9765 | 0.8536 |
| 5%    | 0.8793 | 0.6596             | 0.8511 | 0.9537 | 0.6864 |
| 7%    | 0.9780 | 0.4642             | 0.7283 | 0.8592 | 0.5792 |

### A. SEGMENTATION OF SYNTHETIC BRAIN MR IMAGES

In this section, the synthetic brain MR images will be used for segmentation to show the performance of our algorithm. The first experiment in this section are carried out in 3T-weighted brain MR images. Fig. 2 shows the segmentation result of our algorithm on 3T-weighted brain MR images with 5% noise and 80% intensity inhomogeneity. The first and second row show the initial images with intensity inhomogeneity, estimated bias field, bias corrected images, and the segmentation results, respectively. It can be seen from the results that the intensities within each brain tissue in the bias-corrected images become quite homogeneous. Fig. 2 demonstrate that the results of our method are consistent with the expected tissue regions.

![FIGURE 2. Illustration of three 3T-weighted brain MR images (1st column), the estimated bias field (2nd column), bias-corrected images (3rd column), and the segmentation results of the proposed algorithm (4th column).](image)

We can evaluate the performance of bias field correction by quantifying the intensity inhomogeneities of the bias field corrected images using the coefficient of joint variation (CJV). For two tissues $T_1$ and $T_2$ (GM, WM or CSF), the CJV is defined as

$$\text{CJV} = \frac{\sigma(T_1) + \sigma(T_2)}{\mu(T_1) - \mu(T_2)} \quad (37)$$

where $\sigma(T)$ and $\mu(T)$ are the standard deviation and the mean of the intensities in the tissue $T$. The smaller CJV values indicate better bias field correction results. The CJV values of the three tested methods for the 21 images (N5F80) are plotted in Fig. 3, which shows better performance of our method than two other methods.
FIGURE 3. Comparison of the performance of bias field correction of our method, MICO, and RSCFCM in terms of CJV for 21 brain MR images.

FIGURE 4. Segmentation results on simulated brain MR image with parameter: noise level 5% and intensity inhomogeneity level 0%. (a) The original image. (b) Ground truth image. (c) Standard GMM (JS = 0.6890). (d) DCA-SVFMM (JS = 0.7436). (e) FRSCGMM (JS = 0.8019). (f) BMMSI (JS = 0.7667). (g) Step 1 of Proposed Algorithm (JS = 0.8353). (h) Proposed Algorithm (JS = 0.8425).

FIGURE 5. Segmentation results on simulated brain MR image with parameter: noise level 7% and intensity inhomogeneity level 0%. (a) The original image. (b) Ground truth image. (c) Standard GMM (JS = 0.4652). (d) DCA-SVFMM (JS = 0.6502). (e) FRSCGMM (JS = 0.7614). (f) BMMSI (JS = 0.5746). (g) Step 1 of Proposed Algorithm (JS = 0.7806). (h) Proposed Algorithm (JS = 0.7939).

In the second experiment, we compared our method with above four existing segmentation algorithms on a simulated brain MR images with different noise level to show the robustness of our method to noise. All images in this experiment are selected from BrainWeb (http://brainweb.bic.mni.mcgill.ca/). The BrainWeb provides full three-D synthetic brain MRI data with three modalities (T1, T2 and PD) and it can also provide brain data sets with various slice thickness, noise levels and levels of intensity inhomogeneity. In this experiment, we use the T1-weighted 1mm synthetic brain MR images with no intensity inhomogeneity and a noise level ranging from 3% to 7%.

Fig. 4 and Fig. 5 respectively show the segmentation results on the simulated brain MR images with noise level 5% and 7%. Fig. 4 (a) and Fig. 5 (a) show the original images corrupted by noise and outliers. Fig. 4 (b) and Fig. 5 (b) are the ground truth. We can see that the standard GMM (Fig. 4 (c) and Fig. 5 (c)) performs poor for the image with heavy noise. Due to the lack of spatial constrains, the GMM is very sensitive to noise and outliers. The DCA-FCM (Fig. 4 (d) and Fig. 5 (d)) is a directionally adaptive algorithm and has more complexity on the log-likelihood function. The incorporated local spatial information makes the M-step of EM algorithm cannot be directly applied to the prior distribution. In order to update the contextual mixing proportions, the DCA-FCM needs to solve a quadratic equation for each pixel belong to each cluster. It leads the segmentation results unstable. It can be seen from Fig. 4 (d) and Fig. 5 (d) that the DCA-FCM cannot achieve good segmentation results in high noise environment. The segmentation result of the FRSCGMM (Fig. 4 (e) and Fig. 5 (e)) is much smoother. The FRSCGMM constructs the spatial information based
on the prior and posterior probability, but it only utilizes a smoothing filter over the whole image. Therefore, it may get over-smoothed segmentation results. To improve the segmentation accuracy, the BMMSI (Fig. 4 (f) and Fig. 5 (f)) choose the Student’s t- distribution for its component function and the representation of contextual mixing proportion is closely related to the Gaussian distribution of its neighborhood. However, only image intensity is adopted as feature to apply to segmentation for the BMMSI. It do not consider the region or boundary information of image, hence it cannot achieve the satisfied results. Comparing with these four algorithms, our proposed method can visually produce better segmentation results by using the non-local spatial information and MI maximization method.

For quantitative comparison, we apply above five algorithms to the segmentation of 30 MR images with a noise level ranging from 3% to 7%. The segmentation accuracy is measured by using means and standard deviations of JS values on GM, WM, and CSF. The results are listed in Table 2. This comparison demonstrates that the proposed algorithm produces the most accurate segmentation (with higher means of JS values) and has the best ability and robustness of denoising (with lower standard deviations of JS values). In addition, we can also find that the proposed algorithm has the highest mean JS values for CSF tissue, which shows that our method can retain more image details.

In the third experiment, we evaluate the proposed algorithm and compare it to five existing brain MRI segmentation algorithms, including AS-EM [9], AM-EM [10], BFCFM [2], RSCFCM [22], and MICO [23], to show the robustness of our method to intensity inhomogeneity. The parameters for each method are set with the default values specified in the papers. Please refer to the corresponding references for more details.

In order to show the impact of the intensity inhomogeneity, the segmentation results obtained by applying above six methods to the simulated brain MR images from BrainWeb are displayed in Fig. 6. Fig. 6 (a) is the initial image with 80% intensity inhomogeneity and 3% noise and Fig. 6 (b) is the ground truth. The intensity inhomogeneity level in the original image is very high, which makes the segmentation algorithms hard to obtain accurate results. AS-EM (Fig. 6 (f)) is an iterative algorithm which interleaves classification with bias field estimation. However, AS-EM is constructed by manually selecting representative points of each cluster, which may cause the segmentation results that are not fully objective and reproducible. The AM-EM (Fig. 6 (e)) is an improvement of AS-EM and it uses a digital brain atlas to provide the prior probability maps of brain tissues. However, the parameters initialization of the AM-EM remains very significant, which may result in unsatisfied segmentation results under K-means initializations. Fig. 6 (g) shows the segmentation result of the RSCFCM. The RSCFCM only uses pixels information on one of the four directions in the neighborhood to decrease the influence of noise, hence it cannot retain more image details. MICO (Fig. 6 (d)) can estimate the bias field well and effectively reduce the influence of severe intensity inhomogeneity, but it is sensitive to noise and outliers. The BFCFM (Fig. 6 (c)) can adaptively

| Algorithm | Tissue | 3% | 5% | 7% |
|-----------|--------|----|----|----|
| JS        | GM     | 0.8922 ± 0.0052 | 0.8564 ± 0.0064 | 0.8161 ± 0.0059 |
| JS        | WM     | 0.9109 ± 0.0065 | 0.8755 ± 0.0089 | 0.8369 ± 0.0126 |
| JS        | CSF    | 0.8807 ± 0.0226 | 0.8300 ± 0.0240 | 0.7822 ± 0.0277 |
| GMM       | JS     | 0.8479 ± 0.0095 | 0.6589 ± 0.0139 | 0.4627 ± 0.0315 |
| GMM       | GM     | 0.8472 ± 0.0146 | 0.6900 ± 0.0233 | 0.5455 ± 0.0344 |
| GMM       | WM     | 0.8356 ± 0.0118 | 0.5594 ± 0.0199 | 0.1940 ± 0.0371 |
| GMM       | CSF    | 0.8813 ± 0.0203 | 0.7858 ± 0.0301 | 0.6909 ± 0.0394 |
| DCA_SV    | JS     | 0.8686 ± 0.0045 | 0.7559 ± 0.0084 | 0.6439 ± 0.0090 |
| DCA_SV    | GM     | 0.8613 ± 0.0060 | 0.7387 ± 0.0090 | 0.6261 ± 0.0141 |
| DCA_SV    | WM     | 0.8750 ± 0.0070 | 0.7589 ± 0.0163 | 0.6501 ± 0.0161 |
| DCA_SV    | CSF    | 0.8729 ± 0.0067 | 0.8025 ± 0.0318 | 0.6828 ± 0.0434 |
| BMMSI     | JS     | 0.8753 ± 0.0033 | 0.7577 ± 0.0054 | 0.5716 ± 0.0120 |
| BMMSI     | GM     | 0.8866 ± 0.0066 | 0.7733 ± 0.0139 | 0.6042 ± 0.0221 |
| BMMSI     | WM     | 0.9088 ± 0.0044 | 0.7655 ± 0.0075 | 0.5229 ± 0.0078 |
| BMMSI     | CSF    | 0.8251 ± 0.0325 | 0.7551 ± 0.0472 | 0.6869 ± 0.0427 |
| RSCGMV    | JS     | 0.8699 ± 0.0132 | 0.8125 ± 0.0177 | 0.7652 ± 0.0144 |
| RSCGMV    | GM     | 0.8555 ± 0.0141 | 0.8069 ± 0.0156 | 0.7631 ± 0.0138 |
| RSCGMV    | WM     | 0.8785 ± 0.0124 | 0.8376 ± 0.0185 | 0.7934 ± 0.0208 |
| RSCGMV    | CSF    | 0.8282 ± 0.0300 | 0.7551 ± 0.0472 | 0.6869 ± 0.0427 |
FIGURE 6. Segmentation results on simulated brain MR image with parameter: noise level 3% and intensity inhomogeneity level 80%. (a) The original image. (b) Ground truth image. (c-h) are the results of BCFCM, MICO, AM-EM, AS-EM, RSCFCM and our method, respectively.

Y. Chen et al.: Robust Spatial Information-Theoretic GMM Algorithm

TABLE 3. JS values (mean ± standard deviation) for the segmentation results on simulated brain MR images.

|    | JBCFCM | MICO | AM-EM | AS-EM | RSCFCM | proposed |
|----|--------|------|-------|-------|--------|----------|
| N3 | F20    |      |       |       |        |          |
| JS | 0.8456 ± 0.0260 | 0.8623 ± 0.0390 | 0.8245 ± 0.0148 | 0.6032 ± 0.0129 | 0.8327 ± 0.0065 | 0.8942 ± 0.0048 |
| JS | 0.8373 ± 0.0020 | 0.8559 ± 0.0148 | 0.8371 ± 0.0144 | 0.8371 ± 0.0114 | 0.8942 ± 0.0048 |
| JS | 0.8583 ± 0.0013 | 0.8868 ± 0.0301 | 0.8824 ± 0.0283 | 0.5979 ± 0.0093 | 0.8311 ± 0.0029 | 0.9671 ± 0.0045 |
| JS | 0.8329 ± 0.0035 | 0.8772 ± 0.0281 | 0.8716 ± 0.0257 | 0.8112 ± 0.0125 | 0.8674 ± 0.0064 | 0.8831 ± 0.0048 |
| N3 | F40    |      |       |       |        |          |
| JS | 0.8458 ± 0.0019 | 0.8638 ± 0.0081 | 0.8297 ± 0.0043 | 0.6036 ± 0.0070 | 0.8461 ± 0.0045 | 0.8771 ± 0.0065 |
| JS | 0.8211 ± 0.0595 | 0.8534 ± 0.0275 | 0.8175 ± 0.0415 | 0.5518 ± 0.0275 | 0.8328 ± 0.0149 | 0.8707 ± 0.0016 |
| JS | 0.8370 ± 0.0263 | 0.8609 ± 0.0229 | 0.8592 ± 0.0271 | 0.8218 ± 0.0105 | 0.8381 ± 0.0049 |
| JS | 0.8528 ± 0.0032 | 0.8786 ± 0.0242 | 0.8732 ± 0.0352 | 0.8197 ± 0.0313 | 0.8212 ± 0.0220 |
| JS | 0.8377 ± 0.0024 | 0.8404 ± 0.0173 | 0.8555 ± 0.0247 | 0.8454 ± 0.0137 | 0.8131 ± 0.0074 |
| N3 | F60    |      |       |       |        |          |
| JS | 0.8411 ± 0.0081 | 0.8701 ± 0.0132 | 0.8564 ± 0.0249 | 0.5996 ± 0.0249 | 0.8325 ± 0.0074 | 0.8714 ± 0.0095 |
| JS | 0.8327 ± 0.0029 | 0.8705 ± 0.0254 | 0.8719 ± 0.0519 | 0.8064 ± 0.0027 | 0.8591 ± 0.0034 |
| N3 | F80    |      |       |       |        |          |
| JS | 0.8621 ± 0.0023 | 0.8648 ± 0.0106 | 0.8533 ± 0.0147 | 0.6053 ± 0.0055 | 0.8169 ± 0.0010 | 0.8724 ± 0.0098 |
| JS | 0.8907 ± 0.0049 | 0.8546 ± 0.0109 | 0.8168 ± 0.0042 | 0.5879 ± 0.0145 | 0.8092 ± 0.0026 | 0.8651 ± 0.0028 |
| JS | 0.8941 ± 0.0074 | 0.8716 ± 0.0132 | 0.8601 ± 0.0284 | 0.8493 ± 0.0229 | 0.8341 ± 0.0019 |
| JS | 0.8051 ± 0.0022 | 0.8766 ± 0.0287 | 0.8777 ± 0.0297 | 0.8189 ± 0.0353 | 0.7919 ± 0.0024 | 0.8460 ± 0.0078 |
| N5 | F80    |      |       |       |        |          |
| JS | 0.7859 ± 0.0014 | 0.7481 ± 0.0192 | 0.7577 ± 0.0275 | 0.5353 ± 0.0076 | 0.7491 ± 0.0192 | 0.8032 ± 0.0045 |
| JS | 0.7702 ± 0.0026 | 0.7284 ± 0.0152 | 0.6689 ± 0.0045 | 0.4892 ± 0.0019 | 0.7399 ± 0.0045 | 0.8142 ± 0.0019 |
| JS | 0.8065 ± 0.0029 | 0.7577 ± 0.0032 | 0.6957 ± 0.0017 | 0.5689 ± 0.0058 | 0.7810 ± 0.0023 | 0.8248 ± 0.0029 |
| JS | 0.8471 ± 0.0035 | 0.7966 ± 0.0040 | 0.7676 ± 0.0045 | 0.7676 ± 0.0045 | 0.7966 ± 0.0040 | 0.8340 ± 0.0040 |

segment brain MR images and correct bias field. But it also cannot preserve image details well. Comparing with segmentation results of other five algorithms and the ground truth, our proposed method can effectively overcome the effect of intensity inhomogeneity and noise and retain more image details, which makes the segmentation results more accurate. Similarly, we also apply these algorithms on 21 brain MR images selected from BrainWeb (60th∼80th), which contain 3% noise and the intensity inhomogeneity ranging from 20% to 80%. The mean and standard deviation of JS values for the segmentation of GM, WM, and CSF are listed in Table 3. Both visual and quantitative comparisons demonstrate that our proposed algorithm has the superior performance in dealing with the intensity inhomogeneity.

B. SEGMENTATION OF CLINICAL BRAIN MR IMAGES

In this section, we will segment the clinical brain MR images which are generated from Internet Brain Segmentation Repository (IBSR, http://www.cma.mgh.harvard.edu/ibsr/). Above six algorithms (AS-EM, AM-EM, BCFCM, RSCFCM, and the proposed method) will be applied to the segmentation of these clinical brain MR images in this experiment. Fig. 7 and Fig. 8 respectively show the segmentation results on the clinical brain MR image data sets from IBSR (205_3# and 15_3#). Each data set has intensity inhomogeneity.

Fig. 7 (a) is the initial image (21th of 205_3#) with the low-level intensity inhomogeneity and Fig. 7 (b) is the ground truth. Fig. 7 (c-h) are the segmentation results of above six algorithms. Comparing with the ground truth, our proposed method can effectively overcome the effect of intensity inhomogeneity and noise. Similarly, we also apply these algorithms on 21 brain MR images selected from BrainWeb (60th∼80th), which contain 3% noise and the intensity inhomogeneity ranging from 20% to 80%. The mean and standard deviation of JS values for the segmentation of GM, WM, and CSF are listed in Table 3. Both visual and quantitative comparisons demonstrate that our proposed algorithm has the superior performance in dealing with the intensity inhomogeneity.

The segmentation accuracy (mean ± standard deviation of JS values) of these algorithms on the data sets (205_3# and 15_3#) from IBSR is listed in Table 4. We only list the JS values of GM and WM because the data sets only contain the few pixels belonging to CSF. Table 4 demonstrates that our proposed algorithm can obtain more accurate results.

The proposed algorithm has two main constant parameters: the temperature value \( \beta \) and the smoothing parameter \( h \). We will test the effect of them to show the robustness of our method. The parameters are chosen small enough to prevent the image from losing much of its sharpness and details. But the much smaller \( \beta \) and \( h \) will make the influence of neighbor pixels much smaller than the center pixel, which will decrease...
Y. Chen et al.: Robust Spatial Information-Theoretic GMM Algorithm

FIGURE 7. Segmentation results on a clinical brain MR image (21th of 205_3#) (a) The original image. (b) Ground truth image. (c-h) are the results of BCFCM, MICO, AM-EM, AS-EM, RSCFCM and our method, respectively.

FIGURE 8. Segmentation results on a clinical brain MR image (34th of 15_3#) (a) The original image. (b) Ground truth image. (c-h) are the results of BCFCM, MICO, AM-EM, AS-EM, RSCFCM and our method, respectively.

the robustness to noise and outliers. However, if we choose the much bigger $\beta$ and $h$, it will cause over-smoothing and losing image details. The selection of $\beta$ and $h$ is usually dictated by the considered application of the model.

The effect of the parameter $\beta$ and $h$ in our algorithm is tested on the synthetic MR brain images (3%, 5%, 7% noise and 80% intensity inhomogeneity) from BrainWeb. For the constant parameter, we experiment on the synthetic brain MR images of slice 60th~80th. The average JS values of GM, WM, and CSF are shown in Fig. 9 and Fig. 10. It can be seen from the Fig. 9, when we set $\beta$ as 2, our algorithm can get a better result for segmentation in different noise level. In our paper, the temperature value $\beta$ is manually set to 2. And Fig. 10 shows that we should choose a smaller $h$ in low

| TABLE 4. JS values (mean ± standard deviation) for the segmentation results on clinical T1-weighted brain MR images. (%) |
|---|
|   | BCFCM | MICO | AM-EM | AS-EM | RSCFCM | proposed |
| GM | 0.3687 ± 0.0455 | 0.3926 ± 0.0389 | 0.3414 ± 0.0374 | 0.3390 ± 0.0293 | 0.3995 ± 0.0180 | 0.4816 ± 0.0231 |
| WM | 0.4543 ± 0.0377 | 0.4513 ± 0.0339 | 0.4256 ± 0.0333 | 0.4291 ± 0.0283 | 0.4641 ± 0.0372 | 0.4676 ± 0.0318 |
| GM | 0.5631 ± 0.0211 | 0.5809 ± 0.0448 | 0.5508 ± 0.0337 | 0.5809 ± 0.0453 | 0.5083 ± 0.1151 | 0.6019 ± 0.1266 |
| WM | 0.6021 ± 0.0392 | 0.6912 ± 0.0609 | 0.6544 ± 0.0607 | 0.6739 ± 0.0458 | 0.6702 ± 0.0497 | 0.7061 ± 0.0530 |

FIGURE 9. Illustration of the effect of $\beta$ on a synthetic brain MR image data with parameters: noise level 3%, 5%, 7% and intensity inhomogeneity level 80%.

FIGURE 10. Illustration of the effect of $h$ on a synthetic brain MR image data with parameters: noise level 3%, 5%, 7% and intensity inhomogeneity level 80%.
noise level images to retain details and choose a bigger $h$ in high noise level images to tolerate the noise. In our paper, the smoothing parameter $h$ is set to 50 in 3% noise level images and 75 in 5% and 7% noise level images.

V. CONCLUSIONS
This paper proposed a robust spatial information-theoretic GMM algorithm for the bias field estimation and brain MRI segmentation. This method efficiently incorporates non-local spatial information and integrates the bias field estimation model into a GMM framework to simultaneously estimate the intensity inhomogeneity and segment brain MR images. By incorporating the non-local spatial information, the proposed algorithm can reduce the effect of noise and preserve more details. Meanwhile, it also uses the MI maximization method to improve the robustness to outliers. Therefore, the proposed method overcomes the drawbacks of existing GMM-based clustering algorithms, including limited robustness to outliers, over-smoothness for segmentation and limited accuracy for image details. The experimental results on both synthetic and clinical brain MR images demonstrate that the proposed algorithm can simultaneously overcome the influence of noise and intensity inhomogeneity, and outperforms other several state-of-the-art algorithms.

ACKNOWLEDGMENT
The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES
[1] U. Vovk, F. Pernus, and B. Likar, “A review of methods for correction of intensity inhomogeneity in MRI,” IEEE Trans. Med. Imag., vol. 26, no. 3, pp. 405–421, Mar. 2007.
[2] M. N. Ahmed, S. M. Yamany, A. A. Farag, and T. Moriarty, “Bias field estimation and adaptive segmentation of MRI data using a modified fuzzy C-means algorithm,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 1999, pp. 250–255.
[3] M.-E. Algorri and F. Flores-Mangas, “Classification of anatomical structures in MR brain images using fuzzy parameters,” IEEE Trans. Biomed. Eng., vol. 51, no. 9, pp. 1599–1608, Sep. 2004.
[4] Z. Wu, K. D. Paulsen, and J. M. Sullivan, “Adaptive model initialization and deformation for automatic segmentation of T1-weighted brain MRI data,” IEEE Trans. Biomed. Eng., vol. 52, no. 6, pp. 1128–1131, Jun. 2005.
[5] H.-H. Chang, D. J. Valentina, G. R. Duckwiler, and A. W. Toga, “Segmentation of brain MR images using a charged fluid model,” IEEE Trans. Biomed. Eng., vol. 54, no. 10, pp. 1798–1813, Oct. 2007.
[6] S. K. Michahelles, L. Costaridou, E. Panagiotopoulos, R. Speller, G. Panayiotaakis, and A. Todd-Pokropek. “Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine,” IEEE Trans. Biomed. Eng., vol. 56, no. 9, pp. 2225–2231, Sep. 2009.
[7] C. Ambroise, “The EM algorithm and extensions, by G.M. McLachlan and T. Krishnan,” J. Classification, vol. 15, no. 1, pp. 154–156, 1998.
[8] G. McLachlan and D. Peel, Finite Mixture Models. New York, NY, USA: Springer, 2000.
[9] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, “Adaptive segmentation of MRI data,” in IEEE Trans. Med. Imag., vol. 15, no. 4, pp. 429–442, Aug. 1996.
[10] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetsens, “Automated model-based bias field correction of MR images of the brain,” IEEE Trans. Med. Imaging., vol. 18, no. 10, pp. 885–896, Oct. 1999.
[11] L. Gupta and T. Sorantukul, “A Gaussian mixture based image segmentation algorithm,” Pattern Recognit., vol. 31, no. 3, pp. 315–325, 1998.
[12] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, “A spatially constrained mixture model for image segmentation,” IEEE Trans. Neural Netw., vol. 16, no. 2, pp. 494–498, Mar. 2005.
[13] A. Dipprasios, N. Vlassis, and T. Gevers, “A spatially constrained generative model and an EM algorithm for image segmentation,” IEEE Trans. Neural Netw., vol. 18, no. 3, pp. 798–808, May 2007.
[14] C. Nikou, N. P. Galatsanos, and A. C. Likas, “A class-adaptive spatially variant mixture model for image segmentation,” IEEE Trans. Image Process., vol. 16, no. 4, pp. 1121–1130, Apr. 2007.
[15] T. M. Nguyen and Q. M. J. Wu, “Fast and robust spatially constrained Gaussian mixture model for image segmentation,” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 4, pp. 621–635, Apr. 2013.
[16] Q. Song, “A robust information clustering algorithm,” Neural Comput., vol. 17, no. 12, pp. 2672–2689, Dec. 2005.
[17] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic clustering for image segmentation,” Comput. Vis. Image Underst., vol. 117, no. 10, pp. 1412–1420, Oct. 2013.
[18] Z. Wang, Q. Song, Y. C. Soh, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 69, no. 1, pp. 1–17, Jun. 1998.
[19] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Underst., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[20] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[21] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[22] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[23] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[24] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[25] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
[26] Z. M. Wang, Y. C. Soh, Q. Song, and K. Sim, “Adaptive spatial information-theoretic fuzzy clustering algorithm for image segmentation,” Comput. Vis. Image Understand., vol. 15, no. 4, pp. 281–300, Aug. 2000.
Yunjie Chen received the B.S. and M.S. degrees from the School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing, China, in 2002 and 2005, respectively, and the Ph.D. degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology, in 2009. He is currently a Professor with the School of Mathematics and Statistics, Nanjing University of Information Science and Technology. His research interests include medical image processing, statistical analysis, and machine learning.

Mao Cai received the B.S. degree in information and computing science from the Nanjing University of Information Science and Technology, Nanjing, China, in 2019. His research interests are in pattern recognition, image segmentation, and image processing.

Xinze Zhou received the B.S. degree in mathematics and applied mathematics from the Nanjing University of Information Science and Technology, Nanjing, China, in 2019. His research interests are in machine learning and image segmentation.

Cheng Ning is currently pursuing the M.S. degree in mathematics from the Nanjing University of Information Science and Technology, Nanjing, China. Her current research interest is in image segmentation with statistical model.

Chunzheng Cao received the Ph.D. degree in applied mathematics from Southeast University, Nanjing, China, in 2012. He is currently a Professor of statistics with the School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing. His research interests include statistical learning, functional/longitudinal data analysis, and Bayesian inference.

Jianwei Yang received the M.S. degree in mathematics from Xi’an Jiaotong University, in 1996, and the Ph.D. degree in computer science from Hong Kong Baptist University, Hong Kong, in 2005. He is currently a Professor with the School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing, China. His current research interests include pattern recognition and computer vision.

***