Similarity Measure-Based Possibilistic FCM With Label Information for Brain MRI Segmentation

Xiangzhi Bai, *Member, IEEE*, Yuxuan Zhang, Haonan Liu, and Zhiguo Chen

***Abstract*—Magnetic resonance imaging (MRI) is extensively applied in clinical practice. Segmentation of the MRI brain image is significant to the detection of brain abnormalities. However, owing to the coexistence of intensity inhomogeneity and noise, dividing the MRI brain image into different clusters precisely has become an arduous task. In this paper, an improved possibilistic fuzzy *c*-means (FCM) method based on a similarity measure is proposed to improve the segmentation performance for MRI brain images. By introducing the new similarity measure, the proposed method is more effective for clustering the data with nonspherical distribution. Besides that, the new similarity measure could alleviate the "cluster-size sensitivity" problem that most FCM-based methods suffer from. Simultaneously, the proposed method could preserve image details as well as suppress image noises via the use of local label information. Experiments conducted on both synthetic and clinical images show that the proposed method is very effective, providing mitigation to the cluster-size sensitivity problem, resistance to noisy images, and applicability to data with more complex distribution.**

***Index Terms*—Label information, magnetic resonance imaging (MRI) brain image, possibilistic fuzzy *c*-means (PFCM), segmentation, similarity measure.**

I. Introduction

THE magnetic resonance imaging (MRI) technique is widely used in the field of medicine these days. As one of the most common ways to visualize a sophisticated brain structure, the MRI technique contributes a lot to diagnosis brain disease, such as brain tumor, paralysis, and stroke [1], [2]. Hence, research on segmentation of the MRI brain image has drawn extensive attention [3]. The tissues of primary concern in the MRI brain image include gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Accurate segmentation of those tissues is of great significance in the medical domain, specifically in clinical practice [3]. MRI brain images are nonlinear and nonspherical data, and they are usually accompanied by various artifacts, such as noise, partial volume effect, and intensity inhomogeneity. This makes it hard to segment MRI brain image precisely [3], [4]. Another insurmountable fact is that CSF accounts for a very small proportion of most MRI brain images. This is another obstacle for all fuzzy *c*-means (FCM)-based methods when segmenting MRI brain images [5].

Clustering is an efficient unsupervised method to achieve image segmentation. Clustering methods are very popular for segmentation of the MRI brain images due to their simplicity. Clustering methods aim at decomposing a given dataset into some clusters based on minimizing the objective function iteratively [6]. The procedure of clustering generally splits objects and patterns into different clusters where similarity between samples from the same class is higher than samples from different clusters [7], [8]. Clustering methods can be mainly divided into hard clustering and fuzzy clustering. Fuzzy clustering is more suitable for images with uncertainty and its typical applications cover various areas, such as remote sensing, medical image, etc. [9], [10]. FCM is the classical fuzzy clustering method and it performs well on lots of noise-free images [11]. However, without taking spatial information into consideration, FCM is relatively sensitive to noise. And FCM tends to equalize cluster populations in the clustering process [7], which we call "cluster-size sensitivity." To ameliorate the noise sensitivity, many modified FCM approaches have been proposed.

Ahmed *et al.* [12] proposed an FCM_S algorithm, which adds a neighborhood regularization term to suppress noises [8]. FCM_S is effective for images which contain salt and pepper noise. But image detail preserving and noise resistance are still tough choices for FCM_S. Pal *et al.* [13] proposed a new method called the possibilistic FCM (PFCM) algorithm. PFCM considers not only memberships but also possibilities, which are less sensitive to outliers than FCM. PFCM can be seen as a hybridization of the possibilistic *c*-means (PCM) [14] and FCM. PFCM could describe the data more informatively than FCM [13]. However, the performance of PFCM relies on parameter selection. Park [15] put forward the intuitive FCM (IFCM) model. To alleviate the effect of noise, IFCM introduces the intuition level to modify the form of membership. IFCM can identify the inliers to some extent [15]. However, IFCM is still sensitive to noise because it does not consider spatial information.

To control the influence of local information, Krinidis and Chatzis [8] presented a robust fuzzy local

Manuscript received November 15, 2017; revised March 3, 2018; accepted April 24, 2018. This work was supported by the National Natural Science Foundation of China under Grant U1736217. This paper was recommended by Associate Editor P. P. Angelov. *(Corresponding author: Xiangzhi Bai.)*

X. Bai is with the Image Processing Center, Beihang University, Beijing 100191, China, also with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China, and also with the Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China (e-mail: jackybxz@buaa.edu.cn).

Y. Zhang, H. Liu, and Z. Chen are with the Image Processing Center, Beihang University, Beijing 100191, China (e-mail: zhangyuxuan1996@buaa.edu.cn; lhnmailbox@qq.com; chenzhiguo6313@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2830977

information c-means (FLICM) algorithm. FLICM utilizes gray-level information by defining a new fuzzy factor. FLICM is more robust to noisy images and preserves image details well. Gong *et al.* [16] made an improvement on FLICM and raised up the KWFLICM algorithm. KWFLICM introduces a kernel measure and utilizes more local context information, which further reduces the effect of noises and outliers [16]. KWFLICM could produce satisfactory segmentation results sometimes. However, KWFLICM is unstable and time-consuming. FLICM and its variants usually achieve good clustering results, but it has been proved that the mathematical deduction is not very rigorous in their frameworks. In [17] and [18], it indicates that the iterative updating formula in FLICM could not minimize the objective function actually.

Zhu *et al.* [19] imposed constraints on the membership function and proposed the GIFP_FCM algorithm. GIFP_FCM performs well for segmenting textures in noisy images. Zhao proposed a kernel variant of GIFP_FCM in [20] (KGFCM), which accelerates the convergence speed for images corrupted by noise. Recently, by integrating prior knowledge of cluster centers in the objective function, Deng *et al.* [21] proposed ETFCM to guide the clustering centers into the optimal positions. All of the aforementioned methods use the Euclidean distance or kernel induced distance as the similarity measure. Those methods perform better if the accurate clustering centers could be attained. However, they may not be appropriate for the hyperspherically distributed data clustering, especially for data with obscure centers.

Density-based approaches are also very popular for clustering [22]–[24]. Ester *et al.* [22] proposed an algorithm which is the density-based spatial clustering of applications with noise (DBSCAN). DBSCAN is effective in discovering clusters of an arbitrary shape. Neighborhood radius and the minimum cluster size have crucial impacts on the performance of DBSCAN [23]. Rodriguez and Laio [24] proposed the density peak (DP) algorithm. This algorithm is based on the idea that cluster centers have a higher density than their neighbors and have a relatively large distance from points with higher densities. The DP algorithm could suppress the effect of outliers automatically [24]. However, the DP algorithm relies on manual selection of cluster centers from the decision graph [24].

To compensate the applicability to manifold-structured nonconvex clusters, Ding *et al.* [25] revisited the formulation of FCM and brought forward a new similarity criterion. This similarity criterion is based on an average weighted similarity from a point of a certain cluster to any member of the same cluster. This method makes FCM free of cluster centers, which is called center-free FCM (CFFCM). In CFFCM, the assignment of a pixel depends on all of the pixels. CFFCM makes it possible to cluster sophisticated data. However, CFFCM is sensitive to noise and initialization [25].

In this paper, we design a novel similarity measure, which makes our method more suitable for nonlinearly separable data clustering. The calculation of point-to-point similarity in our similarity measure considers the local and global spatial information, which is more effective than CFFCM for clustering nonlinear data. A little few of this paper has been presented in [26].

In addition, we ameliorate the method from another two aspects and propose a novel method called similarity measure-based PFCM with local label information. First, we take the interclass similarity and intraclass similarity into consideration, allowing that a specific cluster has not only self-similarity (or called confidence degree) but also some similarities with other clusters. This improvement, taking advantage of more correlations between clusters, would make the proposed method more robust to noise. Second, we utilize the local label information in a novel style. This benefits the image detail preservation and noise suppression. Apart from that, by defining a novel similarity measure and neighbor factor, the proposed method could also mitigate the cluster-size sensitivity problem.

Moreover, we compare the proposed method with another nine existing clustering methods from both the qualitative and quantitative aspects. Experimental results show that the proposed method is effective and performs well for the segmentation of the MRI brain image, especially for the clinical MRI brain image.

The remainder of this paper is organized as follows. Section II briefly describes the fundamental principle of PFCM. Details of the proposed method are described in Section III. Various results of experiments conducted on images including artificial images, simulated images, synthetic, and clinical images are presented in Section IV. Concluding remarks are drawn in Section V.

II. Background Works

A. FCM Method

The FCM method was first introduced by Dunn [11] and later was extended by Bezdek [27]. FCM deals with the problem of clustering n measured data points or objects into c clusters through an iterative minimization process. The objective function of FCM is defined as [27]

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \|x_k - v_i\|^2 = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m d_{ik}^2 \tag{1}$$

where n is the number of pixels, c is the number of clusters, m is a fuzzy factor and is usually set as 2, x_k is the kth measured data point or object, and v_i is the ith cluster prototype. Minimizing the objective function J_m can be obtained through an iterative process with the following updated equations:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \tag{2}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m} \tag{3}$$

with the constraint $\sum_{i=1}^{c} u_{ik} = 1$.

B. PFCM Method

Clustering methods that only considered the membership value usually lead to inadequate interpretation of the data [13], [28]. Those methods perform worse when clustering data contain outliers [13].

To circumvent the aforementioned circumstances, Pal *et al.* [13] proposed the PFCM model. PFCM could avoid coincident clusters because it considers typicality values in clustering. PFCM can be regarded as the hybridization of FCM and PCM [29] by introducing two parameters a and b. a and b could be considered as the relative importance measuring factors for FCM and PCM, respectively. The objective function of PFCM is defined as [13]

$$J_{m,\eta} = \sum_{i=1}^{c}\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)d_{ik}^{2} + \sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}(1-t_{ik})^{\eta} \tag{4}$$

where n is the number of pixels. c is the number of clusters. u_{ik} and t_{ik} are membership and typicality, respectively. u_{ik} and t_{ik} is subject to the constraints $\sum_{i=1}^{c} u_{ik} = 1 \forall k$, and $0 \le u_{ik}, t_{ik} \le 1$. Here, the constants a and b are non-negative. In (4), η is the same as m, is a fuzzy factor, and is usually set as 2. γ_i is the user-defined constant.

By applying the Lagrange multiplier theorem to (4), after solving the optimization problem, the calculation of the membership partition matrix, typicality matrix, and the cluster centers are obtained as follows [13]:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \tag{5}$$

$$t_{ik} = \frac{1}{\left(1+\frac{b}{\gamma_i}d_{ik}^{2}\right)^{\frac{1}{\eta-1}}} \tag{6}$$

$$v_i = \frac{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)x_k}{\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)} \tag{7}$$

where $1 \le i \le c$, $1 \le k \le n$. γ_i is recommended to be defined as follows:

$$\gamma_i = K\frac{\sum_{k=1}^{n} u_{ik}^{m} d_{ik}^{2}}{\sum_{k=1}^{n} u_{ik}^{m}}. \tag{8}$$

Here, K is chosen to be 1 [13].

PFCM ameliorates problems suffered by FCM and PCM. For instance, PFCM is less sensitive to outliers than FCM, because PFCM assigns a relatively small typicality value to outliers in clustering [13]. Since the memberships generated by FCM are relative values, they may not be suitable to represent "typicality" [30], [31]. The usage of t_{ik} in PFCM could be interpreted as one noise class per good cluster. And the typicality in the noise cluster is the complement of the data class [31]. In brief, PFCM performs well for images with outliers. However, just like FCM, PFCM is not suitable for the nonlinearly separable data due to the use of Euclidean distance.

In PFCM, the use Euclidean distance, specifically the distance between data points and cluster centers, limits the performance of PFCM for nonlinearly separable data. This is because Euclidean distance emphasizes spherical clusters, which tends to partition data points not far from each other to the same cluster. For complicated nonspherical data, PFCM usually could not acquire a satisfactory result. If we design a more appropriate similarity measure, PFCM may be more effective for clustering the complex data like MRI brain images.

III. Proposed Method

In this section, we describe our method to improve PFCM exhaustively and analyze the advantage of the proposed method for MRI brain image segmentation. The improvements of the proposed method include two parts: 1) considering the intrinsic relationship between clusters represented by the intraclass similarity and interclass similarity and 2) using the local label information in a novel way to replenish the local and nonlocal spatial constraints.

A. Introducing the Similarity Measure

In this paper, the proposed similarity measure is to compute the absolute similarity between the data point and cluster. The absolute similarity consists of direct similarity and indirect similarity. The direct similarity is associated with the intraclass similarity while the indirect similarity depends on the interclass similarity.

Before defining the final form of the similarity measure, we first define W_{lk} to compute the similarity between point l and k as follows:

$$W_{lk} = \exp\left(-\left(\frac{d_{lk}}{H}\right)^{2} - \frac{\mathrm{dis}_{lk}}{\mathrm{Diag}}\right) \tag{9}$$

where H is a bandwidth derived from [32]. The value of H reflects the distribution of the data. The specific calculations and definitions of H can refer to [32]. d_{lk} and dis_{lk} represent the intensity Euclidean distance and coordinate Euclidean distance between pixel l and pixel k, respectively. Diag is interpreted as the diagonal length of an image. If the proposed method is applied for point clustering, d_{lk} and dis_{lk} are both the Euclidean distance of coordinates between point l and point k. In this case, Diag is the farthest distance of two points among all points. The second term $\exp(-\mathrm{dis}_{lk}/\mathrm{Diag})$ is intended to consider the spatial information. Consistent with the common sense, the farther the distance between two pixels is, the less the similarity between them is.

The calculations of the intraclass similarity and interclass similarity depend on the relative similarity between the data point and cluster. The similarity between data point and cluster is useful for image clustering [25]. In this paper, following the characteristics of PFCM, we use W_{lk} to define the relative similarity between point k and class i, denoted by ρ_{ik}, as follows:

$$\rho_{ik} = \frac{\sum_{l=1}^{n}\left(au_{il}^{m} + bt_{il}^{\eta}\right)W_{lk}}{\sum_{l=1}^{n}\left(au_{il}^{m} + bt_{il}^{\eta}\right)}, \quad \begin{matrix} l \in \Omega_i \\ d_{lk} < w \end{matrix} \tag{10}$$

where W_{lk} is calculated by (9). a and b are the same factors as PFCM. u_{il} and t_{il} are membership and typicality, respectively, and ρ_{ik} could be regarded as a weighted average of W_{lk}, which is associated with both the membership and typicality. In our experiment, a and b are constrained to $a+b=1$ to eliminate one parameter. Ω_i represents the set of pixels belonging to the ith cluster. Here, d_{lk} is the distance between point l and point k, and w is a threshold. That is, to say, only the points

belonging to the ith cluster and not far away from point k are used to compute ρ_{ik}. This strategy makes ρ_{ik} more representative for the meaning of similarity, since the pixel clustering mainly depends on the pixels of the considering cluster, rather than all of the pixels. For the MRI brain image, when we compute the membership of a pixel belonging to a certain tissue, the pixels in this tissue should play the most important roles. Moreover, this strategy reduces the computational complexity.

We proceed to discuss the interclass similarity $S_{ij}(i \neq j)$ and the intraclass similarity $S_{ii}(i = j)$. The interclass similarity values are decided by samples from different classes, while the calculation of every intraclass similarity depends only on samples from the same class. In this paper, we use the relatively similarity ρ_{ij} to define S_{ij} in a unified form as follows:

$$S_{ij} = \frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk}}{\sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}}}, k \in \left\{ \arg\max_{k} \{\rho_{ik} \text{ or } \rho_{jk}\} \right\}. \tag{11}$$

Here, only those points which maximize either ρ_{ik} or $\rho_{jk}(k \in \{\arg_k \max\{\rho_{ik} \text{ or } \rho_{jk}\}\})$ are used to compute S_{ij}, because those points are more representative for the ith class or jth class. Dealing with k in such restriction, $S_{ij}(i \neq j)$ makes more sense to represent the interclass similarity. Meanwhile, if $i = j$, the intraclass similarity S_{ii} depends only on those points belonging to the ith class.

Then we will illustrate the reason for defining S_{ij} as such a formula.

Proposition 1: $S_{ij} \in (0, 1), \forall i, j$.

Proof: From (9), it can be seen that W_{lk} is an exponential function whose exponential part is a negative value. According to the property of the exponential function, we can easily obtain that $W_{lk} \in (0, 1)$. Equation (10) indicates that ρ_{ik} is a form of the weighted average of W_{lk}. It can be easily deduced that ρ_{ik} cannot exceed the maximum value of $W_{lk}(\rho_{ik} \leq \max(W_{lk}),\ \forall k, \forall l \in \Omega_i\ )$. So $\rho_{ik} \in (0, 1) \forall i, \forall k$.

Utilizing the Cauchy inequality

$$\begin{aligned} &\sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}} \geq \sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}} \\ &\qquad \text{if } \rho_{ik}, \rho_{jk} \in (0, 1) \text{ then } \sqrt{\rho_{ik}\rho_{jk}} > \rho_{ik}\rho_{jk} \\ &\Rightarrow \sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}} \geq \sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}} > \sum_{k=1}^{n} \rho_{ik}\rho_{jk} \\ &\qquad \Rightarrow S_{ij} = \frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk}}{\sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}}} \in (0, 1). \end{aligned} \tag{12}$$

Therefore, S_{ij} is between 0 and 1. S_{ij} could represent the similarity between the ith class and the jth class. ■

Proposition 2: S_{ij} has the maximum likelihood to gain the maximum value if ρ_{ik} and ρ_{jk} are approximately equal and close to 1.

Proof: Using the Cauchy inequality for S_{ij}, we obtain

$$\begin{aligned} S_{ij} &= \frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk}}{\sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}}} \\ &= \frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk} \cdot \sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}}}{\sqrt{\sum_{k=1}^{n} \rho_{ik} \sum_{k=1}^{n} \rho_{jk}} \cdot \sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}}} \\ &\leq \frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk}}{\sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}}} \left(\text{equal when } \frac{\rho_{ik}}{\rho_{jk}} = \text{const} \right). \end{aligned} \tag{13}$$

Then, using the inequality of arithmetic and geometric means, we obtain the following inequality:

$$\frac{\sum_{k=1}^{n} \rho_{ik}\rho_{jk}}{\sum_{k=1}^{n} \sqrt{\rho_{ik}\rho_{jk}}} \in \left(\overline{\sqrt{\rho_{ik}\rho_{jk}}}, 1 \right). \tag{14}$$

In (14), $\overline{\sqrt{\rho_{ik}\rho_{jk}}}$ is the average value of $\sqrt{\rho_{ik}\rho_{jk}}$. According to the equal condition of inequalities (13) and (14), if ρ_{ik} and ρ_{jk} are approximately equal and close to 1, S_{ij} has the maximum likelihood to gain the maximum value. ■

We consider different situations to analyze whether the conditions could be satisfied as follows.

1) For different clusters ($i \neq j$) and a certain point k, the relative similarity ρ_{ik} and ρ_{jk} cannot be a large value near 1 simultaneously, because a certain point is impossible, belonging to two different classes at the same time. So, the above conditions cannot occur. This means the interclass similarity ($S_{ij}(i \neq j)$) between clusters will not be a large value.
2) For a particular cluster ($i = j$), ρ_{ik} and ρ_{jk} are equal. According to inequalities (13) and (14), if most $\rho_{ik}(k \in \arg\max\{\rho_{ik} \text{ or } \rho_{jk}\})$ are closer to 1, the intraclass similarity (S_{ii}) of the ith class is closer to 1. It indicates that a cluster has a higher intraclass similarity value if points assigned to the cluster more certainly.
3) If the relative similarities of points to the ith class are small (ρ_{ik} for most k), then the interclass similarity ($S_{ij}(i \neq j)$) is relatively small. This is because ρ_{ik} and ρ_{jk} have a relatively large difference under such circumstances. However, the intraclass similarity (S_{ii}) primarily depends on the kth points constrained to $k \in \arg\max\{\rho_{ik}\}$. In this situation, the small-size cluster is less affected by the large-size clusters, because they compute the intraclass similarity separately. That is to say, the proposed similarity measure is less sensitive to unbalanced cluster sizes than some other FCM-based methods [7], [33].

We can conclude from the aforementioned analysis that every particular class tends to have a large intraclass similarity value and a relatively small interclass similarity value. The intraclass similarity could be interpreted as the confidence degree of a class. It seems that every class trusts itself more than others. However, because points in the classes are vague, there should be some connectivity between classes. The core idea of the proposed similarity measure could be interpreted as the fuzzification of the clustering classes, while FCM methods are based on the fuzzification of the clustering points.

Then, the absolute similarity between point k and class i can be calculated, which is represented by τ_{ik}. τ_{ik} is defined

as follows:

$$\tau_{ik} = \sum_{j=1}^{c} \rho_{jk} S_{ij} = \rho_{ik} S_{ii} + \sum_{j=1, j\neq i}^{c} \rho_{jk} S_{ij} \tag{15}$$

where τ_{ik} can be seen as the sum of $\rho_{ik}S_{ii}$ and $\sum_{j=1,j\neq i}^{c} \rho_{jk}S_{ij}$. Here, $\rho_{ik}S_{ii}$ and $\sum_{j=1,j\neq i}^{c} \rho_{jk}S_{ij}$ represent the direct and indirect similarity between point k and class i, respectively. In (15), ρ_{ik} is the relative similarity between point k and class i. $S_{ij}(i \neq j)$ is the criteria to measure the interclass similarity. If j is equal to i, S_{ii} could be interpreted as the intraclass similarity (or the confidence degree).

The proposed similarity measure utilizes the spatial information while computing the similarity between two points or pixels (W_{lk}). However, in (9), the second item $\exp(-\text{dis}_{lk}/\text{Diag})$ indicates that the immediate neighbor pixels of a central pixel and the pixels not far away from it have nearly the same influence to the central pixel. This suppresses the importance of the immediate neighbor to some extent. For example, if the image has noises, the immediate neighbor of a central pixel should play a more important role to suppress noises. So in the next section, we advance the method with local label constraints.

B. Introducing the Local Label Information

MRI images contain noises of different levels. It is necessary to utilize the neighbor information for improving the noise tolerance and resistance properties.

Assume that the kth pixel is at the center of a local window (for example, 3×3). By adding a related neighbor factor in the objective function, the neighbor pixels in the local window (N_k) would affect the ultimate attribution of the kth pixel.

We assume that the introduction of local label information should have three characteristics as follows.

1) It should reflect the damping extent of the neighbors with the spatial distance from the central pixel.
2) It should control the tradeoff between image detail retention and image noise resistance.
3) It should reflect that the possibility of the attribution of central pixel is positively correlated with the amount of the neighbor pixels assigned to the same class.

We introduce a new factor f_{ik} that satisfies the aforementioned three characteristics. Before defining f_{ik}, we first count the labels of the neighbor pixels which are around the central pixel k.

To meet the first characteristic mentioned above, we define the neighbor label Z_{ir} as follows:

$$Z_{ir}(r \in N_k) = \begin{cases} \frac{1}{d_{rk}} & \text{if } i = \arg_i \max(u_{ir}) \\ 0 & \text{else} \end{cases} \tag{16}$$

where d_{rk} is the Euclidean distance between the neighbor pixel r and the central pixel k. For every certain ith class, it has a label for each pixel around pixel k. The label is equal to the reciprocal of d_{rk} when u_{ir} achieves the maximum; otherwise, Z_{ir} is zero. For a simple example, assuming there are only two classes, the calculation of Z_{ir} is shown in Fig. 1.

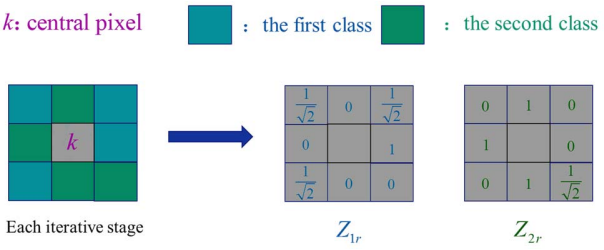


Fig. 1. Calculation of Z_{ir}.

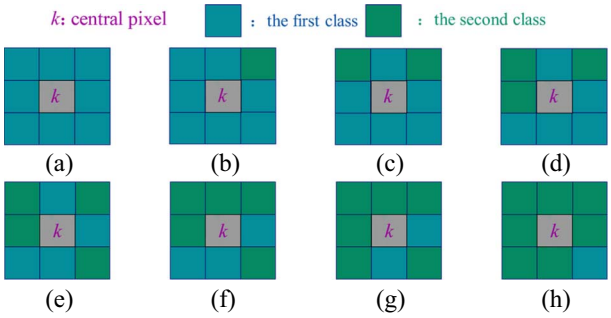


Fig. 2. Computation of f_{ik} for a central pixel k with different neighbor conditions.

Fig. 1 reveals that the central pixel k is actually affected by the labels of the neighbor pixels with distance constraint because of the defuzzification of the u_{ir}.

Then, we calculate the sum of Z_{ir} for a specific i and the r located in a local window around k as follows:

$$Y_{ik} = \sum_{r \in N_k} Z_{ir}. \tag{17}$$

Here, Y_{ik} might translate into the total number of pixels assigned to the ith class. In this definition, Y_{ik} considers the distinct influence of the spatial distance. Then, we normalize the amount of labels of the neighbor pixels belonging to each class, as follows:

$$P_{ik} = \frac{Y_{ik}}{\sum_{i=1}^{c} Y_{ik}}. \tag{18}$$

In (18), P_{ik} is the proportion of the labels around pixel k in a window.

To define the new factor f_{ik}, we choose a parameter β to balance image noise insensitivity and image detail preservation. The formula of f_{ik} is presented as follows:

$$f_{ik} = 1 + \beta P_{ik}. \tag{19}$$

Then we discuss the impact of f_{ik} on u_{ik} as follows.

Property 1: The membership u_{ik} should be positively related to neighbor factor f_{ik}.

Analysis: For simplicity, assume that there are only two classes in the clustering data. For an extreme example, the central pixel is corrupted by noise, while the neighbor pixels are homogenous, and belong to the first class as shown in Fig. 2(a). In this situation, f_{1k} will get the maximum value $1+\beta$, while f_{2k} is equal to 1. So f_{1k} should push the membership value u_{1k} to increase in order to enhance the anti-noise ability. As the possibility of the neighbor pixels belonging to the second class increases, the membership of the noise pixel k belonging to the second class increases, alone with the increase

of f_{2k} and decreasing of f_{1k} as shown in Fig. 2(b)–(h). So the membership u_{ik} should be positively related to the neighbor factor f_{ik}, which could help determine the distribution of noise pixels from the perspective of probability. For multiclass clustering, a similar analysis can be made. As the possibility of the amount of pixels around k belonging to the ith class increases, the possibility of pixel k belonging to class i increases.

We define f_{ik} to preserve more image detail as well as to suppress image noise. However, the prerequisite is that the membership u_{ik} should have a positive correlation with f_{ik}. How we utilize the neighbor information f_{ik} in the proposed method will be elaborated in the following section.

C. Final Form of the Improved PFCM of Our Method

The new similarity measure adopting the intraclass similarity and interclass similarity suppresses the sensitivity to unbalanced cluster sizes. Meanwhile, if we consider the local label information, the proposed method should be more insensitive to noise. In this part, we utilize both the similarity measure and the local label information to improve the performance of PFCM. We multiply the similarity measure τ_{ik} and the neighbor factor f_{ik}, then substitute the d_{ik} in (4). The final objective function of the improved PFCM is obtained as follows:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)\frac{1}{f_{ik}^{2}}\cdot\frac{1}{\tau_{ik}^{2}} + \sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}(1-t_{ik})^{\eta}. \tag{20}$$

Here, τ_{ik} is computed by (15), f_{ik} is computed by (19), and u_{ik} is under the constraint condition

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k. \tag{21}$$

A Lagrange objective function can be constructed as

$$L(u,t) = \sum_{i=1}^{c}\sum_{k=1}^{n}\left(au_{ik}^{m} + bt_{ik}^{\eta}\right)\frac{1}{f_{ik}^{2}}\cdot\frac{1}{\tau_{ik}^{2}} + \sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}(1-t_{ik})^{\eta} - \sum_{k=1}^{n}\lambda_k\left(\sum_{i=1}^{c}u_{ik} - 1\right). \tag{22}$$

To minimize the Lagrange objective function, we take the derivative of L with respect to u_{ik}, t_{ik}, and λ_k, respectively. The following calculations are used:

$$\frac{\partial L}{\partial u_{ik}} = \mathrm{amu}_{ik}^{m-1}\frac{1}{f_{ik}^{2}}\cdot\frac{1}{\tau_{ik}^{2}} - \lambda_k \tag{23}$$

$$\frac{\partial L}{\partial t_{ik}} = b\eta t_{ik}^{\eta-1}\cdot\frac{1}{f_{ik}^{2}\tau_{ik}^{2}} - \gamma_i\eta(1-t_{ik})^{\eta-1} \tag{24}$$

$$\frac{\partial L}{\partial \lambda_k} = 1 - \sum_{i=1}^{c} u_{ik}. \tag{25}$$

Letting $(\partial L/\partial u_{ik}) = 0$, $(\partial L/\partial t_{ik}) = 0$, and $(\partial L/\partial \lambda_k) = 0$, we obtain the iterative formulas of u_{ik} and t_{ik}

$$u_{ik} = \frac{(f_{ik}\tau_{ik})^{\frac{2}{m-1}}}{\sum_{i=1}^{c}(f_{ik}\tau_{ik})^{\frac{2}{m-1}}} \tag{26}$$

$$t_{ik} = \frac{1}{A1 + \left(\frac{b}{\gamma_i(f_{ik}^{2}\tau_{ik}^{2})}\right)^{\frac{1}{\eta-1}}}. \tag{27}$$

According to (26), we know that u_{ik} is the normalized ratio of $f_{ik}\tau_{ik}$, so u_{ik} is positively related to f_{ik}. This meets the fundamental requirement for f_{ik} as described in Section III-B.

Through utilizing u_{ik} and t_{ik}, the clustering of data and segmentation of the MRI brain image could be achieved. The whole procedure of the proposed method is summarized as follows.

Step 1: Set the number of cluster c the fuzzification parameter m and η the window size N_k; the parameters a, b; and the stopping condition ε. For all experiments in this paper, we set $a = 0.65$, $b = 1 - a$.

Step 2: Initialize the membership matrix U using the results of FCM and initialize the typicality matrix T randomly. Compute the absolute similarity τ_{ik} using (9)–(11) and (15) and compute f_{ik} according to (16)–(19). Let $t = 1$.

Step 3: Update the membership U and typicality T according to (26) and (27), respectively.

Step 4: Update τ_{ik} and f_{ik}, respectively, according to (15) and (19).

Step 5: If $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$ then stop; otherwise, $t = t + 1$ and go to step 3.

D. Parameter Discussion

There are three parameters in the proposed method. They are a, w, and β, respectively. Among them, a controls the relative importance of u_{ik} and t_{ik}. The parameter β is established to balance the noises and details. The threshold w is calculated as follows [34]:

$$w = \delta\overline{d}. \tag{28}$$

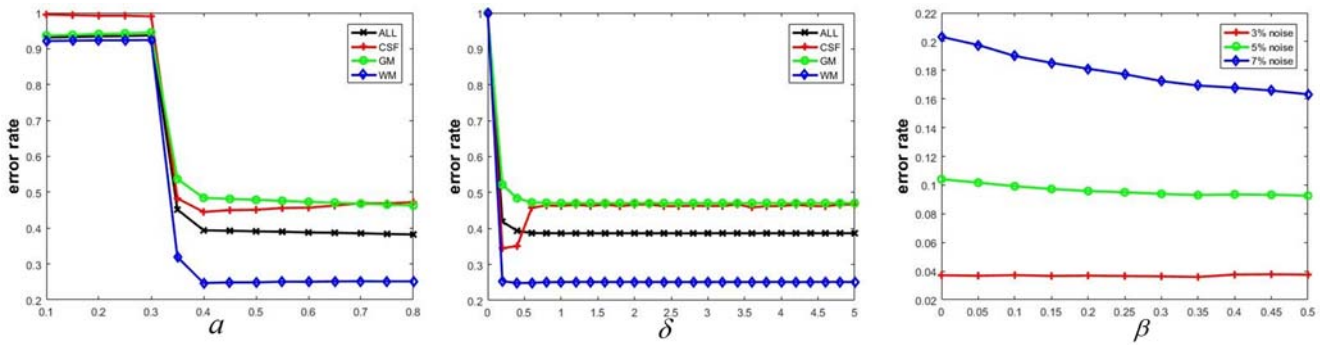
Here, $\overline{d}$ is the mean deviation of the clustering data. For dataset $X = \{x_1, x_2 \ldots x_n\}$, it can be calculated as

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n}x_i\right). \tag{29}$$

Then we will analyze how to select a, δ, and β approximately. We conducted an experiment on MRI images from the Internet brain segmentation repository (IBSR) [35] and set the value of a from 0.1 to 0.8 with the step of 0.05. The error rate is defined as the ratio of the number of misclassified pixels of a tissue (CSF, WM, and GM) to the total numbers of pixels of the same tissue. Fig. 3(a) depicts the varying trend of the error rate of CSF, GM, WM and the total error rate of these three tissues (ALL) using different a. Fig. 3(a) indicates that the error rate of all tissues starts to level off in the position of 0.4. To acquire overall good results, we set $a = 0.65$ in all of our experiments.

Similarly, we set the value of δ from 0 to 5 with the step of 0.2, and the result is shown as Fig. 3(b). We can see from Fig. 3(b) that the error rate tends to be stable at the point of $\delta = 0.5$. We use $\delta = 3$ in our experiments for the segmentation of the MRI brain image.

The parameter β is set to balance noise insensitivity and image detail preservation. We perform the parameter experiment on different levels of noisy MRI images. The noise



(a) (b) (c)

Fig. 3. Varying trend of the error rate under different values of (a) a, (b) δ, and (c) β. The red, green, and blue lines represent the total error rate of MRI of noise level 1%, 3%, and 5%, respectively.

levels are from 3% to 7%, and we set the value of β from 0 to 0.5 with a step length of 0.05. From Fig. 3(c), we can find that in a high noise level, the error rate decreases as β increases. The higher the noise level, the more effective the β behaves. In a lower noise level, the error rate of the clustering result is almost constant. We use a compromise value of β to adapt to different levels of noise. β is set as 0.2 in our experiments.

In Fig. 3, all of the curves gradually level off. We can infer from Fig. 3 that the proposed method is not sensitive to parameters within a certain range.

IV. Experiments and Results

In this section, various experiments are carried out to verify the performance of the proposed method on both linearly and nonlinearly separable data. We compare the proposed method with nine methods, including: FCM_S [12], PFCM [13], IFCM [15], FLICM [8], KWFLICM [16], CFFCM [25], GIFP_FCM [19], KGFCM [20], and ETFCM [21] on the simulated images, artificial images, synthetic, and clinical MRI brain images. The synthetic MRI brain images generated from the BrainWeb [36] and the clinical MRI brain images originated from IBSR [35].

A. Segmentation of Artificial Images

In this section, the artificial images are mainly used to compare the validity of the proposed method with the existing methods on the complicated nonspherical data. Many methods consider the constraint of the immediate neighbor information of a central pixel to resist noise. Unlike that for point clustering, the clustering depends on the distance between discrete coordinates. This will generate additional points that do not require clustering, and may lead to a mistaken result. In this experiment, we compare the proposed method with the above six methods for nonspherical data clustering. The immediate neighbors are not considered in this situation. Then FLICM [8] and FCM_S [12] perform like FCM [27], and KWFLICM [16] degenerates into KFCM [37]. The DP [24] algorithm is also very popular for point clustering, which can recognize clusters regardless of their shape [24]. We also compare the proposed method with the DP algorithm for various manifold-structured point clustering.

Fig. 4 shows the results of those methods. Fig. 4 displays the clustering results of different geometric data. It is obvious that the data points in Fig. 4(a)–(c) are nonlinearly separable. In Fig. 4(d), the data points are linearly separable. However, it is confronted with a problem that the cluster sizes are unbalanced. This unbalanced problem is especially serious in clinical MRI brain images, in which the CSF is much less than GM and WM. So the clustering results of Fig. 4(d) have a certain guiding significance for MRI brain image segmentation on the cluster-size sensitivity problem.

From Fig. 4(a)–(c), we can see that the methods taking the Euclidean norm as similarity measure, such as FCM, PFCM, IFCM, GIFP_FCM, and ETFCM, tend to divide the data points into similar classes linearly. And those methods fail to classify the nonlinearly separable data. KFCM and KGFCM map the data into higher dimensional feature space to actualize nonlinear separation. However, they do not perform very well on nonlinearly separable data. KGFCM even misses one class, as shown in Fig. 4(c). The CFFCM method is proposed for the nonlinearly separable data clustering. However, it is sensitive to noise and it fails to separate data with complex distributions, as shown in Fig. 4(a)–(c). In Fig. 4(d), the size of class located in the middle of the field is nearly a quarter of the size of the other two classes, and the clustering prototypes are not far from each other. In such a case, KFCM, FCM, PFCM, IFCM, and KGFCM are sensitive to cluster sizes, and those methods misclassify a few points. CFFCM can reduce the amount of mistaken points. However, it still has some misclassifications. GIFP_FCM and ETFCM misclassify a few points for the "cloud noise" data as shown in Fig. 4(d). But they could not separate the nonlinear data like Fig. 4(a)–(c). From the results shown in Fig. 4, we can conclude that the proposed method outperforms the comparison methods for nonlinearly separable or unbalanced data. As we all know, MRI brain image is nonlinearly separable and the cluster sizes are unbalanced. Therefore, this preliminary experiment explains that our method may suit for the segmentation of the MRI brain image.

Fig. 5 shows the results of DP algorithm. We choose the cluster centers from the decision graph manually and acquire

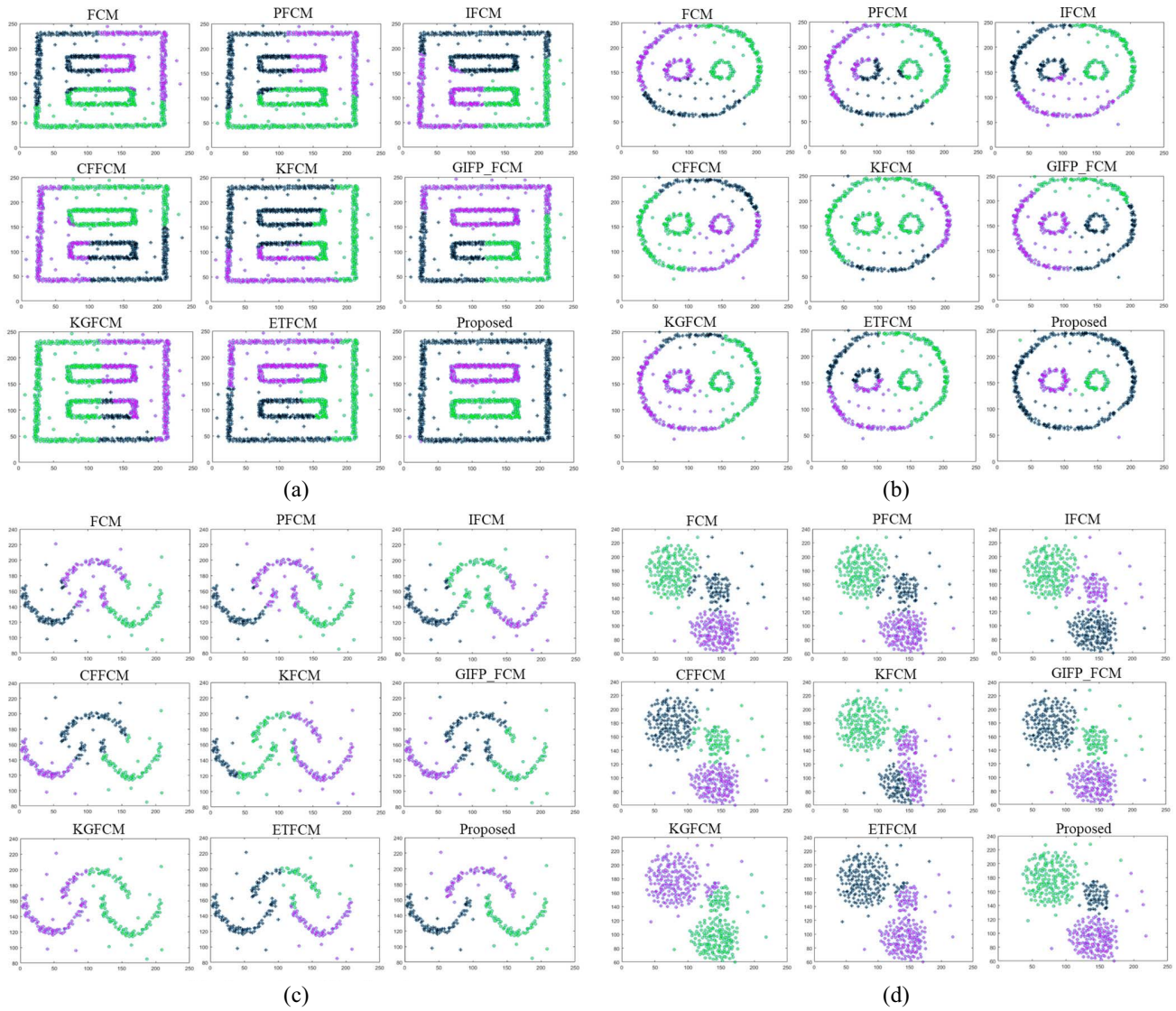


Fig. 4. Comparison results of FCM, PFCM, IFCM, CFFCM, KFCM, GIFP_FCM, KGFCM, ETFCM, and the proposed method. Clustering of (a) rectangle noise, (b) circular noise, (c) arch noise, and (d) cloud noise.

the clustering results. Fig. 5(d) shows that the DP algorithm performs well for spherical distributed data, which could identify outliers and noises to some extent, as shown in Fig. 5(d). However, if the data points have more complicate distributions, the cluster centers chosen from decision graphs could not achieve good partitions as shown in Fig. 5(a)–(c). From Figs. 4 and 5, we could see that the proposed method is more effective than DP algorithm for nonlinearly separable point clustering.

B. Segmentation of Simulated Images

The main challenges of clustering MRI images are the existing ones of intensity inhomogeneity and unbalanced cluster sizes. Especially in the MRI brain image, there are abundant image details with the coexistence of image noises, for the segmentation of the MRI brain image. In this section, two simulated images are generated to simply simulate the intensity inhomogeneity in MRI images and the unbalanced cluster size problem. The 3-D graphical representations are displayed in Fig. 6. The clustering results are shown in Figs. 7 and 8, respectively.

Fig. 7 shows that PFCM, IFCM, and ETFCM are sensitive to intensity inhomogeneity and noise. They fail to classify the pixels correctly. FLICM and FCM_S consider the neighbor information and are not sensitive to noises. However, they lose some image details, or even misclassify more pixels, in the presence of intensity inhomogeneity. The results are shown in Fig. 7(c) and (g). KWFLICM, CFFCM, GIFP_FCM, KGFCM, and the proposed method perform well for clustering the first image. However, CFFCM is sensitive to noise and still has some mistaken pixels, as shown in Fig. 7(f). KWFLICM is time-consuming and loses some image details in edges and corners, as shown in Fig. 7(h). GIFP_FCM performs well. Only a few noisy points are misclassified, as shown in Fig. 7(i). KGFCM also performs well. It only loses a few image details

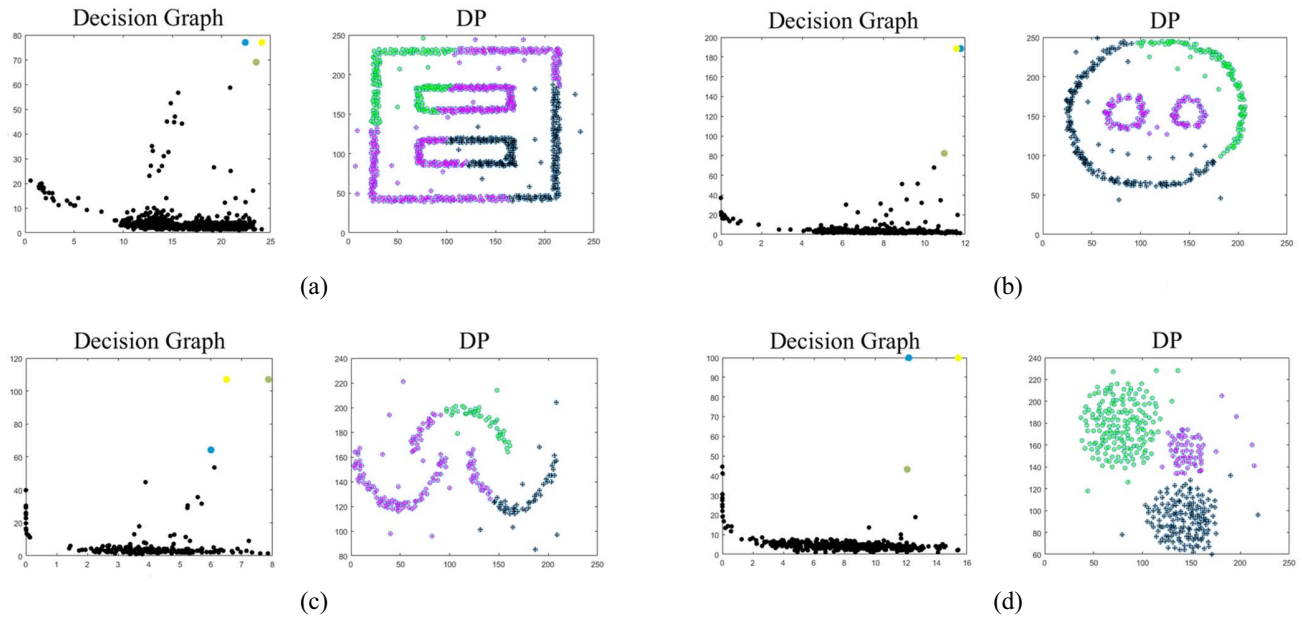


Fig. 5. Decision graphs (left), results of DP algorithm (right). Clustering of (a) rectangle noise, (b) circular noise, (c) arch noise, and (d) cloud noise.

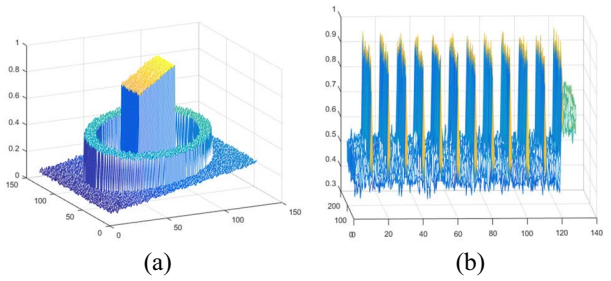


Fig. 6. 3-D display of the two images. (a) First image has a simple intensity inhomogeneity and noise. (b) Second image has unbalanced cluster sizes and noise.

as shown in Fig. 7(j). The proposed method performs the best, as shown in Fig. 7(l).

The second image is simulated to verify the effectiveness of the proposed method and the comparison method for dealing with images with unbalanced cluster sizes. Moreover, the single pixel straight lines in Fig. 6(b) are the image details and there are Gaussian noise in the image. This is a tough task for clustering. PFCM, IFCM, and CFFCM is sensitive to noise and the results are shown as Fig. 8(b), (e), and (f), respectively. FCM_S and KGFCM make the edges and corners of image rougher. As shown in Fig. 8(c) and (j), the single pixel line is widened after clustering by FCM_S or KGFCM. FLICM and KWFLICM are effective for noise suppression. But for image details like the second image, FLICM and KWFLICM may produce unexpected error. Image details like this also exist in the MRI brain image, which may lead to unsatisfactory results when using FLICM and KWFLICM. ETFCM performs well, but it is still sensitive to noise as shown in Fig. 8(k). GIFP_FCM achieves nearly an accurate segmentation as shown in Fig. 8(i). As described in [19], GIFP_FCM performs well for texture images with noise. The proposed method utilizes the local label information in a novel way, and

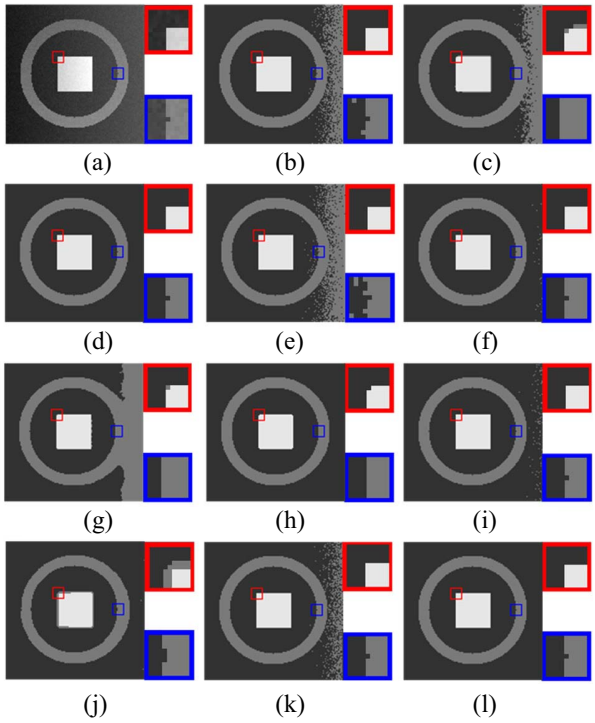


Fig. 7. Comparison results of different methods to the first simulated image which has a simple intensity inhomogeneity and Gaussian noise. (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

preserves image details as well as suppress noise as shown in Fig. 8(l).

C. Segmentation of Synthetic MRI Brain Image

In order to show the robustness of the proposed method, the proposed method and the comparison methods are performed

TABLE I
DC VALUES FOR THE SEGMENTATION RESULTS ON SYNTHETIC MRI BRAIN IMAGES

		PFCM	IFCM	FCM_S	FLICM	KWFLICM	CFFCM	GIFP_FCM	KGFCM	ETFCM	Proposed
N3 F20	CSF	0.8860	0.9494	0.9163	0.8897	0.8788	0.9444	0.9534	0.8894	0.9532	**0.9571**
	GM	0.9143	0.9468	0.9368	0.9284	0.9303	0.9472	0.9490	0.9263	0.9505	**0.9511**
	WM	0.9359	0.9412	0.9389	0.9347	0.9414	0.9480	0.9441	0.9331	0.9464	**0.9482**
N3 F40	CSF	0.8416	0.9379	0.9046	0.8803	0.8752	0.9381	0.9439	0.8788	0.9433	**0.9504**
	GM	0.8723	0.9290	0.9170	0.9079	0.9078	0.9301	0.9327	0.9062	0.9328	**0.9363**
	WM	0.9128	0.9241	0.9190	0.9123	0.9224	0.9295	0.9287	0.9127	0.9284	**0.9326**
N5 F40	CSF	0.7885	0.9093	0.8944	0.8765	0.8745	0.9214	0.9210	0.8749	0.9191	**0.9328**
	GM	0.8080	0.8973	0.9059	0.9047	0.9087	0.9027	0.9040	0.9018	0.9038	**0.9162**
	WM	0.8680	0.8895	0.9066	0.9087	0.9101	0.8951	0.8952	0.9081	0.8949	**0.9102**

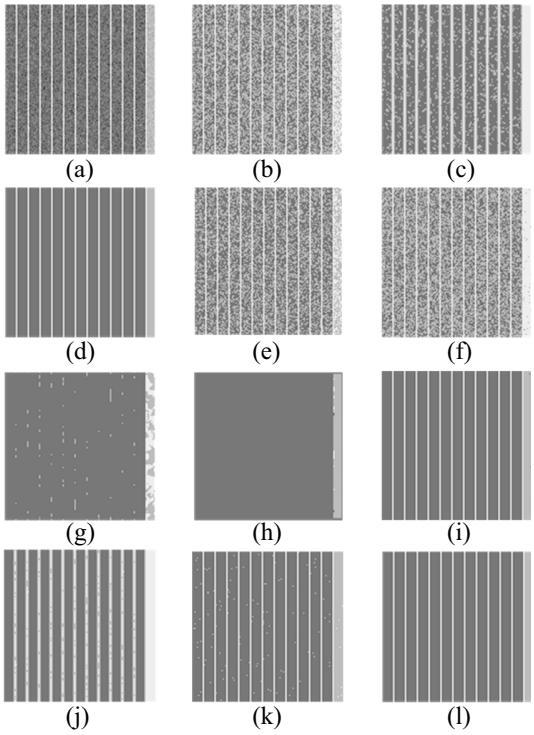


Fig. 8. Comparison results of different methods to the second image which has unbalanced cluster sizes and Gaussian noise. (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

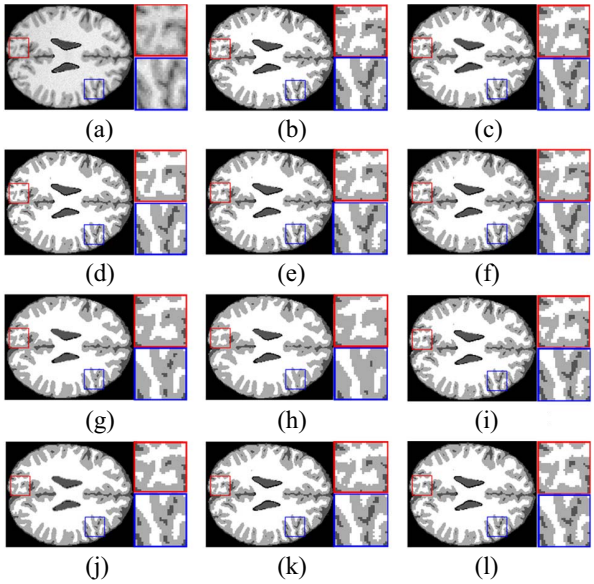


Fig. 9. Comparison segmentation results of synthetic MRI brain images with noise level 3% and intensity inhomogeneity level 20% (N3F20). (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

on synthetic MRI brain images obtained from BrainWeb [36]. BrainWeb provides synthetic MRI brain images with different levels of noise and intensity inhomogeneity. The experiment is conducted on 210 images with different noise levels from 3% to 5% and different intensity inhomogeneity levels from 20% to 40%. We analyze and compare the results from both the qualitative and quantitative aspects.

Figs. 9–11 show some comparison segmentation results of synthetic MRI brain images with different levels of noise and intensity inhomogeneity, respectively.

According to these segmentation results, we can make lateral comparison and longitudinal comparison. As shown in Fig. 9(g), (h), and (j), FLICM, KWFLICM, and KGFCM are very effective for suppressing the effect of noise. However, they lose some important image details which can be observed from the enlarged image. Combined with Figs. 10 and 11, we can infer that FLICM, KWFLICM, and KGFCM would loss more image details as the level of noise or intensity inhomogeneity increases. PFCM, IFCM, CFFCM, and ETFCM are sensitive to intensity inhomogeneity by comparing (b), (e), (f), and (k) in Figs. 9 and 10. Also, Figs. 10 and 11 show that they are sensitive to noise. FCM_S performs well in noise image, but the image is smoothed as shown in Fig. 11(c). GIFP_FCM performs well, but it is a little sensitive to noise as shown in Fig. 11(i). By comparing the segmentation results of all the clustering methods in Figs. 9–11, the proposed method outperforms the other comparison methods and the results of the proposed method are the closest to the ground truth.

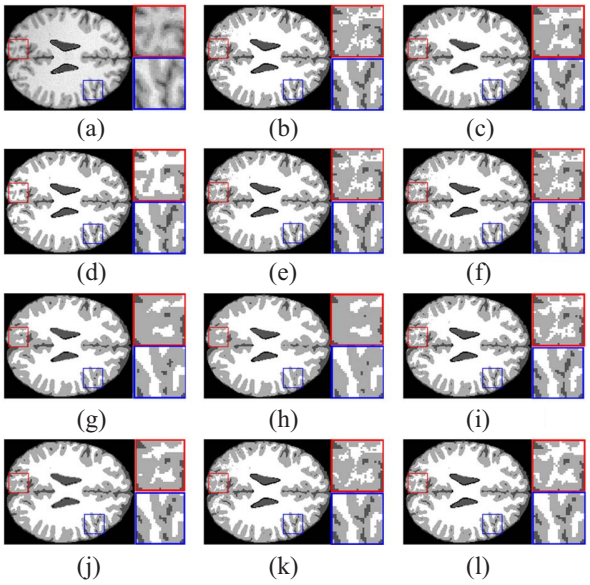


Fig. 10. Comparison segmentation results of synthetic MRI brain images with noise level 3% and intensity inhomogeneity level 40% (N3F40). (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

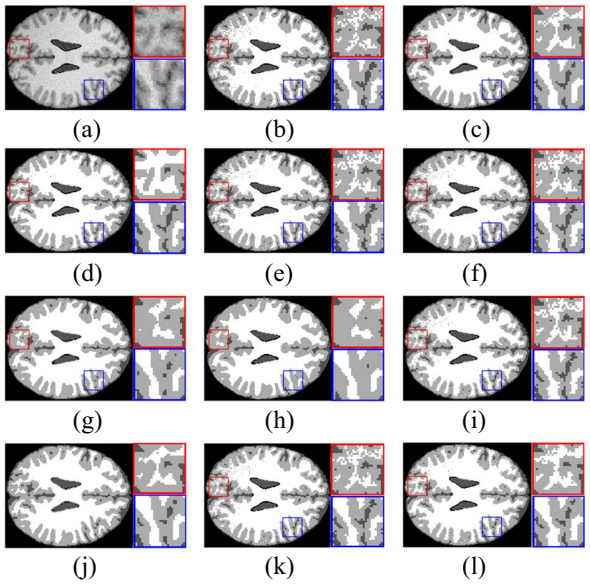


Fig. 11. Comparison segmentation results of synthetic MRI brain images with noise level 5% and intensity inhomogeneity level 40% (N5F40). (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

For quantitative comparative analysis, we use the dice coefficient (DC) value [38] to measure the capability of these methods. The DC value is defined as [38]

$$\mathrm{DC}(SR_1, SR_2) = 2\frac{|SR_1 \cap SR_2|}{|SR_1| + |SR_2|} \tag{30}$$

where SR_1 represents the segmentation result and SR_2 is the ground truth. The higher DC value corresponds to a more accurate segmentation result.

We compute the average DC values of CSF, WM, and GM of images with different levels of noise and intensity inhomogeneity. The results are listed in Table I.

From Table I, we can find that the proposed method has the highest DC values for the segmentation of all the synthetic MRI brain images, which indicates that the proposed method is robust to noise and has some defensive effect to intensity inhomogeneity. As shown in Table I, FLICM, KWFLICM, and KGFCM have better DC values of WM, GM than CSF. This reflects that FLICM, KWFLICM, and KGFCM sacrifice the image details for noise resistance. Although PFCM does not perform well in the DC values, it can help identify outliers, as pointed out in [13]. IFCM performs well in synthetic images, but it is sensitive to noise and intensity inhomogeneity. As the noise level increases, the performance of IFCM descends rapidly. FCM_S is relatively stable as the noise level changes, but it does not produce a satisfactory result. CFFCM could cluster the nonspherical data, such as the MRI brain images. However, CFFCM is sensitive to noise. GIFP_FCM performs well for MRI brain images due to the constraint of membership function, but it is still sensitive to noise. ETFCM could achieve a good segmentation result because ETFCM uses the prior knowledge. However, without considering the spatial information, ETFCM is sensitive to noise. The proposed method introduces interclass similarity with local label constraint and provides a new formula to measure the similarity, which is proven to be effective from Table I.

D. Segmentation of Clinical MRI Brain Image

Our emphasis is on evaluating the effectiveness of our method on the segmentation of 20 groups of clinical brain MRI images (IBSR_20 Normals). IBSR provides the ground truth segmentation results which can help us make a comparison of the proposed method with the comparison methods. The brain MRI image dataset can be acquired from [35]. In our experiments, all the images from 20 groups containing CSF, GM, and WM are used. There are 679 images in total. We analyze and compare the results of the proposed method and the comparison methods from both the qualitative and quantitative aspects too.

For qualitative comparative analysis, we applied six methods and the proposed method on the clinical MRI brain images. The results are shown in Fig. 12.

Fig. 12(a) and (d) shows the original image and the ground truth, respectively. Fig. 12(b) is the segmentation result of PFCM. PFCM is sensitive to noise because PFCM does not consider spatial information. And in Fig. 12(b), many pixels belonging to CSF are misclassified because the original image has only a few CSFs. IFCM introduces the intuition level, which can identify inliers to some extent. However, IFCM ignores the spatial information, which is still sensitive to noise as shown in Fig. 12(e). FCM_S, FLICM, KWFLICM, and KGFCM consider the neighbor information, thus they are less sensitive to noise, as shown in Fig. 12(c), (g), (h), and (j), respectively. However, FCM_S is sensitive to the parameter selection and may omit some important image

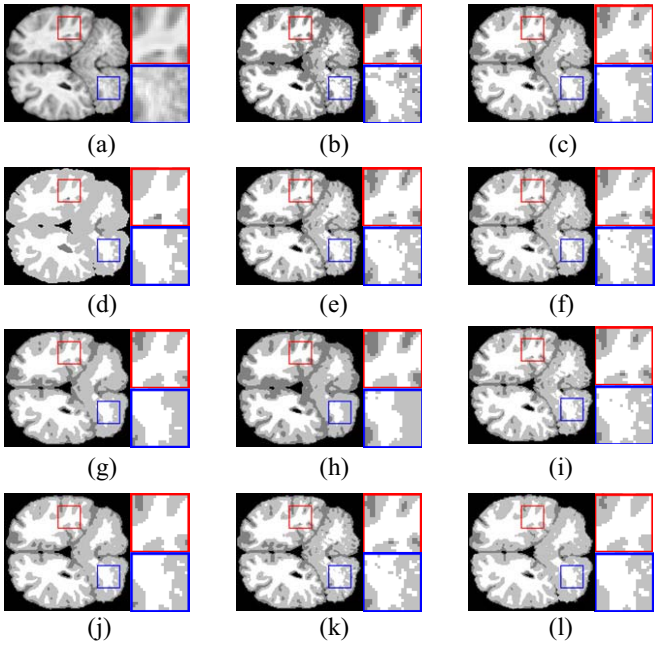


Fig. 12. Comparison segmentation results of clinical MRI brain images of different algorithms. (a) Original image. (b) PFCM. (c) FCM_S. (d) Ground truth. (e) IFCM. (f) CFFCM. (g) FLICM. (h) KWFLICM. (i) GIFP_FCM. (j) KGFCM. (k) ETFCM. (l) Proposed.

details. Although FLICM is free of parameters and is effective for noise images, it misses some details. If there are many details in the image, FLICM usually does not perform well. KWFLICM takes more neighbor information into account and makes the image smooth. KWFLICM has an adaptive parameter σ, which affects the performance for different images. KGFCM loses some important image details. Moreover, FCM_S, PFCM, IFCM, and FLICM adopt the Euclidean distance as the similarity measure, which may not be suitable for nonlinearly separable data like the MRI brain image. The results of CFFCM, GIFP_FCM, and ETFCM have retained more details but are still sensitive to noise. The proposed method introduces the interclass similarity and local label information which can cluster the pixels more accurately. This strategy protects more details as well as avoids noise. The results show that the proposed method is superior to the comparison methods for the segmentation of the clinical MRI brain image.

For quantitative comparison analysis, we conduct our experiments on the clinical MRI brain images from IBSR [35], where images of 20 groups containing CSF, WM, and GM are all used. The accuracy of the clustering results is measured by the average DC values on CSF, WM, and GM of all the images. The comparison results of those six methods and the proposed method are listed in Table II.

Table II shows that the proposed method has the highest average DC values, which indicates that the proposed method is more suitable for the segmentation of clinical MRI brain image. Especially, DC values of CSF of our method are also the highest, which indicates that the proposed method is less sensitive to cluster sizes than other methods. Although CSF is only a few in clinical MRI brain image, the proposed

TABLE II
DC VALUES FOR THE SEGMENTATION RESULTS ON CLINICAL MRI BRAIN IMAGES

Method	WM	GM	CSF
PFCM	0.8156	0.5317	0.0566
FCM_S	0.8224	0.6601	0.0886
IFCM	0.8222	0.6466	0.0764
CFFCM	0.8186	0.6968	0.0984
FLICM	0.8107	0.6756	0.1031
KWFLICM	0.8035	0.6904	0.1198
GIFP_FCM	0.8220	0.6797	0.0879
KGFCM	0.8193	0.7047	0.1123
ETFCM	0.8186	0.6587	0.0796
Proposed	**0.8227**	**0.7484**	**0.1384**

method can preserve more image details of CSF than other methods. This means the proposed method could well handle the problem of cluster-size sensitivity. The DC values of GM of the proposed method are higher than other methods. This verifies that the proposed method is more robust to noise. Generally, for the segmentation of clinical MRI brain image, the methods considering neighbor information would be less sensitive to noise, and thus have a higher DC value of GM at the cost of damaging image detail. Then the DC values of WM decline, such as FLICM, KWFLICM, and KGFCM. However, the proposed method improves the accuracy of GM and CSF as well as maintains the highest DC values of WM. Our method outperforms all the comparison methods.

V. CONCLUSION

In this paper, we propose the similarity measure-based PFCM method with local label information for the segmentation of MRI brain image. This method is based on PFCM that can identify the outliers in an image. And by defining the similarity measure based on intraclass similarity and interclass similarity, the proposed method could cluster both the linearly and nonlinearly separable data. The proposed method could also ameliorate the cluster-size sensitivity problem. Moreover, the proposed method utilizes the local label information to preserve more image details as well as to suppress the effect of noise. Experiments on artificial and simulated images show that our method is suitable for the data with complex situations, such as intensity inhomogeneity, noise, and unbalanced cluster sizes. Those interferences are exactly the difficulties for the segmentation of the MRI brain image. But the proposed method could handle those interferences well. The average clustering results on synthetic and clinical MRI images indicate that the proposed method is effective and performs better than all the comparison methods for the segmentation of MRI brain image.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments, which have enhanced this paper.

REFERENCES

[1] S. R. Kannan, R. Devi, S. Ramathilagam, and K. Takezawa, "Effective FCM noise clustering algorithms in medical images," *Comput. Biol. Med.*, vol. 43, no. 2, pp. 73–83, 2013.

[2] J. Zhang *et al.*, "Incorporating MRI structural information into bioluminescence tomography: System, heterogeneous reconstruction and in vivo quantification," *Biomed. Opt. Express*, vol. 5, no. 6, pp. 1861–1876, 2014.

[3] Y. Chen, H. Zhang, Y. Zheng, B. Jeon, and Q. M. J. Wu, "An improved anisotropic hierarchical fuzzy c-means method based on multivariate student t-distribution for brain MRI segmentation," *Pattern Recognit.*, vol. 60, pp. 778–792, Dec. 2016.

[4] M. B. Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J. P. Thiran, "Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images," *IEEE Trans. Med. Imag.*, vol. 24, no. 12, pp. 1548–1565, Dec. 2005.

[5] C.-C. Wong, C.-C. Chen, and M.-C. Su, "A novel algorithm for data clustering," *Pattern Recognit.*, vol. 34, no. 2, pp. 425–442, 2001.

[6] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. New York, NY, USA: Academic Press, 2008.

[7] P.-L. Lin, P.-W. Huang, C. H. Kuo, and Y. H. Lai, "A size-insensitive integrity-based fuzzy c-means method for data clustering," *Pattern Recognit.*, vol. 47, no. 5, pp. 2042–2056, 2014.

[8] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.

[9] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[10] X. Bai, Z. Chen, Y. Zhang, Z. Liu, and Y. Lu, "Infrared ship target segmentation based on spatial information improved FCM," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3259–3271, Dec. 2016.

[11] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1973.

[12] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193–199, Mar. 2002.

[13] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.

[14] H. Timm and R. Kruse, "A modification to improve possibilistic fuzzy cluster analysis," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Honolulu, HI, USA, 2002, pp. 1460–1465.

[15] D.-C. Park, "Intuitive fuzzy c-means algorithm for MRI segmentation," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Seoul, South Korea, 2010, pp. 1–7.

[16] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy c-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573–584, Feb. 2013.

[17] T. Celik and H. K. Lee, "Comments on 'a robust fuzzy local information c-means clustering algorithm,"' *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1258–1261, Mar. 2013.

[18] L. Szilágyi, "Lessons to learn from a mistaken optimization," *Pattern Recognit. Lett.*, vol. 36, pp. 29–35, Jan. 2014.

[19] L. Zhu, F.-L. Chung, and S. Wang, "Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 3, pp. 578–591, Jun. 2009.

[20] F. Zhao, L. Jiao, and H. Liu, "Kernel generalized fuzzy c-means clustering with spatial information for image segmentation," *Digit. Signal Process.*, vol. 23, no. 1, pp. 184–199, 2014.

[21] Z. Deng *et al.*, "Transfer prototype-based fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1210–1232, Oct. 2016.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Int. Conf. Knowl. Disc. Data Min.*, Portland, OR, USA, 1996, pp. 226–231.

[23] J. Hou, H. Gao, and X. Li, "DSets-DBSCAN: A parameter-free clustering algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3182–3193, Jul. 2016.

[24] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[25] J. Ding, R. Ma, X. Hu, J. Yang, and S. Chen, "Fuzzy c-means revisited: Towards a cluster-center-free reformulation," in *Proc. Chin. Conf. Pattern Recognit.*, Chongqing, China, 2010, pp. 1–5.

[26] X. Bai, Z. Chen, M. Liu, and Y. Zhang, "Center-free PFCM for MRI brain image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, 2015, pp. 656–660.

[27] J. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.

[28] R. Hettiarachchi and J. F. Peters, "Voronoï region-based adaptive unsupervised color image segmentation," *Pattern Recognit.*, vol. 65, pp. 119–135, May 2017.

[29] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.

[30] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 100, no. 1, pp. 9–34, 1999.

[31] L. A. Zadeh, "A theory of approximate reasoning," in *Machine Intelligence*, vol. 9, J. Hayes, D. Michie, and L. Mikulich, Eds. New York, NY, USA: Halstead Press, 1979, pp. 149–194.

[32] R. N. Dave and R. Krishnapuram, "Robust clustering methods: A unified view," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 2, pp. 270–293, May 1997.

[33] J. C. Noordam, W. H. A. M. van den Broek, and L. M. C. Buydens, "Multivariate image segmentation with cluster size insensitive fuzzy c-means," *Chemometr. Intell. Lab. Syst.*, vol. 64, no. 1, pp. 65–78, 2002.

[34] D.-M. Tsai and C.-C. Lin, "Fuzzy c-means based clustering for linearly and nonlinearly separable data," *Pattern Recognit.*, vol. 44, no. 8, pp. 1750–1760, 2011.

[35] NITRC. (2016). *The Internet Brain Segmentation Repository*. [Online]. Available: https://www.nitrc.org/projects/ibsr/

[36] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans. (2016). *BrainWeb: Online Interface to a 3D MRI Simulated Brain Database*. [Online]. Available: http://www.bic.mni.mcgill.ca/brainweb/

[37] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[38] 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.

Authors' photographs and biographies not available at the time of publication.