

# Non-local-based spatially constrained hierarchical fuzzy C-means method for brain magnetic resonance imaging segmentation

Yunjie Chen<sup>1</sup>, Jian Li<sup>1</sup>, Hui Zhang<sup>2</sup>✉, Yuhui Zheng<sup>2</sup>, Byeungwoo Jeon<sup>3</sup>, Qingming Jonathan Wu<sup>4</sup>

<sup>1</sup>School of Math and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China

<sup>2</sup>School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China

<sup>3</sup>College of Information and Communication Engineering, Sungkyunkwan University, Sungkyunkwan, Korea

<sup>4</sup>Department of Electrical and Computer Engineering, University of Windsor, Windsor, Canada

✉ E-mail: nrzhanghui@gmail.com

ISSN 1751-9659

Received on 19th April 2016

Accepted on 16th May 2016

doi: 10.1049/iet-ipr.2016.0271

www.ietdl.org

**Abstract:** Owing to the existence of noise and intensity inhomogeneity in brain magnetic resonance (MR) images, the existing segmentation algorithms are hard to find satisfied results. In this study, the authors propose an improved fuzzy C-mean clustering method (FCM) to obtain more accurate results. First, the authors modify the traditional regularisation smoothing term by using the non-local information to reduce the effect of the noise. Second, inspired by the mechanism of the Gaussian mixture model, the distance function of FCM is defined by using the form of certain exponential function consisting of not only the distance but also the covariance and the prior probability to improve the robustness. Meanwhile, the bias field is modelled by using orthogonal basis functions to reduce the effect of intensity inhomogeneity. Finally, they use the hierarchical strategy to construct a more flexibility function, which considers the improved distance function itself as a sub-FCM, to make the method more robust and accurate. Compared with the state-of-the-art methods, experiment results based on synthetic and real MR images demonstrate its accuracy and robustness.

## 1 Introduction

Accurate segmentation of brain magnetic resonance (MR) images to three main tissues: grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) is fundamental in brain diseases diagnosis. In automatic analysis for brain MR images, segmentation algorithms [1] using computer vision and pattern recognition plays an important role. However, brain MR images are often corrupted by some classical image deterioration, such as noise and intensity inhomogeneity (also named as bias field), which affect the effectiveness of the segmentation methods.

Fuzzy C-means method proposed by Dunn [2] is one of the most widely used clustering methods for images segmentation. It allows the clustering procedure maintain more information from image than hard clustering methods such as K-means [3] and obtain more accurate results. However, it has been proved that [4] the fuzzy C-mean clustering method (FCM) is sensitive to noise without considering any spatial information. Furthermore, the FCM is not robust enough for its Euclidean distance.

Recently, various improved FCM-type clustering schemes [2–11] have been proposed by incorporating spatial constraints to reduce the effect of the noise. Krinidis and Chatzis [4] proposed an algorithm called fuzzy local information C-means (FLICM) by using a fuzzy local similarity measure to reduce the effect of the noise. Pham [12] modified the FCM objective function by introducing a spatial penalty term to estimate the spatially smooth membership.

Gaussian mixture model (GMM) [6, 11, 13–23] is another widely used method for image segmentation, which models the pixel intensities by using a mixed Gaussian distribution. In order to reduce the segmentation sensitivity to noise, Markov random field (MRF) theory has been used to impose spatial information. In [16–18], the complex smoothing prior information is used to reduce the effect of the noise; however, the M-step of expectation maximisation (EM) algorithm cannot be applied directly to the prior distribution. In order to overcome this drawback, Nguyen *et al.* introduced a novel factor to incorporate spatial information between neighbouring pixels into MRF distribution [4]. To

improve the robustness of FCM, many manuscripts [6, 9, 24] modified the Euclidean distance by using Gaussian distribution.

Recently, many manuscripts [6, 12, 25–27] have proved that the effect of the bias field is harder to reduce than that of noise. Pham [12] proposed first- and second-order regularisation terms to reduce the effect of the bias field. However, the coefficients of the regularisation terms are hard to adjust for satisfied results. Li *et al.* [27] proposed a coherent local intensity clustering criterion function to evaluate the classification and the bias field estimation, and used a Gaussian convolution operation to preserve the smoothness of the estimated bias field. Another kind of bias field estimation method is based on the basis functions [6, 28, 29]. The main idea of these methods is selecting basis functions to form linear combination for modelling the bias field.

Following these ideas, Ji *et al.* [6] proposed a robust spatially constrained fuzzy C-means (RSCFCM) algorithm to obtain more accurate results. The RSCFCM used a novel spatial factor, which the proposed spatial factor is constructed based on the posterior probabilities and prior probabilities, and takes the spatial direction into account, to overcome the impact of noise. The RSCFCM can estimate the bias field meanwhile segmenting images. However, the RSCFCM only uses intensity information of pixels on one direction of horizontal, vertical and two diagonal, which makes it may lose details when segmenting object with slender topological structure.

To obtain more robust results, the hierarchical strategy has been proposed to improve mixture models [7, 8]. The hierarchical mixture classifier can provide class conditional density estimates as flat mixtures. Inspired by these ideas, we propose a non-local-based spatially constrained hierarchical fuzzy C-means (NLSCHFCM) algorithm. In NLSCHFCM, a new factor constructed by weighted combination of posterior probabilities and prior probabilities of neighbourhood is incorporated to regularisation term, which makes the NLSCHFCM preserve more abundant details in brain MR images while reducing the effect of the noise effectively. In order to further improve the ability to identify the segmentation for each pixel, the distance function is constructed by using Gaussian distribution. In order to obtain more robust results, we represented the improved distance function by a

sub-FCM of two or three sub-components. The experiments on both synthetic and real brain MR images show that the proposed model can successfully reduce the effect of the noise and bias fields. The NLSCHFCM can obtain more accurate results than those of several state-of-the-art methods.

## 2 Related work

### 2.1 Fuzzy C-means framework

Let  $I$  denote an observed image composed a set of pixels  $\{x_i \in I | i = 1, \dots, N\}$  with dimension  $D$ . The FCM [1] classifies  $I$  into  $K$  clusters based on minimising the energy function

$$J_{\text{FCM}} = \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m x_i - v_k 2^2 \quad (1)$$

where  $v_k$  is the cluster centroid of the  $k$ th class;  $u_{i,k}$  is the corresponding membership, which can measure the membership ratios of pixel  $i$  belonging to the  $k$ th class, and satisfies  $u_{i,k} \in [0, 1]$ ,  $\sum_{k=1}^K u_{i,k} = 1$ ;  $m \in (1, \infty)$  is the fuzzy coefficient.

From (1), we can observe that the standard FCM only uses the intensity information of each pixel, which makes it sensitive to noise and intensity inhomogeneity without any spatial information taking into consideration. In order to reduce the effect of noise, Pham [12] proposed a model called robust fuzzy  $C$ -means algorithm (R\_FCM) by using neighbourhood information of each pixel. The energy function can be written as

$$\begin{aligned} J_{\text{R_FCM}} &= J_{\text{FCM}} + J_{\text{Reg}} \\ &= \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m x_i - v_k 2^2 + \frac{\beta}{2} \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m \sum_{j \in \Omega_i} \sum_{l \in L_j} u_{j,l}^m \end{aligned} \quad (2)$$

where  $\Omega_i$  is the neighbourhood of pixel  $i$ ,  $L_j = \{1, \dots, k-1, k+1, \dots, K\}$ . The parameter  $\beta$  controls the trade-off between the data term and the regularisation term. The R\_FCM can reduce the effect of the noise; however, the size of the neighbourhood is hard to choose when segmenting images with different noise. Furthermore, R\_FCM may loss details in subtraction region easily for using isotropic neighbour information. In order to preserve more detail information, Caldairou *et al.* [5] improved FCM by using the non-local framework (NL\_R\_FCM) (see (3)) where  $W_{i,j}$  is the weighted parameter measuring the similarity between two patches  $P_i$  and  $P_j$

$$W_{i,j} = \frac{1}{H_i} \exp \left( -\frac{|X(P_i) - X(P_j)|^2}{h^2} \right) \quad (4)$$

$P_i$  is the neighbour patch centred at pixel  $i$  with radius size  $S$ ;  $H_i$  is a normalisation to ensure  $\sum W_{i,j} = 1$ ;  $h$  is a non-negative constant; and  $X(P_i)$  is the vector representing the intensity information of patch  $P_i$ . The non-local information used in NL\_R\_FCM contains local region information, which makes it preserve more detail information than single intensity information and makes NL\_R\_FCM can obtain more accurate results. Furthermore, the NL\_R\_FCM can reduce the effect of intensity inhomogeneity by using non-local-based regularisation term. However, from (3), we can find that the distance function is based on Euclidean distance and only use the cluster centroid information, which makes the method inaccurate.

### 2.2 GMMs framework

Clustering algorithms based on finite mixture model have become increasingly popular in recent years. Given the image sampled from continuous random distribution with unknown density  $f(x)$ . A mixture of multivariate normal component densities is typically used to describe the data. Thus,  $f(x)$  can be estimated by using a GMM [6, 13]

$$f(x_i | \theta) = \sum_{k=1}^K \pi_{i,k} \phi(x_i | \theta_k) \quad (5)$$

where  $\pi_{i,k}$  represents the mixing probabilities, which satisfies  $\pi_{i,k} > 0$ ,  $\sum_k \pi_{i,k} = 1$ .  $\phi(x_i | \theta_k)$  is the Gaussian density function

$$\phi(x_i | \theta_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_k|^{1/2}} \exp \left( -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \quad (6)$$

with parameter  $\theta_k = \{\mu_k, \Sigma_k\}$ ,  $k = (1, 2, \dots, K)$ .  $\mu_k$  is the mean and the  $\Sigma_k$  is the covariance. Note that  $x_i$  in (5) is independent, the joint condition density of the image data set can be modelled as

$$p(X | \Pi, \Theta) = \prod_{i=1}^N f(x_i | \Pi, \Theta) = \prod_{i=1}^N \sum_{k=1}^K \pi_{i,k} \phi(x_i | \theta_k). \quad (7)$$

As shown in (5), we can find that the GMM is sensitive to the noise without any spatial information. To deal with this shortcoming, MRF theory [9] is widely used to incorporate the spatial information into classification methods

$$p(\Pi) = \frac{1}{Z} \exp \left( -\frac{1}{T} U(\Pi) \right) \quad (8)$$

where  $U(\Pi)$  is the smoothing prior,  $Z$  is a normalising constant and  $T$  is the temperature constant. With the Bayes' rules, the log-likelihood function can be written as

$$\begin{aligned} L(\Pi, \Theta | X) &= \log(p(\Pi, \Theta | X)) \\ &\propto \log(p(X | \Pi, \Theta)p(\Pi))' \\ &= \sum_{i=1}^N \log \left\{ \sum_{k=1}^K \pi_{i,k} \phi(x_i | \theta_k) \right\} + \log(p(\Pi)) \\ &= \sum_{i=1}^N \log \left\{ \sum_{k=1}^K \pi_{i,k} \phi(x_i | \theta_k) \right\} - \log(Z) - \frac{1}{T} U(\Pi) \end{aligned} \quad (9)$$

and can be maximised by using EM algorithm. Many different ways to select energy  $U(\Pi)$  have been adopted when segmenting different kind of images. Nguyen and Wu [13] have pointed that the prior distribution  $\pi_{i,k}$  cannot be calculated by using the M-step of EM algorithm directly. Thus, the calculation of the prior distribution  $\pi_{i,k}$  needs some computationally complex algorithms. In order to deal with this problem, they proposed a novel factor  $G_{i,k}$  by using prior distributions and posterior probability as

$$G_{i,k} = \exp \left[ \frac{\beta}{2N} \sum_{n \in \partial_i} (z_{n,k}^t + \pi_{n,k}^t) \right] \quad (10)$$

where  $z_{n,k}$  is the posterior probability and  $\beta$  is non-negative constant to control the smoothing prior.  $\partial_i$  is the neighbourhood of the  $i$ th pixel. Then, they introduced the smoothing prior  $U(\Pi)$  by the following equation:

---


$$\begin{aligned} J_{\text{NL_R_FCM}} &= J_{\text{NL_FCM}} + J_{\text{NL_Reg}} \\ &= \sum_{i=1}^N \sum_{k=1}^K \sum_{j \in \Omega_i} W_{i,j} u_{i,k}^m x_i - v_k 2^2 + \frac{\beta}{2} \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m \sum_{j \in \Omega_i} W_{i,j} \sum_{l \in L_j} u_{j,l}^m \end{aligned} \quad (3)$$

$$U(\Pi) = - \sum_{i=1}^N \sum_{k=1}^K G_{i,k}^t \log \pi_{i,k}^{t+1} \quad (11)$$

where  $t$  indicates the iteration step. With this factor, Nguyen proposed a robust spatially constrained GMM (FRSCGMM), which can reduce the effect of noise efficiently. However, from (10), we can find that the method can reduce the effect of the noise by using neighbourhood information easily. However, each pixel in the neighbourhood has same weighting, which makes the improved method easily lose detail information when segmenting objects with slim structure such as CSF in brain MR images.

To obtain more accurate results, Ji *et al.* [6] introduced a new factor  $F_{i,k}$  considering the spatial direction

$$F_{i,k}^t = \exp \left[ \frac{\beta}{2N_i} \sum_{n \in \partial_i^{S_k}} (z_{n,k}^t + \pi_{n,k}^t) \right] \quad (12)$$

where,  $\partial_i^{S_k}$  is the neighbourhood of pixel  $i$  at direction  $S_k$ , which is given by the following equation:

$$S_k^* = \arg \min_{s=[1, S]} \sum_{n \in \partial_i^s} \text{dist}(x_i, v_k^t) \quad (13)$$

$\text{dist}(x_i, v_k^t)$  is the Euclidean distance between pixel  $i$  and cluster centre  $v_k$ .  $\partial_i^s$  is the neighbourhood at direction  $s$  (four directions horizontal, vertical and two diagonal directions). Then, they proposed a fuzzy clustering-type objective function (RSCFCM) by using the new factor and can obtain more accurate results than FRSCGMM. From (12), it can be found that the improved factor only use the information of pixels on one of the four direction and the pixels in each direction have same weights, which makes RSCFCM still lose detail information.

### 3 Non-local-based spatially constrained fuzzy C-means (NLSCFCM)

Motivated by the use of non-local information in NL\_R\_FCM [5] and the constructions of factor in RSCFCM [6], we propose an NLSCFCM algorithm for brain MR image segmentation by introducing a novel non-local-based factor

$$\text{NLF}_{i,k}^t = \exp \left[ \frac{\beta}{2N_i} \sum_{n \in \partial_i} W_{i,n} (z_{n,k}^t + \pi_{n,k}^t) \right] \quad (14)$$

where  $W_{i,n}$  is the weighted parameter calculated by using (4). Therefore, we proposed an improved fuzzy clustering-type objective function based on the novel factor  $\text{NLF}_{i,k}$

$$J_{\text{NLSCFCM}} = \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m (-\log(\pi_{i,k} \phi(x_i | \theta_k))) + \sum_{i=1}^N \sum_{k=1}^K \text{NLF}_{i,k}^t \log(\pi_{i,k}) \quad (15)$$

In [27, 30], the bias field is reconstructed by using the liner combination of basis functions and can be written as

$$B_i = \sum_{l=1}^L q_l \varphi_l(i) = Q^\top \Psi(i) \quad (16)$$

where  $q_l \in R$ ,  $l = 1, \dots, L$ , are the combination coefficients.  $\varphi_l$  is the orthogonal basis function and satisfies:  $\int_\Omega \varphi_l(x) \varphi_j(x) dx = \delta_{l,j}$ ,

$\delta_{i,j} = 1$  for  $i=j$  and  $\delta_{i,j} = 0$  for  $i \neq j$ . Following the idea of [25, 28], we use the orthogonal Legendre polynomials as the basis functions. The size of the coefficients is  $L = (n+1)(n+2)/2$  for two-dimensional (2D) case and  $L = (n+1)(n+2)(n+3)/6$  for 3D case. Here,  $n$  is the degree of Legendre polynomials and depend on prior knowledge of the coil and smoothness of the bias field.

Then, the objective function can be written as

$$J_{\text{NLSCFCM}} = \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m (-\log(\phi(x_i | \theta_k, B_i))) - \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m \log(\pi_{i,k}) + \sum_{i=1}^N \sum_{k=1}^K \text{NLF}_{i,k}^t \log(\pi_{i,k}) \quad (17)$$

where

$$\phi(x_i | \theta_k, B_i) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_k|^{1/2}} \exp \left( -\frac{1}{2} (x_i - B_i \mu_k)^\top \Sigma_k^{-1} (x_i - B_i \mu_k) \right).$$

### 4 Non-local-based spatially constrained hierarchical fuzzy C-means

In this subsection, we introduce a more flexible fuzzy algorithm called NLSCHFCM. The idea is straightforward and easy to implement. We assume the distance function is a sub-fuzzy model and (17) can be written as

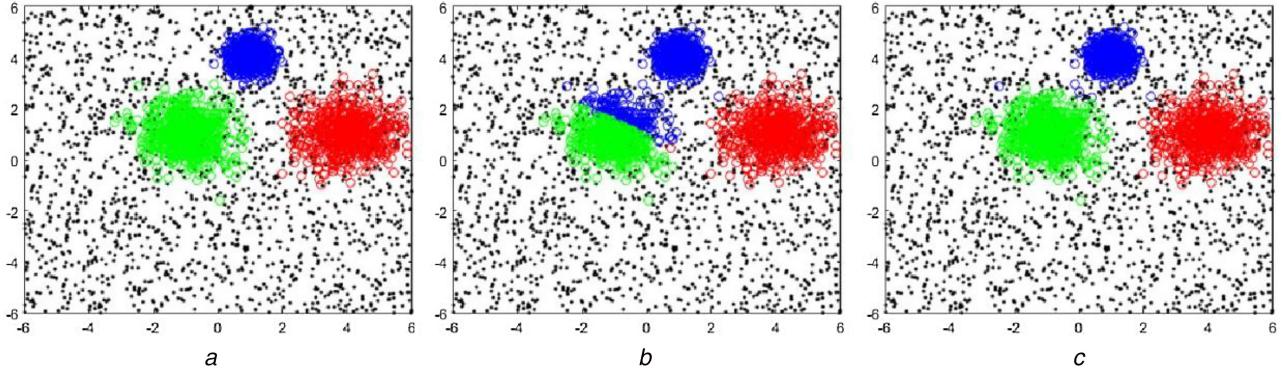
$$J_{\text{NLSCHFCM}} = \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m \sum_{o=1}^O v_{i,k,o}^n d_{iko} + \sum_{i=1}^N \sum_{k=1}^K \text{NLF}_{i,k}^t \log(\pi_{i,k}) \quad (18)$$

where  $d_{iko}$  is the sub-distance function and defined as (see (19))  $v_{i,k,o}$  is the sub-membership and satisfies  $\sum_{o=1}^O v_{i,k,o} = 1$ . It can be seen that the sub-membership  $v_{i,k,o}$  represents the  $o$ th sub-class that belongs to the  $k$ th class. Equation (18) can be considered that the model has two levels: the image data is classified into  $K$  classes in the first level; the data in the  $k$ th class is generated by  $O$  sub-clusters in the second level. When  $o$  in the second level is set as 1, then hierarchical fuzzy C-means (HFCM) is degraded to FCM. So, the HFCM can be regarded as an extension of standard FCM. Furthermore, in HFCM, each point belongs to which class not only based on the distance function, but also based to the sub-component information.

To show the robustness of HFCM, we compared HFCM with standard FCM on a synthetic data, which includes three classes of points from three Gaussian components. Each class has 700 points and the parameters of these three Gaussian distributions are:  $\mu_1 = (-1, 1)^\top$ ,  $\mu_2 = (1, 4)^\top$ ,  $\mu_3 = (4, 1)^\top$ ,  $\Sigma_1 = \text{diag}(1/2, 1/2)$ ,  $\Sigma_2 = \text{diag}(1/7, 1/7)$ , and  $\Sigma_3 = \text{diag}(1/2, 1/2)$ , where  $\mu_i$  is the mean and  $\Sigma_i$  is the covariance. The data is noised by 2100 outliers, which follows the uniform distribution and located in  $[-6, 6]$ . The initial data and outliers (black points) are shown in Fig. 1a. Fig. 1b shows the classification result of FCM. From the results, we can find that due to the effect of the outliers, some points that belong to green class have been misclassified into blue class. Fig. 1c shows the result of HFCM. By using the sub-component information, the HFCM can obtain more accurate results. We use the misclassification error (MCR) to measure the accuracy and the MCRs of FCM and HFCM are 9.8 and 0.28%, respectively. Then, we can conclude that the HFCM improves the classification performance significantly by containing hierarchical information.

---


$$d_{iko} = -\log(\pi_{i,k} \phi(x_i | \theta_{ko}, B_i)) = -\log \left( \pi_{i,k} \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma_{ko}|^{1/2}} \exp \left( -\frac{1}{2} (x_i - B_i \mu_{ko})^\top \Sigma_{ko}^{-1} (x_i - B_i \mu_{ko}) \right) \right) \quad (19)$$



**Fig. 1** Classification results on synthetic data

(a) Original data (three classes) with outliers, (b) Solution of FCM with MCR 9.8%, (c) Solution of HFCM with MCR 0.28%

**Remark 1:** In RSCFCM, from (14), we can find that only one direction can be considered in  $F_{i,k}^t$ , and only the intensity of each pixel in neighbour is used, which makes RSCFCM efficiently, but inaccurate. In our proposed factor, we use the patch information to contain more detailed information. In  $NLF_{i,k}$ , all the pixels in the neighbour have been considered and have different weights, which are calculated by using patch information, to reduce the effect of noise and preserve more details. The calculation of the weights may make NLSCFCM inefficient. From (4), it can be found that the weights can be pre-calculated to reduce the computational complexity.

**Remark 2:** From (3), it can be sound that the non-local information in NL\_R\_FCM [5] is used on traditional Euclidean distance. The dissimilarity function in our proposed model is defined by using the negative log-posterior, which can improve the ability to identify the class for each pixel. The NL\_R\_FCM use the local neighbour information to reduce the effect of intensity inhomogeneity, when the intensity inhomogeneity level is severe, the method is hard to find accurate result. In our method, the bias field estimation has been coupled into the model, which makes the proposed method can estimate the bias field meanwhile segmenting images even with severe intensity inhomogeneity.

**Remark 3:** The proposed method uses the HFCM to improve the robustness of FCM and makes the method less sensitive to the effect of outliers.

The objective function  $J_{NL SCHFCM}$  can be minimised similarly to the standard FCM algorithm. We take the first derivatives of  $J_{NL SCHFCM}$  with respect to  $u$ ,  $\mu$ ,  $\Sigma$  and  $Q$  to zero results.

Using the Lagrange multiplier method, the membership estimator  $u$  and  $v$  can be written as

$$u_{i,k}^t = \frac{\left(\sum_{o=1}^O v_{i,k,o}^n (-\log(\pi_{i,k}\phi(x_i|\theta_{ko}, B_i)))\right)^{1/(1-m)}}{\sum_{j=1}^K \left(\left(\sum_{o=1}^O v_{i,j,o}^n (-\log(\pi_{i,j}\phi(x_i|\theta_{jo}, B_i)))\right)^{1/(1-m)}\right)} \quad (20)$$

$$v_{i,k,o}^t = \frac{(-u_{i,k}^m \log(\pi_{i,k}\phi(x_i|\theta_{ko}, B_i)))^{1/(1-n)}}{\sum_{j=1}^O (-u_{i,j}^m \log(\pi_{i,j}\phi(x_i|\theta_{jo}, B_i)))^{1/(1-n)}} \quad (21)$$

Solving  $(\partial J_{NL SCHFCM}/\partial \mu_{ko})=0$  and  $(\partial J_{NL SCHFCM}/\partial \Sigma_{ko})=0$ , we can obtain

$$\mu_{k,o}^t = \frac{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n B_i x_i}{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n B_i^2} \quad (22)$$

$$\Sigma_{k,o}^t = \frac{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n (x_i - B_i \mu_{k,o})(x_i - B_i \mu_{k,o})^T}{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n} \quad (23)$$

The computation of the conditional expectation values  $z_{i,k}$  in the iteration step  $t$  can be written as

$$z_{i,k}^t = \frac{\pi_{i,k} \phi(x_i|\theta_{ko}, B_i)}{\sum_{j=1}^K \pi_{i,j} \phi(x_i|\theta_{jo}, B_j)} \quad (24)$$

Solving  $(\partial J_{NL SCHFCM}/\partial \pi_{i,k})=0$  with the constraint  $\sum_{k=1}^K \pi_{i,k} = 1$  by using the Lagrange's multiplier method, it can be found

$$\pi_{i,k}^t = \frac{u_{i,k}^m + NLF_{i,k}}{\sum_{j=1}^K u_{i,j}^m + NLF_{i,j}} \quad (25)$$

Solving  $(\partial J_{NL SCHFCM}/\partial Q)=0$ , the combination coefficients can be calculated by the following equation:

$$Q^t = \left( \sum_{i=1}^N \Psi(i) \Psi(i)^T e_1(i) \right)^{-1} \sum_{i=1}^N \Psi(i) e_2(i) \quad (26)$$

where

$$e_1(i) = \sum_{k=1}^K \sum_{o=1}^O v_{i,k,o}^n u_{i,k}^m \mu_{ko}^T \Sigma_{ko}^{-1} \mu_{ko} e_2(i) = \sum_{k=1}^K \sum_{o=1}^O v_{i,k,o}^n u_{i,k}^m x_i^T \Sigma_{ko}^{-1} \mu_k.$$

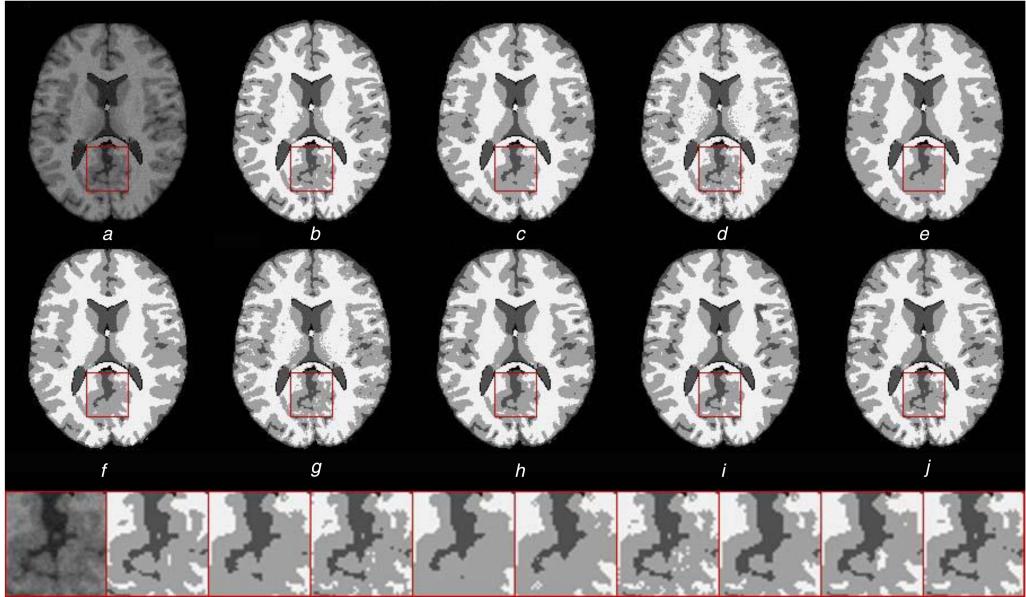
The  $L \times L$  matrix  $\sum_{i=1}^N \Psi(i) \Psi(i)^T e_1(i)$  is non-singular. The detail of derivation of (20)–(26) is given in Appendix 1.

For a deep understanding of our method, we summarise the process as follows:

- Step 1: Initialisation of  $u$ ,  $v$ ,  $\mu$ ,  $\Sigma$  and  $Q$ .
- Step 2: Calculate  $W_{ij}$  for all pixels in the image by using (4).
- Step 3: Update membership function  $u_{i,k}$  by using (20).
- Step 4: Update membership function  $v_{i,k,o}$  by using (21).
- Step 5: Updating conditional expectation value  $z_{i,k}$  by using (24).
- Step 6: Update prior probability  $\pi_{i,k}$  by using (25).
- Step 7: Update the novel factor  $NLF_{i,k}$  by using (14).
- Step 8: Update combination coefficients of the bias field by using (26).
- Step 9: Update centroids and covariance matrices by using (22) and (23).
- Step 10: Check convergence criterion. If convergence has been reached, stop the iteration, otherwise, go to step 3.

## 5 Experimental results

In this section, we segment synthetic and clinical 3T brain MR images into WM, GM and CSF by using the proposed NLSCFCM algorithm. Unless otherwise specified, the parameters used in our experiments are set as follows: radius size of non-local patch is set as  $S=1$ . Radius size  $r$  of searching is 3. The non-negative constant  $h$  is set as 4. Temperature value is set as  $\beta=3$ . The degree of basis function is set as  $n=4$  and then the



**Fig. 2** Segmentation results on the 85th transaxial image of a simulated image data set with the parameter: noise level 3% and intensity inhomogeneity level 0% (N3F0)

(a) Initial image, (b) Ground truth, (c–j) Are the segmentation results of SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively. The bottom shows the details of each red rectangles (from left to right is initial image, ground truth, SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively.)

number of the basis function  $L$  is 15. The fuzzy factor  $m$  and  $n$  take their default values 2.

### 5.1 Evaluation with synthetic data

To show the improvement from our method, we compare our method with seven existing segmentation methods, including three GMM-based methods (SCGM\_EM AQ7 [31], CA\_SVFMM [32], and FRSCGMM [13]) and four FCM-based methods (TMTFCM [10], FLICM [4], RSCFCM [6] and NL\_R\_FCM [5]). The parameters for each algorithm are set with the default values with the default values shown in the papers, which can be seen in Table 1. All the methods are initialised by using  $K$ -means method.

Since SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM and FLICM have not considered the effect of intensity inhomogeneity, we first test on a synthetic brain magnetic resonance imaging (MRI) data set from BrainWeb with the parameter: noise level 3% and intensity inhomogeneity level 0% (N3F0). Fig. 2 shows segmentation results of the 85th transaxial image. Figs. 2a and b show the initial image and the ground truth. Figs. 2c–j show the segmentation results of SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively.

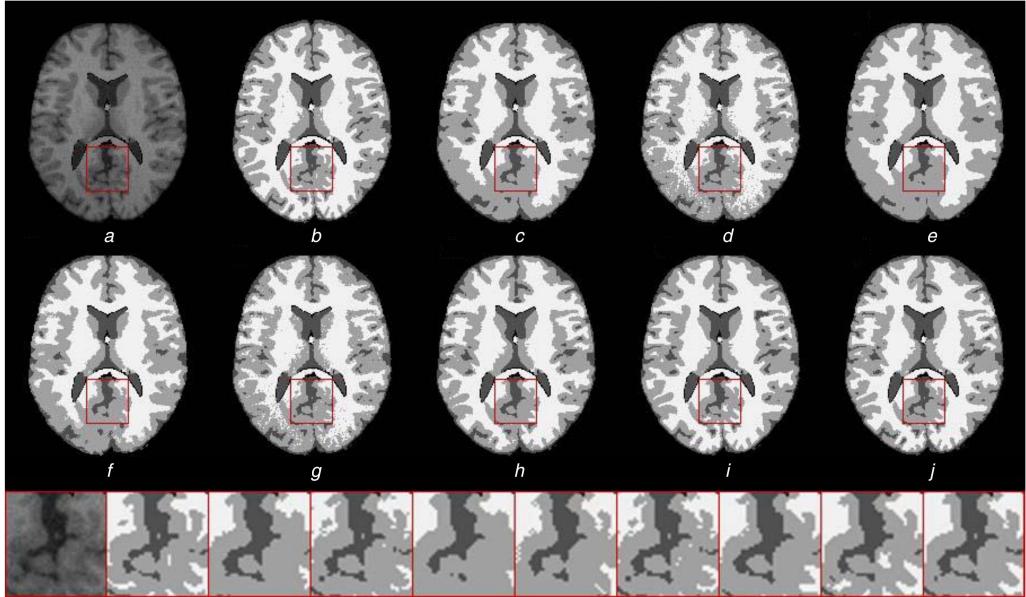
The SCGM\_EM and FRSCGMM use isotropic spatial information, which makes the segmentation results are not satisfied enough. In order to add direction information into finite mixture model (FMM), the CA\_SVFMM (Fig. 2d) improves the log-likelihood function by using local information, however, in order to update the contextual mixing proportions, the CA\_SVFMM needs

to solve a second degree equation, which may make the results unsatisfied [6]. FLICM (Fig. 2g) proposed a factor by incorporating local spatial and local grey level information to reduce the effect of noise. However, the weights of pixels are based on spatial distance and the similar function is based on Euclidean distance, which makes it inaccurate. TMTFCM (Fig. 2f) uses the  $t$ -distribution as the distance function to improve the accuracy. However, it used isotropic neighbourhood information, which makes the method lose details. In order to reduce the effect of noise, the RSCFCM (Fig. 2h) uses neighbourhood information to construct a new factor. However, only one direction of the neighbourhood can be considered and each pixel in the direction has same weight, which makes the method lose details. Fig. 2i is the segmentation result of NL\_R\_FCM. From the result, it can be found some pixels belong to GM have been misclassified into CSF. The bottom of Fig. 2 shows the details of each red rectangles in each segmentation results. Comparing with segmentations obtained by using other algorithms, the NLSCHFCM can visually obtain the best result.

The second experiment is tested on the 85th transaxial image with parameter: N3F60. From the results shown in Fig. 3, we can find that SCGM\_EM (Fig. 3c), CA\_SVFMM (Fig. 3d), FRSCGMM (Fig. 3e), TMTFCM (Fig. 3f) and FLICM (Fig. 3g) are sensitive to intensity inhomogeneity; RSCFCM, NL\_R\_FCM and NLSCHFCM can reduce the effect of intensity inhomogeneity. The RSCFCM and NLSCHFCM use basis functions to estimate the bias field. In NLSCHFCM, we use non-local patch information and hierarchical information to reduce the effect of noise and preserve more detailed information, which makes the bias estimation more accurate than that of RSCFCM. From the results, we can find that our method is better than RSCFCM. The NL\_R\_FCM reduces the effect of intensity inhomogeneity by using non-local information; however, it is based on Euclidean distance, which makes the method inaccurate. The bottom of Fig. 2 shows the details of each red rectangles in each segmentation results. From the details, we can find that our method can obtain more accurate results.

To facilitate the visions, we use Jaccard similarity (JS) [27] as the metric to quantitatively evaluate the segmentation accuracy. The JS is the ratio between intersection and union of the segmented volume  $S_1$  and ground truth volume  $S_2$

$$JS(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \quad (27)$$



**Fig. 3** Segmentation results on the 85th transaxial image of a simulated image data set with the parameter: noise level 3% and intensity inhomogeneity level 60% (N3F60)

(a) Initial image, (b) Ground truth, (c–j) Are the segmentation results of SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively. The bottom shows the details of each red rectangles (from left to right is initial image, ground truth, SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively.)

We apply above eight methods on whole synthetic brain MR image data sets with N3F0, N3F60, N5F60 and N3F80. The mean JS values of WM, GM and CSF are listed in Table 2. The results demonstrate that our method produces the most accurate segmentation (with higher JS values). Our method is more robust to the noise (with higher JS values when noise level is increasing) and has higher robustness to details (with higher JS values for CSF tissue).

## 5.2 Evaluation with clinical data

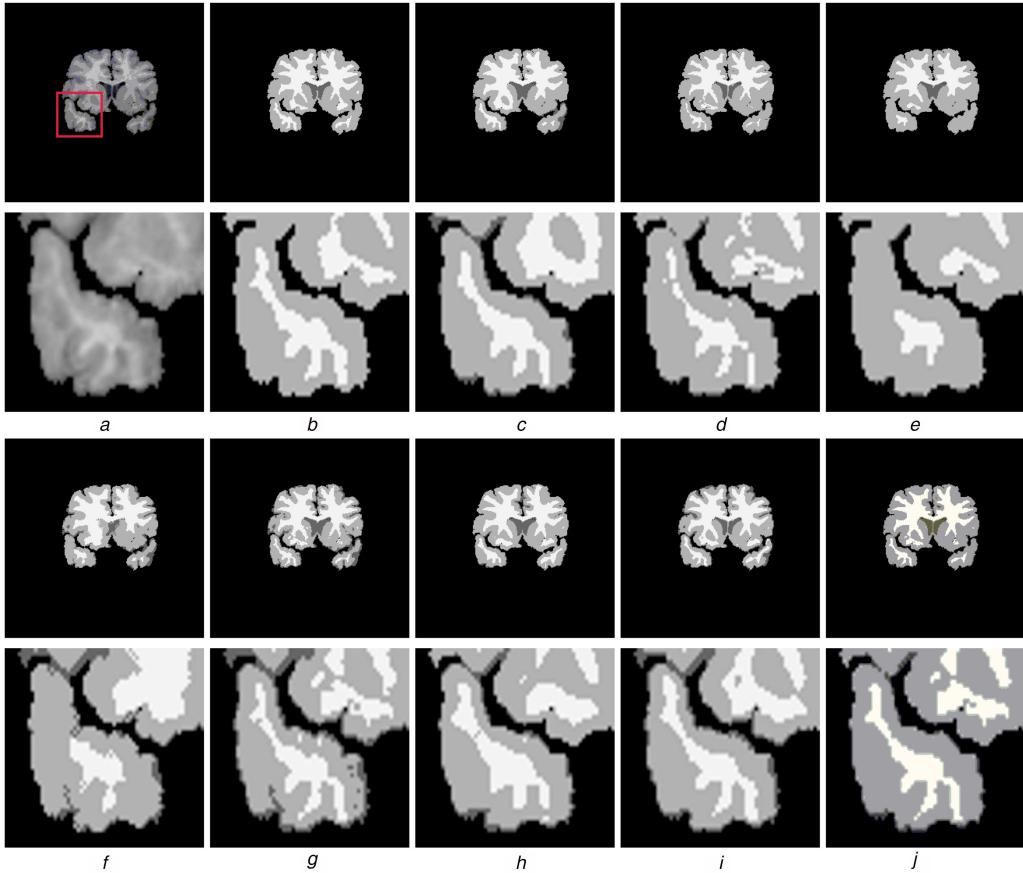
To show the excellence of our method, we compared our method with other methods on a clinical 3 T MR brain image data set from Internet Brain Segmentation Repository (IBSR) (12\_3#). The segmentation results are shown in Fig. 4. Fig. 4a is the 39th of the 12\_3#, which has noise, severe intensity inhomogeneity and weak edges. Fig. 4b is the ground truth. Fig. 4c–j show the segmentation results of SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively. Owing to the effect of the intensity inhomogeneity, SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM and FLICM failed to obtain results. The RSCFCM can reduce the effect of intensity inhomogeneity and obtain more accurate result; however, the weights of pixels in neighbour are same, which makes the method sensitive to weak edges. The NL\_R\_FCM only uses global

centroid information, when the image has weak edges, NL\_R\_FCM is hard to find accurate results. It can be seen from the results in the rectangular region, NL\_R\_FCM misclassified some GM pixels into WM. Comparing with segmentations obtained by using other seven methods, our method can visually obtain the best result and the mean of JS values of the eight methods on ten total IBSR data sets are shown in Table 3. Since there are only small pixels belong to CSF are contained in IBSR data sets, we only calculate the JS values for WM and GM. From the values shown in Table 3, we can find that our method can obtain more accurate results.

To show the robustness of our method, we compared our method with the popular softwares: SPM and FSL on a clinical data. Fig. 5a shows the initial image. It can be found that the initial image has strong noise and severe intensity inhomogeneity. Fig. 5b shows the segmentation result of FSL. The FSL uses the MRF to reduce the effect of noise, however, it is sensitive to the intensity inhomogeneity. The segmentation result of SPM is shown in Fig. 5c. The SPM uses atlas information to reduce the effect of weak edges. Furthermore, the SPM applied intensity inhomogeneity correction to reduce the effect of intensity inhomogeneity, which makes SPM can obtain more accurate result than that of FSL. However, from the result, we can find that many pixels belong to CSF have been misclassified into GM. Fig. 5d

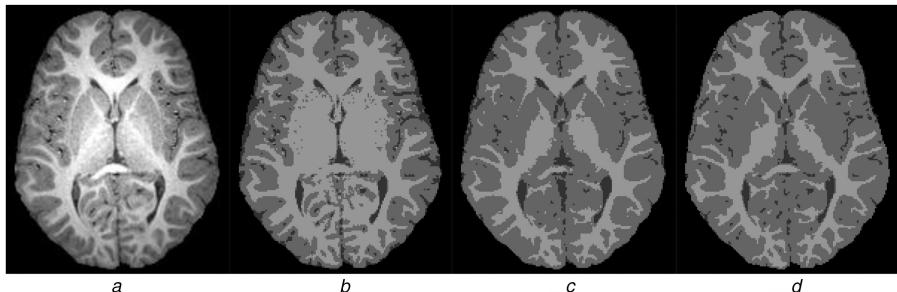
**Table 2** Mean JS values of GM, WM and CSF segmentation obtained by applying eight methods to synthetic brain MR image data sets

		SCGM_EM	CA_SVFMM	FRSCGMM	TMTFCM	FLICM	RSCFCM	NL_R_FCM	NLSCHFCM
N3F0	WM	0.849	0.875	0.810	0.763	0.877	0.843	0.820	0.893
	GM	0.873	0.867	0.851	0.839	0.864	0.846	0.868	0.881
	CSF	0.908	0.892	0.884	0.903	0.893	0.902	0.901	0.915
N3F60	WM	0.813	0.857	0.796	0.753	0.851	0.839	0.836	0.878
	GM	0.778	0.794	0.773	0.764	0.796	0.835	0.876	0.876
	CSF	0.800	0.804	0.783	0.813	0.816	0.889	0.920	0.921
N5F60	WM	0.784	0.787	0.763	0.729	0.776	0.798	0.803	0.853
	GM	0.751	0.706	0.747	0.729	0.706	0.775	0.837	0.843
	CSF	0.780	0.745	0.765	0.797	0.746	0.842	0.891	0.894
N3F80	WM	0.767	0.825	0.780	0.720	0.808	0.839	0.810	0.871
	GM	0.731	0.751	0.742	0.726	0.745	0.829	0.864	0.874
	CSF	0.756	0.760	0.749	0.777	0.765	0.881	0.916	0.917



**Fig. 4** Segmentation results on the clinical MR image

(a) Initial image, (b) Ground truth, (c–j) Are the segmentation results of SCGM\_EM, CA\_SVFMM, FRSCGMM, TMTFCM, FLICM, RSCFCM, NL\_R\_FCM and our method, respectively



**Fig. 5** Segmentation results on the real MR image

(a) Initial image, (b–d) Are the segmentation results of FSL, SPM and our method, respectively

shows the segmentation result of our method. It can be found that our method can obtain more accurate result than FSL and SPM.

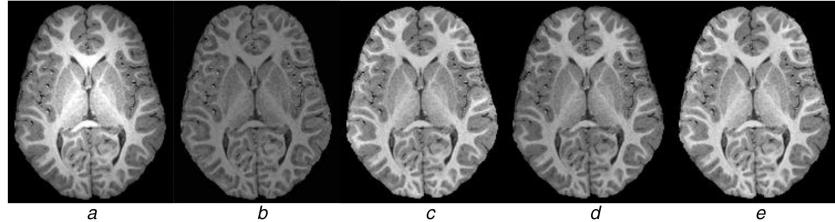
### 5.3 Intensity inhomogeneity correction

To show the ability of bias field correction, we compared our method with Wells method [33], RSCFCM [6] and N3 [34]. Fig. 6 shows the intensity inhomogeneity corrected results on a clinical data. Fig. 6a shows the initial image. Fig. 6b shows the corrected result of Wells method. The Wells method uses a low-pass filter to preserve the smoothness of the bias field, which makes the contrast lower. Fig. 6c shows the corrected result of RSCFCM. The RSCFCM can obtain more accurate result than Wells method by using basis functions. Fig. 6d is the result of N3. Fig. 6e shows the

result of our method. In order to compare the ability of the bias correction with other methods, we use coefficient of variance (CV) as a metric to evaluate the performance of the algorithms. CV is defined as a quotient between standard deviation and mean value of selected tissue class. A good algorithm for bias correction and segmentation should give low CV values for the bias-corrected intensities within each segmented region. The CV values of these images are listed in Table 4. In this experiment, we use FCM to segment the corrected images to calculate CV values. The results shown in Table 4 reflect that the CV values of our method are lower than those of the Wells method, RSCFCM and N3, which indicate that the bias-corrected images of our method are more homogeneous than those of the other two methods.

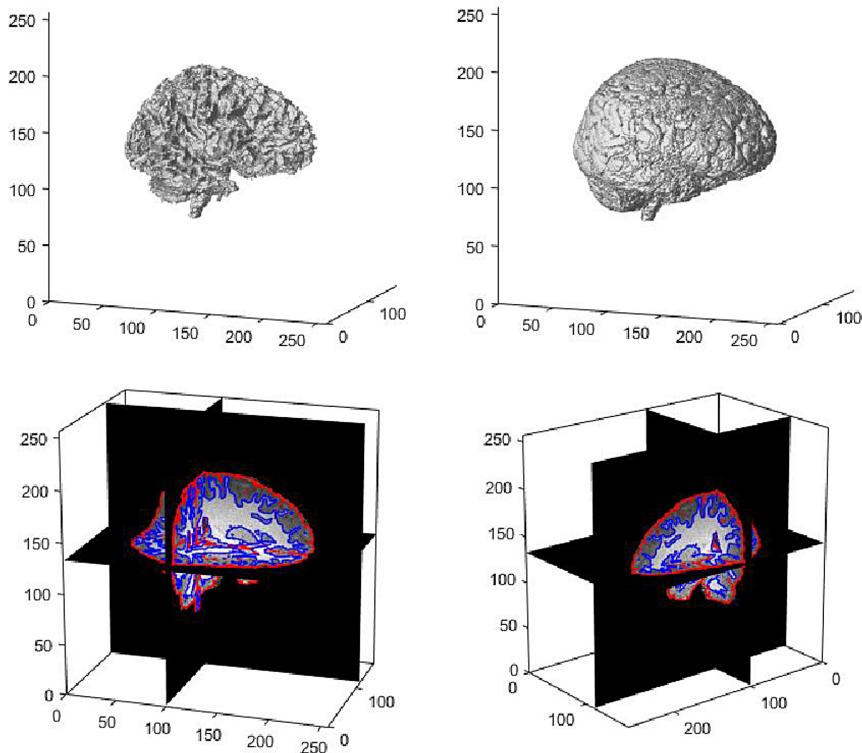
**Table 3** Mean JS values of GM, WM and CSF segmentation obtained by applying eight methods to clinical brain MR image data sets

	SCGM_EM	CA_SVFMM	FRSCGMM	TMTFCM	FLICM	RSCFCM	NL_R_FCM	NLSCHFCM
clinical data	WM 0.840	0.888	0.875	0.724	0.737	0.852	0.840	0.893
	GM 0.810	0.815	0.759	0.742	0.844	0.828	0.751	0.853



**Fig. 6** Intensity inhomogeneity correction on the real MR image

(a) Initial image, (b-d) Are corrected results of Wells method, RSCFCM, N3 and our method, respectively



**Fig. 7** 3D segmentation results of the GM and WM on clinical brain MR data set

#### 5.4 Segmentation results on clinical 3D images

Fig. 7 shows the 3D segmentation results of our method for the clinical data, which has severe intensity inhomogeneity and noise (Fig. 5a is the 133th of the data). The first two rows show the evolutions of the WM and GM surfaces, respectively. To better view the intermediate results, we also present the edges of the result for three slices of different axis, as shown in the second row. It can be observed that satisfactory result has been obtained by our method.

## 6 Discussion

In our experiments, we only analysed the segmentation results on the skull stripped synthetic and real MR images since the skull stripped image can avoid the interference of inter subject variation of non-brain structures. All the brain MR images are all skull stripped by using the method proposed by Shi *et al.* [35]. Fig. 8 shows the segmentation of our method on three skull stripped MR slices generated from BrainWeb (N3F80), together with the estimated bias fields, bias corrected images and segmentation results. From the results, we can find that the intensities within each brain tissue in the bias corrected images become quite homogeneous. Fig. 9 shows the segmentation of our method on

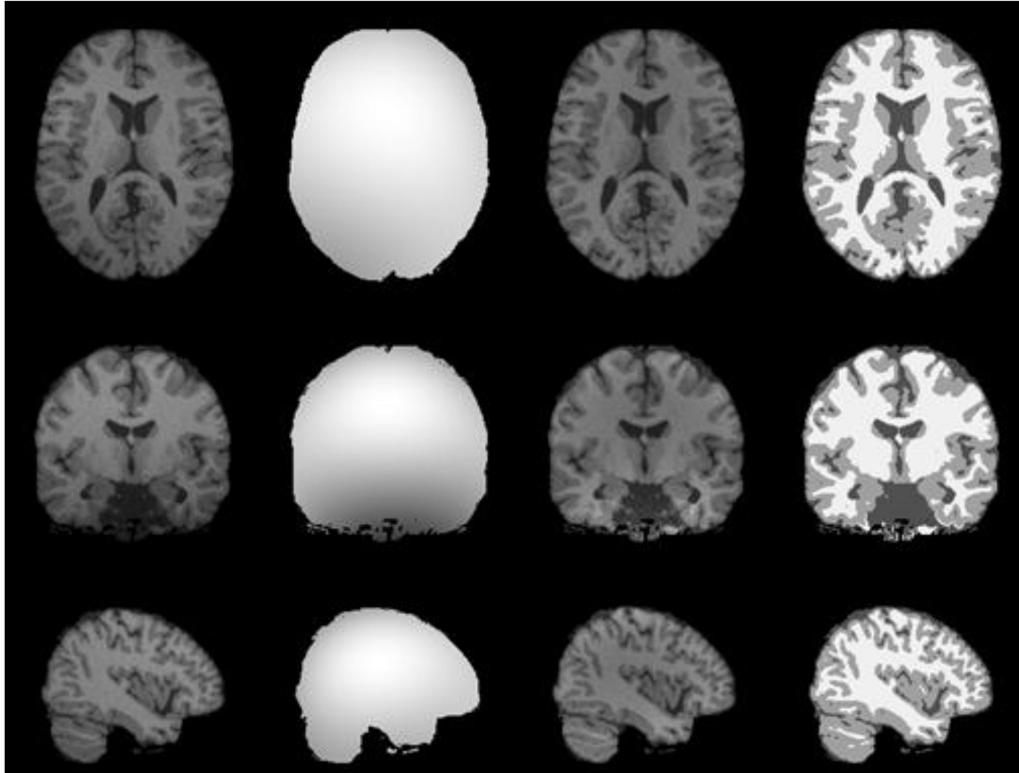
three MR slices with skulls. It is clear that our method can still obtain satisfactory results without being influenced by the skulls.

The degree of basis function determines the accuracy and stability of the calculated bias field. A much lower degree will make the estimated bias field inaccurate when image has severe intensity inhomogeneity. A too large degree will make our method inefficient, unstable and easily trapped into local optima. Our experiments showed that the degree of basis functions up to the four degree sufficiently model the bias field.

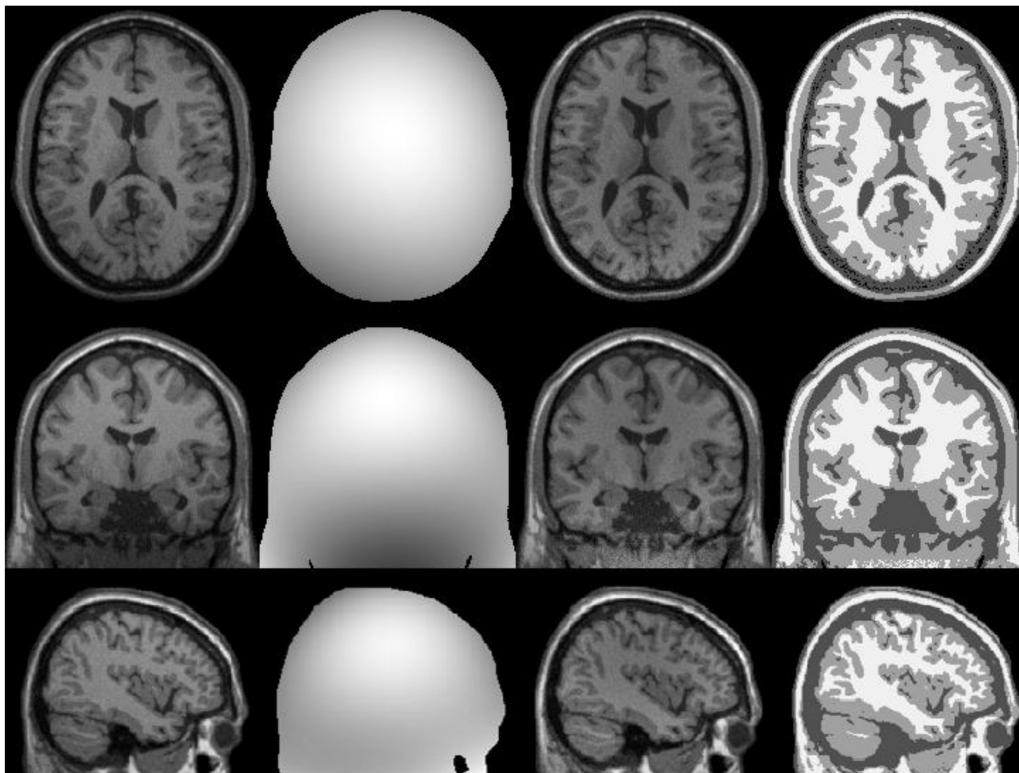
In our method, the non-local information is used to reduce the effect of the noise. The weight  $W_{ij}$  in (4) will never change after it has been calculated, so it needs to be calculated only once. We also analysed the relationship between the parameters of  $W_{ij}$  and the misclassification error (MCR). In this paper, we set the size of non-local patch  $S$  as 1 and the size of searching size  $r$  as 3. We analysed the effect of the parameters on a whole simulated image data set with noise level 4% and inhomogeneity level 60% and the results are shown in Fig. 10a. From the results, we can find that the MCR of the results with the different local region parameters and different search region parameters can affect the accuracy of our method and when  $S=1$ , our method can obtain more accurate results. This is because the brain tissues have more topological changes in the images. We also analysed the effect of the parameter

**Table 4** Coefficient of variation (%)

Figure	Tissue	Wells	RSCFCM	N3	NLSCHFCM
Fig. 6	WM	7.97	7.53	8.28	7.14
	GM	8.82	8.79	9.03	8.73



**Fig. 8** Illustration of three skull stripped 3T-weighted brain MR images (first column), the estimated bias fields (second column), bias-corrected images (third column) and the segmentation results of our method (fourth column)

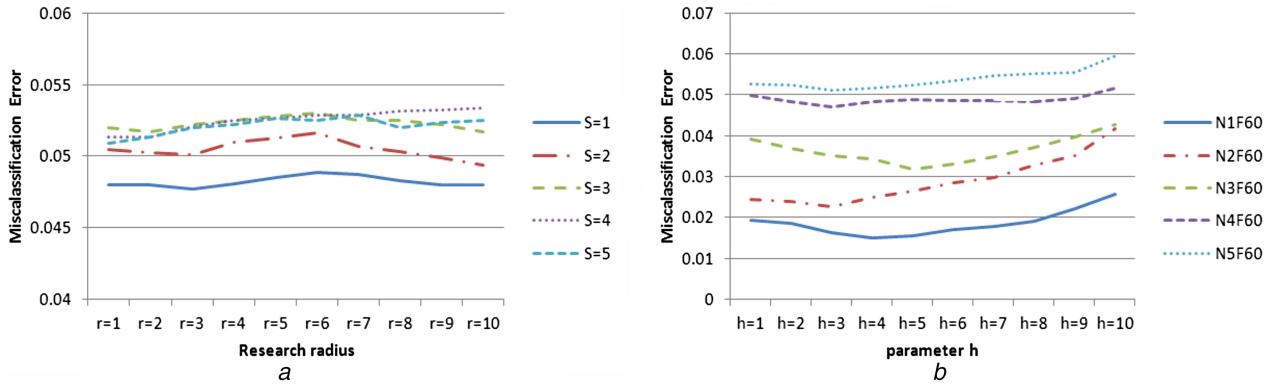


**Fig. 9** Illustration of three 3T-weighted brain MR images (first column), the estimated bias fields (second column), bias-corrected images (third column) and the segmentation results of our method (fourth column)

$h$  in (4) on synthetic MRI data sets: N1F60 (noise level is 1% and inhomogeneity level 60%), N2F60, N3F60, N4F60 and N5F60. In this experiment, we set  $S=1$  and  $r=3$ . The results are shown in Fig. 10b and we can find that when  $h$  is located in [3, 6], our method can obtain satisfactory results.

## 7 Conclusion

In this paper, we proposed the NLSCHFCM algorithm for brain MR image segmentation. The proposed algorithm can reduce the effect of noise by introducing a novel factor considering the non-local information and uses the negative log-posterior as the dissimilarity function to improve the accuracy of the method. Furthermore, our method can estimate the bias field meanwhile segmenting the image and has the ability of preserving the details in brain MR image. In order to obtain more robust and accurate



**Fig. 10** Misclassified error of the segmentation results on simulated MR images with different parameters of the non-local framework  
(a) MCRs of simulated data set N4F60 with different  $p$  and  $r$ , (b) MCRs of simulated data sets N1F60, N2F60, N3F60, N4F60 and N5F60 with different  $h$

results, we use the hierarchical strategy to construct a more flexibility function, which considers the improved distance function itself as a sub-FCM. The proposed method can overcome the draw backs of over-smoothness for segmentations. Experimental results on both synthetic and clinical images have shown that our method outperforms several state-of-the-art segmentation methods when segmenting images with intensity inhomogeneities and noise.

## 8 Acknowledgment

This work was supported in part by the National Nature Science Foundation of China 61572257 and 61402235; the National Research Foundation of Korea under grant NRF-2013K2A2S2000777; the Natural Science Foundation of Jiangsu Province BY2014007-04 and the University Science Research Project of Jiangsu Province 13KJB520016.

## 9 References

- [1] Bezdek, J.C.: ‘Pattern recognition with fuzzy objective function algorithms’ (Kluwer Academic Publishers, Norwell, MA, USA, 1981)
- [2] Dunn, J.C.: ‘A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters’, *J. Cybern.*, 1973, **3**, (3), pp. 32–57
- [3] Mignotte, M.: ‘A de-texturing and spatially constrained K-means approach for image segmentation’, *Pattern Recognit. Lett.*, 2011, **32**, pp. 359–367
- [4] Krinidis, S., Chatzis, V.: ‘A robust fuzzy local information C-means clustering algorithm’, *IEEE Trans. Image Process.*, 2010, **5**, (19), pp. 1328–1337
- [5] Calairou, B., Passat, N., Habas, P.A., et al.: ‘A non-local fuzzy segmentation method: application to brain MRI’, *Pattern Recognit.*, 2011, **44**, pp. 1916–1927
- [6] Ji, Z., Liu, J., Cao, G., et al.: ‘Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation’, *Pattern Recognit.*, 2014, **47**, pp. 2454–2466
- [7] Titsias, M.K., Likas, A.: ‘Mixture of experts classification using a hierarchical mixture model’, *Neural Comput.*, 2002, **14**, (9), pp. 2221–2244
- [8] Di Zio, M., Guarnera, U., Rocci, R.: ‘A mixture of mixture models for a classification problem: the unity measure error’, *Comput. Stat. Data Anal.*, 2007, **51**, (5), pp. 2573–2585
- [9] Chatzis, S., Varvarigou, T.A., Fuzzy, A.: ‘Clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation’, *IEEE Trans. Fuzzy Syst.*, 2008, **16**, pp. 1351–1361
- [10] Zhang, H., Jonathan, Q.M., Wu, T.M.N.: ‘A robust fuzzy algorithm based on Student’s t-distribution and mean template for image segmentation application’, *IEEE Signal Process. Lett.*, 2013, **20**, pp. 117–120
- [11] Ji, Z.X., Xia, Y., Sun, Q.S., et al.: ‘Fuzzy local Gaussian mixture model for brain MR image segmentation’, *IEEE Trans. Inf. Technol. Biomed.*, 2012, **16**, pp. 339–347
- [12] Pham, D.L.: ‘Spatial models for fuzzy clustering’, *Comput. Vis. Image Underst.*, 2001, **84**, (2), pp. 285–297
- [13] Nguyen, T., Wu, Q.: ‘Fast and robust spatially constrained Gaussian mixture model for image segmentation’, *IEEE Trans. Circuits Syst. Video Technol.*, 2013, **23**, pp. 621–635
- [14] Ju, Z., Liu, H.: ‘Fuzzy Gaussian mixture models’, *Pattern Recognit.*, 2012, **45**, pp. 1146–1158
- [15] Dong, F., Peng, J.: ‘Brain MR image segmentation based on local Gaussian mixture model and nonlocal spatial regularization’, *J. Vis. Commun. Image Represent.*, 2014, **25**, pp. 827–839
- [16] Li, S.Z.: ‘Markov random field modeling in imaging analysis’ (Springer-Verlag, New York, 2009)
- [17] Sanjay, G.S., Hebert, T.J.: ‘Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm’, *IEEE Trans. Image Process.*, 1998, **7**, (7), pp. 1014–1028
- [18] Blekas, K., Likas, A., Galatsanos, N.P., et al.: ‘A spatially constrained mixture model for image segmentation’, *IEEE Trans. Neural Netw.*, 2005, **16**, (2), pp. 494–498
- [19] Zhang, H., Jonathan, Wu, Q.M., Nguyen, T.M., et al.: ‘Synthetic aperture radar image segmentation by modified student’s t-mixture model’, *IEEE Trans. Geosci. Remote Sens.*, 2014, **52**, (7), pp. 4391–4403
- [20] Zhang, H., Jonathan, Wu, Q.M., Nguyen, T.M.: ‘Incorporating mean template into finite mixture model for image segmentation’, *IEEE Trans. Neural Netw. Learn. Syst.*, 2013, **24**, (2), pp. 328–335
- [21] Jia, J., Li, X., Yang, B., et al.: ‘Segmentation-based image copy-move forgery detection scheme’, *IEEE Trans. Inf. Forensics Sec.*, 2015, **10**, (3), pp. 507–518
- [22] Pan, Z., Zhang, Y., Kwong, S.: ‘Efficient motion and disparity estimation optimization for low complexity multiview video coding’, *IEEE Trans. Broadcast.*, 2015, **61**, (2), pp. 166–176
- [23] Fu, Z., Ren, K., Shu, J., et al.: ‘Enabling personalized search over encrypted outsourced data with efficiency improvement’, *IEEE Trans. Parallel Distrib. Syst. (TPDS)*, doi: 10.1109/TPDS.2015.2506573
- [24] Zheng, Y., Jeon, B., Xu, D., et al.: ‘Image segmentation by generalized hierarchical fuzzy C-means algorithm’, *J. Intell. Fuzzy Syst.*, 2015, **28**, (2), pp. 961–973
- [25] Vovk, U., Permut, F., Likar, B.: ‘A review of methods for correction of intensity inhomogeneity in MRI’, *IEEE Trans. Med. Imaging*, 2007, **26**, pp. 405–421
- [26] García-Sebastián, M., Fernández, E., Graña, M., et al.: ‘A parametric gradient descent MRI intensity inhomogeneity correction algorithm’, *Pattern Recognit. Lett.*, 2007, **28**, (13), pp. 1657–1666
- [27] Li, C., Xu, C., Anderson, A., et al.: ‘MRI tissue classification and bias field estimation based on coherent local intensity clustering: a unified energy minimization framework’, *Information Processing in Medical Imaging*, Springer, 2009, pp. 288–299
- [28] Zhu, S.-C., Yuille, A.: ‘Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (9), pp. 884–900
- [29] Styner, M., Brechbuhler, C., Szekely, G., et al.: ‘Parametric estimate of intensity in homogeneities applied to MRI’, *IEEE Trans. Med. Imaging*, 2000, **19**, (3), pp. 153–165
- [30] Chen, Y., Zhang, J., Wang, S., et al.: ‘Brain magnetic resonance image segmentation based on an adapted non-local fuzzy c-means method’, *IET Comput. Vis.*, 2012, **6**, (6), pp. 610–625
- [31] Diplaros, A., Vlassis, N., Gevers, T.: ‘A spatially constrained generative model and an EM algorithm for image segmentation’, *IEEE Trans. Neural Netw.*, 2007, **18**, pp. 798–808
- [32] Nikou, C., Galatsanos, N., Likas, A.: ‘A class-adaptive spatially variant mixture model for image segmentation’, *IEEE Trans. Image Process.*, 2007, **16**, pp. 1121–1130
- [33] Wells, W., Grimson, W., Kikinis, R.: ‘Adaptive segmentation of MRI data’, *IEEE Trans. Image Process.*, 1996, **15**, (4), pp. 429–442
- [34] Sled, J., Zijdenbos, A., Evans, A.: ‘A nonparametric method for automatic correction of intensity nonuniformity in MRI data’, *IEEE Trans. Imag. Process.*, 1998, **17**, (1), pp. 87–97
- [35] Shi, F., Wang, L., Dai, Y., et al.: ‘Pediatric brain extraction using learning-based meta-algorithm’, *Neuroimage*, 2012, **62**, pp. 1975–1986

## 10 Appendix

### 10.1 Appendix 1: Derivation of gradient flow

In this Appendix, we give the formal derivation for (20)–(26). The fuzzy membership and sub-membership can be obtained by minimising the objective function  $J_{\text{NLSCHFCM}}$  over  $u_{i,k}$  and  $v_{i,k,o}$  under the constraints  $\sum_{k=1}^K u_{i,k} = 1$ ,  $\sum_{o=1}^O v_{i,k,o} = 1$ . Then, we can obtain

$$J'_{\text{NLSCHFCM}} = J_{\text{NLSCHFCM}} + \lambda_1 \left( 1 - \sum_{k=1}^K u_{i,k} \right) + \lambda_2 \left( 1 - \sum_{o=1}^O v_{i,k,o} \right) \quad (28)$$

Taking the derivative of  $J_{\text{NLSCHFCM}}$  with respect to  $u$  and  $v$  setting the result to zero, we have

$$\frac{\partial J'_{\text{NLSCHFCM}}}{\partial u_{i,k}} = m u_{i,k}^{m-1} \sum_{o=1}^O v_{i,k,o}^n d_{iko} - \lambda_1 = 0 \quad (29)$$

Then, we can obtain

$$u_{i,k} = \left( \frac{\lambda_1}{m \sum_{o=1}^O v_{i,k,o}^n d_{iko}} \right)^{(1/(m-1))} \quad (30)$$

As  $\sum_{k=1}^K u_{i,k} = 1$ ,  $\lambda_1$  can be calculated as

$$\lambda_1 = \left( \frac{1}{\sum_{k=1}^K ((1/m \sum_{o=1}^O v_{i,k,o}^n d_{iko})^{1/(m-1)})} \right)^{m-1} \quad (31)$$

Substituting (31) into (30), the zero-gradient condition for the membership  $u$  can be rewritten as

$$u_{i,k}^t = \frac{(\sum_{o=1}^O v_{i,k,o}^n (-\log(\pi_{i,k} \phi(x_i | \theta_{ko}, B_i))))^{1/(1-m)}}{\sum_{j=1}^K ((\sum_{o=1}^O v_{i,j,o}^n (-\log(\pi_{i,j} \phi(x_i | \theta_{jo}, B_i))))^{1/(1-m)})} \quad (32)$$

Similarly, processing on sub-membership, we can obtain

$$v_{i,k,o}^t = \frac{(-u_{i,k}^m \log(\pi_{i,k} \phi(x_i | \theta_{ko}, B_i)))^{1/(1-n)}}{\sum_{j=1}^O (-u_{i,j}^m \log(\pi_{i,j} \phi(x_i | \theta_{jo}, B_i)))^{1/(1-n)}} \quad (33)$$

For fixed  $u$  and  $v$ , taking the derivative of  $J_{\text{NLSCHFCM}}$  with respect to  $\mu_{ko}$  and setting the result to zero, we have

$$\begin{aligned} & \frac{\partial J_{\text{NLSCHFCM}}}{\partial \mu_{ko}} \\ & \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m \sum_{o=1}^O v_{i,k,o}^n (-\log(\pi_{i,k} (1/(2\pi)^{(D/2)})(1/|\Sigma_{ko}|^{1/2})) \\ & \exp(-(1/2)(x_i - B_i \mu_{ko})^\top \Sigma_{ko}^{-1} (x_i - B_i \mu_{ko}))) \\ & \frac{\partial}{\partial \mu_{ko}} \\ & u_{i,k}^m v_{i,k,o}^n B_i (x_i - B_i \mu_{ko}) = 0 \end{aligned}$$

Solving for  $\mu_{ko}$ , we have

$$\mu_{k,o}^t = \frac{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n B_i x_i}{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n B_i^2 x_i} \quad (34)$$

Taking the derivative of  $J_{\text{NLSCHFCM}}$  with respect to  $\Sigma_{ko}^{-1}$ , we have

$$\begin{aligned} & \frac{\partial J_{\text{NLSCHFCM}}}{\partial \Sigma_{ko}^{-1}} \\ & \sum_{i=1}^N u_{i,k}^m \sum_{o=1}^O v_{i,k,o}^n (-\log(\pi_{i,k} (1/(2\pi)^{(D/2)})(1/|\Sigma_{ko}|^{1/2})) \\ & \exp(-(1/2)(x_i - B_i \mu_{ko})^\top \Sigma_{ko}^{-1} (x_i - B_i \mu_{ko})) \\ & \frac{\partial}{\partial \Sigma_{ko}^{-1}} \\ & \partial (\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n ((1/2) \log(|\Sigma_{ko}|) + (1/2)(x_i - B_i \mu_{ko})^\top \Sigma_{ko}^{-1} (x_i - B_i \mu_{ko}))) \\ & \frac{\partial}{\partial \Sigma_{ko}^{-1}} \\ & \Rightarrow \sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n (x_i - B_i \mu_{ko}) (x_i - B_i \mu_{ko})^\top = \sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n \Sigma_{i,k,o} \end{aligned}$$

Solving for  $\Sigma_{k,o}$ , we have

$$\Sigma_{k,o}^t = \frac{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n (x_i - B_i \mu_{ko}) (x_i - B_i \mu_{ko})^\top}{\sum_{i=1}^N u_{i,k}^m v_{i,k,o}^n} \quad (35)$$

The conditional expectation value  $z_{i,k}$  is the posterior probability of  $i$ , then it can be calculated as

$$\begin{aligned} z_{i,k} &= p(k | i) = \frac{p(k,i)}{p(i)} = \frac{p(k)p(i|k)}{\sum_k^K p(k)p(i|k)} \\ &= \frac{\pi_{i,k} \phi(x_i | \theta_{ko}, B_i)}{\sum_{j=1}^K \pi_{i,j} \phi(x_i | \theta_{jo}, B_j)} \end{aligned} \quad (36)$$

Solving  $(\partial J_{\text{NLSCHFCM}} / \partial \pi_{i,k}) = 0$  with the constraint  $\sum_{k=1}^K \pi_{i,k} = 1$  by using the Lagrange's multiplier method, it can be found

$$\begin{aligned} \frac{\partial J_{\text{NLSCHFCM}}}{\partial \pi_{i,k}} &= \frac{\partial \left( \sum_{i=1}^N \sum_{k=1}^K u_{i,k}^m (-\log(\pi_{i,k})) \right)}{\partial \pi_{i,k}} = \\ & \frac{u_{i,k}^m}{\pi_{i,k}} + \frac{\text{NLF}_{i,k}}{\pi_{i,k}} - \lambda_3 = 0 \end{aligned} \quad (37)$$

Then, we can obtain

$$\pi_{i,k}^t = \frac{u_{i,k}^m + \text{NLF}_{i,k}}{\lambda_3} \quad (38)$$

As  $\sum_{k=1}^K \pi_{i,k} = 1$ ,  $\lambda_3$  can be calculated as

$$\lambda_3 = \sum_{j=1}^K u_{i,j}^m + \text{NLF}_{i,j} \quad (39)$$

Substituting (39) into (28), the zero-gradient condition for  $\pi_{i,k}$  can be rewritten as

$$\pi_{i,k}^t = \frac{u_{i,k}^m + \text{NLF}_{i,k}}{\sum_{j=1}^K u_{i,j}^m + \text{NLF}_{i,j}} \quad (40)$$

Solving  $(\partial J_{\text{NLSCHFCM}} / \partial Q) = 0$ , the combination coefficients can be calculated by the following equation:

$$\begin{aligned} & \frac{v_{i,k,o}^n (-\log(\pi_{i,k} (1/(2\pi)^{(D/2)})(1/|\Sigma_{ko}|^{1/2})))}{\partial Q} \\ & \sum_{k=1}^K u_{i,k}^m \sum_{o=1}^O v_{i,k,o}^n (x_i - Q^\top \Psi(i) \mu_{ko})^\top \Sigma_{ko}^{-1} (x_i - Q^\top \Psi(i) \mu_{ko}) \\ & \frac{\partial}{\partial Q} \\ & ) \Psi(i)^\top \sum_{k=1}^K \sum_{o=1}^O u_{i,k}^m v_{i,k,o}^n \mu_{ko}^\top \Sigma_{ko}^{-1} \mu_{ko} Q = \sum_{i=1}^N \Psi(i) \sum_{k=1}^K \sum_{o=1}^O u_{i,k}^m v_{i,k,o}^n \end{aligned}$$

Then, we can obtain

$$Q^t = \left( \sum_{i=1}^N \Psi(i) \Psi(i)^\top e_1(i) \right)^{-1} \sum_{i=1}^N \Psi(i) e_2(i) \quad (41)$$

where

$$\begin{aligned} e_1(i) &= \sum_{k=1}^K \sum_{o=1}^O v_{i,k,o}^n u_{i,k}^m \mu_{ko}^\top \Sigma_{ko}^{-1} \mu_{ko} e_2(i) = \sum_{k=1}^K \sum_{o=1}^O \\ & v_{i,k,o}^n u_{i,k}^m x_i^\top \Sigma_{ko}^{-1} \mu_{ko} \end{aligned}$$

The  $L \times L$  matrix  $\sum_{i=1}^N \Psi(i) \Psi(i)^\top e_1(i)$  is non-singular and the proof can be seen in [27].