# Robust Self-Sparse Fuzzy Clustering for Image Segmentation

**XIAOHONG JIA**[1], **TAO LEI**[2], **(Senior Member, IEEE), XIAOGANG DU**[2], **SHIGANG LIU**[3],
**HONGYING MENG**[4], **(Senior Member, IEEE), AND ASOKE K. NANDI**[4], **(Fellow, IEEE)**

[1]School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China
[2]School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China
[3]School of Computer Science, Shaanxi Normal University, Xi'an 710119, China
[4]Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, U.K.

Corresponding author: Tao Lei (leitao@sust.edu.cn)

**ABSTRACT** Traditional fuzzy clustering algorithms suffer from two problems in image segmentations. One is that these algorithms are sensitive to outliers due to the non-sparsity of fuzzy memberships. The other is that these algorithms often cause image over-segmentation due to the loss of image local spatial information. To address these issues, we propose a robust self-sparse fuzzy clustering algorithm (RSSFCA) for image segmentation. The proposed RSSFCA makes two contributions. The first concerns a regularization under Gaussian metric that is integrated into the objective function of fuzzy clustering algorithms to obtain fuzzy membership with sparsity, which reduces a proportion of noisy features and improves clustering results. The second concerns a connected-component filtering based on area density balance strategy (CCF-ADB) that is proposed to address the problem of image over-segmentation. Compared to the integration of local spatial information into the objective functions, the presented CCF-ADB is simpler and faster for the removal of small areas. Experimental results show that the proposed RSSFCA addresses two problems in current fuzzy clustering algorithms, i.e., the outlier sensitivity and the over-segmentation, and it provides better image segmentation results than state-of-the-art algorithms.

**INDEX TERMS** Fuzzy c-means clustering (FCM), image segmentation, sparse membership, over-segmentation.

## I. INTRODUCTION

Clustering is an unsupervised classification algorithm and it aims to classify data into several disjoint subsets depending on data features [1]. As one of the important techniques of data analysis and data mining, clustering has been widely used for text classification, biometric features recognition, image segmentation, etc. In these applications, image segmentation based on clustering are very popular and a large number of algorithms have been proposed and used in the fields of medicine [2], remote sensing [3], intelligent transportation [4], etc. Popular clustering algorithms mainly include hierarchical clustering [5], spectral clustering [6], fuzzy clustering [7], affinity propagation clustering [8], density peak clustering [9], etc. In this paper, we mainly focus

The associate editor coordinating the review of this manuscript and approving it for publication was Byung Cheol Song.

on fuzzy clustering based on object function optimization for image segmentation.

Fuzzy c-means clustering (FCM) often suffers from two problems for image segmentation: firstly, it is sensitive to noise [10]; secondly, it lacks the local spatial information of images leading to over-segmentation [11]. To address these drawbacks, researchers proposed many improved fuzzy clustering algorithms. They mainly adopted two strategies; one is to utilize regularization approaches to achieve self-optimization of FCM, the other one is to incorporate local spatial information into the object functions of FCM to improve image segmentation.

To improve the robustness of FCM against noise, a lot of improved FCM algorithms based on self-optimization have been presented. Ahmed *et al.* [12] proposed bias correction FCM (BCFCM) that integrates a bias field into its objective function. Although the bias field is able to correct

some pixels corrupted by noise, it shows low robustness for different kinds of noisy images since the bias field is often not sparse. Based on this work, Zhang *et al.* [13] proposed a deviation-sparse fuzzy c-means (DSFCM) that utilizes regularization with sparsity constraint to mitigate significantly the defect of BCFCM. Though DSFCM is able to provide accurate clustering centers, it is sensitive to regularization parameters and thus shows low robustness. Motivated by the Network lasso [14], Guo *et al.* [15] further explored the affinity-based regularization and proposed the membership affinity lasso regularized fuzzy c-means clustering (MalFCM). The MalFCM generates better classification results than DSFCM via building an affinity matrix, but MalFCM requires high computational complexity since the alternating direction method of multipliers (ADMM) [16] is used to optimize the affinity matrix. However, both DSFCM and MalFCM depend on a condition that the sum of membership values of each pixel is equal to 1. To loosen this condition, Pal *et al.* [17] proposed possibilistic FCM (PFCM) by considering memberships and possibilities, which hybridizes FCM and possibilistic c-means [18] with looser constraint on membership. Although PFCM can mine more informative descriptions of data, it lacks robustness against data with non-spherical distribution. To settle the shortcoming, Bai *et al.* [19] proposed similarity measure-based PFCM. Continuing in direction of membership, Zhou *et al.* [20] reported a membership scaling FCM (MSFCM) based on triangle inequality. The MSFCM not only effectively improves the convergence speed of the model but also maintains the accuracy of data clustering.

In most improved FCM algorithms mentioned above, the fuzzifier exponent is often considered as a constant and its value is usually equal to 2. However, in practical applications, the change of the value of fuzzifier exponent easily leads to different segmentation effect [21], [22]. To address this problem, Miyamoto and Mukaidono [23] proposed maximum entropy FCM (MEFCM) that avoids the selection of the value of fuzzifier exponent and simplifies its calculation. Inspired by MEFCM, both [24] and [25] introduced relative entropy and kernel distance to improve their objective function. However, these FCM algorithms based on entropy theory do not consider the sparsity of memberships, which leads to the misclassification for outliers. Recently, Xenaki *et al.* [26], Koutroumbas *et al.* [27] verified a conclusion that the sparsity of fuzzy memberships can deal well with closely located clusters. Similarly, Xu *et al.* [28] also demonstrated that the sparsity can avoid performance degradation of clustering algorithms. These previous reports show that fuzzy memberships with sparsity can positively improve clustering results. However, it is still a challenge to obtain reasonable sparsity of fuzzy membership. Moreover, it is difficult to apply those algorithms mentioned above to image segmentation due to the lack of local spatial information.

To reduce over-segmentation caused by fuzzy clustering algorithms, the local spatial information of images is usually incorporated into their objective functions [29]–[33].

The utilization of local spatial information can be grouped into two categories: obtaining the neighboring information in a fixed-size window or in an adaptive neighboring window. Early improved FCM algorithms usually employ a fixed-size window of size $w \times w$ to obtain local spatial information such as FCM_S [12], FCM_S1 [29], FCM_S2 [29], FLICM [30], KWFLICM [33], DSFCM_N [13] etc. Similar to the above algorithms, recently, Mishro *et al.* [34] proposed a novel type-2 adaptive weighted spatial FCM (AWSFCM) clustering algorithm that employs a fuzzy linguistic fuzzifier and spatial information of membership to reduce misclassification of pixels. However, in practical applications, a large window usually leads to rich spatial information but high computational cost; on the contrary, a small window leads to low computational cost but limited spatial information. Therefore, a middle-size window like $3 \times 3$ or $5 \times 5$ is popular for these algorithms since the window achieves a balance between spatial information and computational cost. To overcome the limitation of fixed-size windows, some researchers employ superpixel techniques to obtain adaptive neighboring information such as Liu's algorithm [35], FDCM_SSR [36], SFFCM [37], AFCA [38], etc. Although the second strategy improves segmentation accuracy due to the utilization of adaptive neighboring information, these algorithms seriously depend on the selection of superpixel algorithms [39]–[41]. A good superpixel algorithm not only improves the final segmentation result, but also reduces computational cost efficiently. However, the segmentation result will be worse than one obtained by FCM if the superpixel algorithm is unsuitable for images to be segmented. In particular, superpixel algorithms are unavailable for images with low-contrast or blurred edges.

For the reduction of over-segmentation, image filtering is also a popular strategy. In [42], Szilagyi *et al.* proposed enhanced FCM (EnFCM) by introducing histogram to its objective function and applying a local linear-weight filtering to each pixel. As the number of gray-levels is generally much smaller than the number of pixels in a grayscale image, the EnFCM achieves high computational efficiency in image segmentation. On the basis of this work, Cai *et al.* [43] proposed the fast generalized FCM (FGFCM) by integrating a bilateral filter to its objective function. Furthermore, Zhao *et al.* [44] presented a FCM algorithm with self-tuning non-local spatial information [45]. However, the computation of non-local spatial information is time-consuming. Inspired by FGFCM, Guo *et al.* [46] proposed a noise detecting FCM (NDFCM) with auto-tuning parameters by measuring local variance of gray levels. More recently, Lei *et al.* [47] proposed a fast and robust FCM (FRFCM) based on morphological reconstruction and membership filtering. FRFCM achieves good segmentation results and requires short execution time for different kinds of grayscale images.

According to those studies mentioned above, currently popular fuzzy clustering algorithms still suffer from two challenges for image segmentation tasks. The first is that these algorithms lack immunity to outliers due to the non-sparsity

of fuzzy memberships. The second is that it is difficult to overcome effectively over-segmentation for them. To solve these challenges, we employ a novel regularization to obtain sparse fuzzy memberships, and use a connected-component filtering algorithm based on area density balance strategy (CCF-ADB) to achieve region merging adaptively, and finally propose a robust self-sparse fuzzy clustering algorithm (RSSFCA) for image segmentation. The proposed RSSFCA[1] has two advantages:

- The RSSFCA utilizes regularization under Gaussian metric to obtain proper sparse memberships that can effectively reduce non-homogenous interference and achieve better classification than popular self-optimized FCM algorithms.
- The RSSFCA employs the CCF-ADB to achieve automatically region merging, which is superior to the strategy of incorporating local spatial information and helps RSSFCA to achieve better image segmentation than improved FCM algorithms within local spatial information.

The organization of this paper is presented as follows. We first discuss our motivations in Section II. Then, we introduce the proposed algorithm and analyze its advantages in Section III. Thirdly, we conduct experiments and discuss experimental results in Section IV. Finally, we provide conclusion in Section V.

## II. MOTIVATION

To improve fuzzy clustering algorithms for image segmentation, researchers usually adopt two strategies. One is introducing regularization terms into objective functions to improve clustering results, the other one is integrating local spatial information into object functions to overcome over-segmentation. As classification results of pixels are decided by fuzzy memberships, a good objective function should meet the constraint of sparsity of fuzzy memberships. However, it is difficult to obtain sparse fuzzy memberships for popularly self-optimized FCM algorithms such as DSFCM and MalFCM. Motivated by this, we introduce a new regularization to the objective function of FCM. Most of improved FCM algorithms employ a fixed-size window or adaptive neighboring windows to overcome over-segmentation, but the utilization of local spatial information often leads to smooth boundaries or high computational cost. In order to address the problem, we present a connected-component filtering based on area density balance strategy to achieve adaptive region merging, which can maintain boundary accuracy and requires low computational cost.

### A. SELF-OPTIMIZED FCM

Let $X = \{x_1, x_2, \ldots, x_n\} \in \Re^{D \times n}$ be an unlabeled data set, and we aim to spilt $X$ into $c$ disjoint clusters, with the corresponding clustering centers $V = \{v_1, v_2, \ldots, v_c\}$. The FCM algorithm [7] uses Lagrange multiplier technique

[1] Source code is available at https://github.com/SUST-reynole/RSSFCA

to find optimal solutions with respect to sum of squared Euclidean distance error. The objective function of FCM is defined as:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \|x_j - v_i\|^2, \tag{1}$$

where $c$ denotes the number of clusters, $n$ denotes the number of samples, $u_{ij}$ is the degree of membership of $x_j$ with respect to the clustering center $v_i$, $0 \le u_{ij} \le 1$ and $\sum_{i=1}^{c} u_{ij} = 1$, $m$ is the fuzzification exponent for the partition matrix $u_{ij}$, and $\|\cdot\|$ represents the Euclidean norm.

It is well-known that FCM is sensitive to noise and outliers in data, and it lacks robustness for non-spherical data clustering due to the employment of square loss in (1). Regularization-based FCM algorithms address the problem through adding a constraint, i.e., a regularization term, to their objective functions. The regularization term is able to help FCM to improve clustering accuracy for data corrupted by noise or non-spherical data. Zhang *et al.* [13] introduced the deviation-sparse to FCM (DSFCM) to construct a novel objective function that is defined as:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \|x_j - e_j - v_i\|^2 + \lambda \sum_{j=1}^{n} |e_j|, \tag{2}$$

where $e_j$ is the deviation between $x_j$ and its the theoretical value, and $\lambda$ is the regularization scalar. It is easy to obtain a sparse matrix $E = \{e_1, e_2, \ldots, e_n\} \in \Re^{D \times n}$ by using soft-thresholding [48]. The DSFCM can improve classification accuracy for noisy data. On the one hand, it uses large deviation values of $e_j$ to revise $x_j$ corrupted by noise, and on the other hand, it uses $x_j$ with small value to maintain the original $x_j$ uncorrupted by noise. Therefore, the DSFCM can improve the robustness of FCM for noisy data clustering because of the introduction of regularization term into its objective function. However, it is difficult to set the value of $\lambda$ for different kinds of data.

Inspired by the Network lasso [14], Guo *et al.* [15] developed a new regularization of membership affinity lasso (MalFCM) and presented the objective function as follows:

$$J = \sum_{i=1}^{c} \left( \sum_{j=1}^{n} u_{ij}^m \|x_j - v_i\|^2 + \sum_{j=1}^{c} \sum_{k=1}^{n} w_{jk} |u_{ij} - u_{ik}| \right), \tag{3}$$

where $w_{jk}$ is the affinity of the $j$th point $x_j$ to the $k$th point $x_k$. The MalFCM employs the alternating direction method of multipliers (ADMM) [16] to optimize the membership affinity lasso. Consequently, it achieves accurate classification for complex data due to the consideration of membership similarity. However, the MalFCM requires high computational cost because ADMM is time-consuming.

Noted that both DSFCM and MalFCM require to set the value of parameter $m$. To reduce the influence of parameters, Miyamoto and Mukaidono [23] integrated a penalty term of

maximum entropy (MEFCM) into its objective function:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} \|x_j - v_i\|^2 + \lambda \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} ln(u_{ij}). \quad (4)$$

Here, the entropy term works as the degree of fuzzifier. As a result, the MEFCM not only avoids the parameter setting on fuzzification exponent $m$, but also minimizes the intra-class dispersion and maximizes the inter-class negative weight entropy. However, it is still difficult to set the value of $\lambda$ for MEFCM.

Regularization-based FCM algorithms can improve clustering results for complex data, but they face two difficulties for complex non-spherical data. The first is that the Euclidean distance shows poor robustness for non-spherical data, and the second is that fuzzy membership is non-sparse, easily leading to the ambiguity for identified data. To address the difficulties, we use the Gaussian distribution instead of the Euclidean distance, and integrate the appropriate regularization term to our objective function to make sure that fuzzy memberships are sparse. We will present detailed description of the proposed strategy in Section III. A.

### B. FCM WITH LOCAL SPATIAL INFORMATION
Image segmentation algorithms based on FCM easily cause over-segmentation since FCM ignores the local spatial information of images. To overcome the drawback, a large number of improved FCM algorithms have been proposed by incorporating local spatial information into their objective functions [29-33]. These algorithms can be categorized into two groups: The first group requires long execution time due to the computation of local spatial information in each iteration, while the second group shows high computational efficiency since the local spatial information is computed only once before iterations. Generally, the objective function of the first group can be abstracted as follows:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \|x_j - v_i\|^2 + \sum_{i=1}^{c} \sum_{j=1}^{n} G_{ij}, \quad (5)$$

where $G_{ij}$ denotes fuzzy factor that is used to balance noise suppression and the edge detail preservation. The value of $G_{ij}$ depends on pixels within neighboring windows and it can be changed in each iteration. If $x_j$ is a pixel corrupted by noise, then it will be replaced with $G_{ij}$ to reduce the influence of outliers. Therefore, the introduction of constraint terms into objective functions improves the noise tolerance and outliers resistance for image segmentation, and different forms of $G_{ij}$ lead to variously improved FCM algorithms, such as FCM_S, FCM_S1/S2, FLICM, etc.

Although those algorithms mentioned above can improve image segmentation effect, they are impractical due to high computational complexity. To reduce the computational cost, researchers designed fast FCM algorithms that avoid the redundant computation of local spatial information. The objective function of fast FCM algorithms is defined as

follows:

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n'} r_j u_{ij}^m \|x_j' - v_i\|^2, \quad (6)$$

where $n'$ denotes the number of gray-level of test images, and $n' \ll n$. The $x_j'$ denotes the $j$th pixel in the filtered image. We can define $x_j' = F(x_{r \in j})$, where $F$ represents a filter, $x_r$ stands for a pixel fallowing into the neighboring window of $x_j$. Based on different filters, a lot of FCM algorithms used for fast image segmentation are proposed such as EnFCM, FGFCM, NDFCM, FRFCM, etc.

It is clear that the introduction of local spatial information is very popular for improving image segmentation effect, but it suffers from high computational cost. Although many fast FCM algorithms are proposed by incorporating gray-level histogram or superpixels into their objective functions [37], it is still a challenge to obtain excellent segmentation results. For this problem, Comaniciu and Meer [39] presented an auxiliary strategy in mean-shift algorithm, namely eliminating spatial regions containing the number of pixels less than $M$, where $M$ is a threshold. Inspired by this idea, we propose the CCF-ADB to help FCM to overcome over-segmentation. The CCF-ADB is simple and effective, it is superior to the strategy of integrating local spatial information, and achieves automatic region merging. We will comprehensively describe the CCF-ADB in Section III. B.

## III. METHODOLOGY
In this study, we propose a novel self-sparse fuzzy c-means clustering algorithm (SSFCA) based on regularization approaches to obtain suitable sparse fuzzy memberships. Meanwhile, we utilize the CCF-ADB to merge useless small regions. The proposed RSSFCA can effectively overcome outlier sensitivity and over-segmentation, and thus improve segmentation results.

### A. SELF-SPARSE FUZZY C-MEANS
Based on the analysis in Section II. A, DSFCM, MalFCM, and MEFCM cannot obtain sparse fuzzy memberships. To address this issue, we introduce a novel regularization approach by considering $u_{ij}^2$ as a penalty term. We define the objective function of SSFCA as:

$$\widetilde{J} = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} \Phi(x_j|v_i, \Sigma_i) + \gamma \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2, \quad (7)$$

where $\Phi(x_j|v_i, \Sigma_i)$ represents the distance function between $x_j$ and $v_i$, $\gamma$ is a balance factor used for controlling the sparsity of memberships. By changing the value of $\gamma$, the objective function shows different degrees of robustness to outliers or noises.

In (7), if the fuzzy membership is sparse, the first term of $\widetilde{J}$ will be small while the second term will be large. The SSFCA often requires more iterations than $k$-means but fewer iteration than FCM for optimal computation. It is clear that

the novel objective function $\widetilde{J}$ achieves a balance between $k$-means and FCM. The obtained fuzzy membership is more sparse than that provided by FCM. In contrast with $k$-means, some fuzzy membership values are not 0 or 1. Besides, $\Phi(x_j|v_i, \Sigma_i)$ is defined as:

$$\Phi(x_j|v_i, \Sigma_i) = ln(-\rho(x_j|v_i, \Sigma_i)), \tag{8}$$

where $\rho(x_j|v_i, \Sigma_i)$ is the Gaussian density function, and it is defined as:

$$\rho(x_j|v_i, \Sigma_i) = \frac{exp(-\frac{1}{2}(x_j - v_i)^T \Sigma_i^{-1}(x_j - v_i))}{(2\pi)^{(D/2)}|\Sigma_i|^{(1/2)}}, \tag{9}$$

where $D$ denotes the dimension of input data, $\Sigma_i$ denotes covariance matrix that describes the intra-class dispersion of the $i$th class. Substituting (9) into (8), we obtain:

$$\Phi(x_j|v_i, \Sigma_i) = \frac{1}{2}((x_j - v_i)^T \Sigma_i^{-1}(x_j - v_i) + ln|\Sigma_i| + Dln(2\pi)), \tag{10}$$

Note that the distance metric in (10) is different from the Mahalanobis distance, since the former includes the variable $ln|\Sigma_i|$ that may be a large negative value for dense distribution data. The $\Phi(x_j|v_i, \Sigma_i)$ may be unsatisfied for the constraint of nonnegative values due to the effect of $ln|\Sigma_i|$ as shown in Fig. 1.
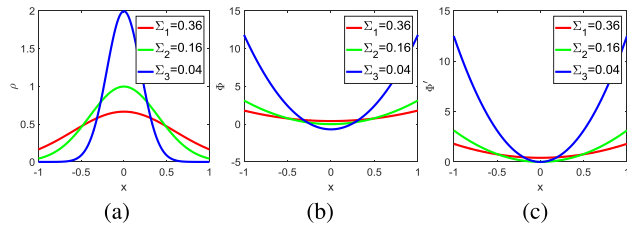


**FIGURE 1.** Distribution curve and corresponding distance curve. (a) Normal distribution curve $\rho$. (b) Distance curve $\Phi$. (c) Corrected distance curve $\Phi'$.

In Fig. 1(a), there are three groups of data provided by (9) with parameters $v_1 = 0$, and $\Sigma_1 = 0.36$; $v_2 = 0$, and $\Sigma_2 = 0.16$; and $v_3 = 0$, and $\Sigma_3 = 0.04$. It can be easily found that a smaller value of covariance corresponds to a compact curve, i.e., compact data distribution. Fig. 1(b) shows the result provided by (10). Note that the blue curve includes negative values, which violates the requirement of positive values on distance measure metric. With the decrease of the covariance value, more negative values lead to serious errors in distance measure and misclassification. To solve this problem, we use $\Phi'(x_j|v_i, \Sigma_i)$ instead of $\Phi(x_j|v_i, \Sigma_i)$, where

$$\Phi'(x_j|v_i, \Sigma_i) = \begin{cases} \Phi - min(\Phi) & min(\Phi) < 0 \\ \Phi & otherwise. \end{cases} \tag{11}$$

Fig. 1(c) shows that $\Phi'(x_j|v_i, \Sigma_i)$ satisfies the non-negative constraint of distance. Substituting $\Phi'(x_j|v_i, \Sigma_i)$ into (7), the final objective function is defined as:

$$\widetilde{J}' = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}\Phi'(x_j|v_i, \Sigma_i) + \gamma \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2. \tag{12}$$

For each sample $x_j$, the $\widetilde{J}'$ can be separated into $c$ sub-problems with constraint conditions $0 \leq u_{ij} \leq 1$ and $\sum_{i=1}^{c} u_{ij}^m = 1$. Then we get

$$\widetilde{J}'_j = min \sum_{i=1}^{c} (u_{ij}\Phi'(x_j|v_i, \Sigma_i) + \gamma u_{ij}^2). \tag{13}$$

through simplification, $\widetilde{J}'_j$ can be rewritten as:

$$\widetilde{J}'_j = min\|u_{ij} - h_{ij}\|^2, \tag{14}$$

where $h_{ij} = -\Phi'(x_j|v_i, \Sigma_i)/2\gamma$. Utilizing the optimization strategy proposed in [49] to solve (14), we obtain fuzzy memberships with different degrees of sparsity by tuning the value of $\gamma$.

Similarly, the $\widetilde{J}'_i$ can be separated into $n$ independent sub-problems to obtain the clustering center $v_i$ by solving $\frac{\partial \widetilde{J}'_j}{\partial v_i} = 0$,

$$\frac{\partial \widetilde{J}'_i}{\partial v_i} = \sum_{j=1}^{n} u_{ij} \left( \frac{\partial \begin{bmatrix} (x_j - v_i)^T \Sigma_i^{-1}(x_j - v_i) \\ + ln|\Sigma_i| + Dln(2\pi) \end{bmatrix}}{\partial v_i} \right)$$

$$= \sum_{j=1}^{n} u_{ij}(x_j - v_i)$$

$$= 0,$$

we get

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}x_j}{\sum_{j=1}^{n} u_{ij}}. \tag{15}$$

Furthermore, the updated covariance matrix $\Sigma_i$ is obtained by solving $\frac{\partial \widetilde{J}'_i}{\partial \Sigma_i} = 0$,

$$\frac{\partial \widetilde{J}'_i}{\partial \Sigma_i} = \sum_{j=1}^{n} u_{ij} \left( \frac{\partial \begin{bmatrix} (x_j - v_i)^T \Sigma_i^{-1}(x_j - v_i) \\ + ln|\Sigma_i| + Dln(2\pi) \end{bmatrix}}{\partial \Sigma_i} \right)$$

$$= \sum_{j=1}^{n} u_{ij} \left( (x_j - v_i)^T \Sigma_i^{-2}(x_j - v_i) + \Sigma_i^{-1} \right)$$

$$= 0,$$

the solution yields

$$\Sigma_i = \frac{\sum_{j=1}^{n} u_{ij}(x_j - v_i)^T (x_j - v_i)}{\sum_{j=1}^{n} u_{ij}}. \tag{16}$$

To reduce the number of iterations, the membership matrix, clustering centers, and the covariance matrix are initialized by the FCM algorithm. We summarize the procedure of the proposed SSFCA as follows:

(1) Set the number of clusters $c$, regularization parameter $\gamma$, convergence threshold $\eta$, and maximum iteration number $T$.
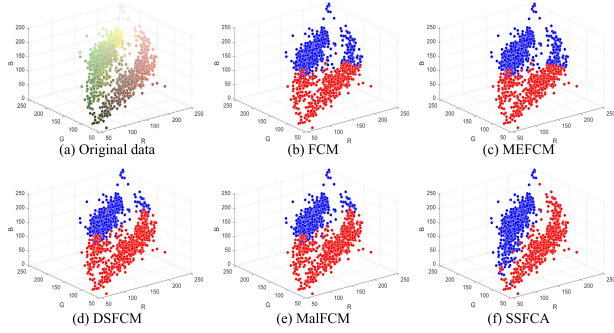
**FIGURE 2.** Results of different algorithms for image data with outliers.

(2) Initialize the membership $U^{(0)}$, the clustering centers $V^{(0)}$, and the covariance matrix $\Sigma^{(0)}$ using the FCM algorithm.

(3) Set the loop counter $t = 1$.

(4) Update $U^{(t)}$, $V^{(t)}$, $\Sigma^{(t)}$ using (14), (15), and (16), respectively.

(5) Update the objective function $\widetilde{J}'^{(t)}$ using (12).

(6) If $\max |\widetilde{J}'^{(t)} - \widetilde{J}'^{(t-1)})| \leq \eta$ or $t \geq T$, stop; otherwise, update $t = t + 1$, and go to step 4.

To demonstrate the effectiveness of SSFCA, we apply the SSFCA to data clustering with outliers. The test data is generated by sampling pixels at regular grid in an image "113044" from Berkeley Segmentation Dataset (BSDS500) [50]. Fig. 2 shows the comparison of clustering results provided by different algorithms. It can be easily seen that the SSFCA with $\gamma = 0.2$ achieves the best clustering result, which demonstrates that SSFCA is more robust for outliers than comparative algorithms based on regularization.

Fig. 3 further illustrates the effectiveness of the SSFCA on image segmentation using heatmap visualization. It is clear that MEFCM, DSFCM and SSFCA attract more attention on "horses" than FCM. Simultaneously, the SSFCA achieves the best result due to the Gaussian metric and self-sparse optimization. Note that we do not provide the result generated by MalFCM since the algorithm requires to construct an affinity membership matrix that is very large for the image "113044".

## B. OVER-SEGMENTATION REDUCTION

FCM uses pixel classification to achieve image segmentation, where each pixel of images is viewed as an independent sample. Therefore, FCM often causes the problem of over-segmentation, i.e., segmentation results include a large number of isolated small areas as shown in Fig. 4. It can be seen that DSFCM misclassifies more pixels than FCM and MEFCM because of inappropriate sparse deviation. Although the SSFCA partly suppresses the interference of non-homogenous pixels and obtains better visual effect, it still suffers from over-segmentation as shown in Fig. 4(e). Improved FCM algorithms based on the incorporation of local spatial information can alleviate over-segmentation by removing small areas, which is insufficient as shown in Fig. 5.

Although the mean-shift can effectively alleviate over-segmentation by eliminating small regions containing pixels less than $M$, the value of $M$ is often adjusted manually for different images. In this work, we propose a novel CCF-ADB to improve the SSFCA. Fig. 6 shows the framework of the CCF-ADB. It is clear that Fig. 6(b) includes many useless small regions that reduces the final segmentation accuracy. According to the proposed CCF-ADB, we firstly compute the area of all connected components in Fig. 6(b), and then sort these connected components in descending order as shown in Fig. 6(c). Because it is difficult to obtain the value of threshold $M$ depending on the sorting result, the ADB strategy is used to improve the sorting result as shown in Fig. 6(d). Based on the improvement, it is easy to obtain the maximum interval of Fig. 6(d). The maximal interval corresponds to a region whose area is considered as the value of $M$. Then we can eliminate small connected components as shown in Fig. 6(e) and (f) using the obtained $M$. Finally, Fig. 6(g) shows the segmentation result using the CCF-ADB.

To illustrate the strategy of ADB in details, let the $\chi_p$ denote the $p$th point, where $0 \leq \chi_p \leq 1$ and $1 \leq p \leq (K+1)$, $p \in N^+$. The $\alpha_q$ denotes the normalized area of the $q$th region, the $\xi_p$ denotes the number of $\alpha_q$ around $\chi_p$ under radius $\varepsilon$, which is presented as follows

$$\xi_p = \sum_{q=1}^{Q} \varphi_q, \qquad (17)$$

$$\varphi_q = \begin{cases} 1 & \|\alpha_q - \chi_p\| \leq \varepsilon \\ 0 & otherwise. \end{cases} \qquad (18)$$
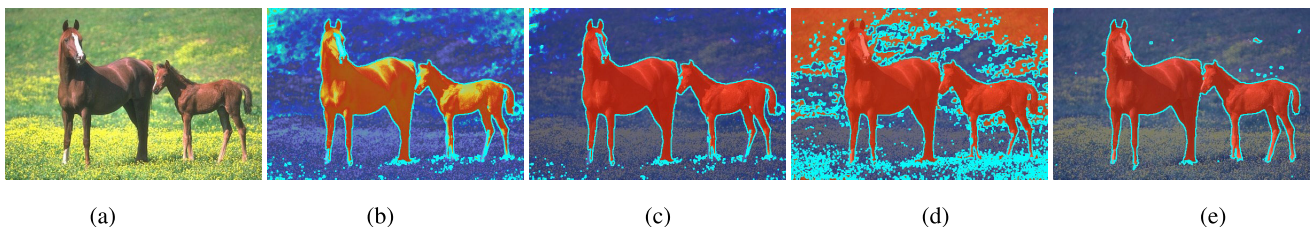


**FIGURE 3.** Comparison of heatmap visualization using different algorithms. (a) The original image. (b) FCM. (c) MEFCM. (d) DSFCM. (e) SSFCA.
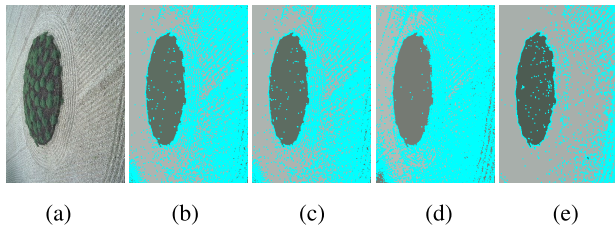
**FIGURE 4.** Comparison of segmentation results using different clustering algorithms with self-optimization. (a) Original image. (b) FCM. (c) MEFCM. (d) DSFCM. (e) SSFCA.
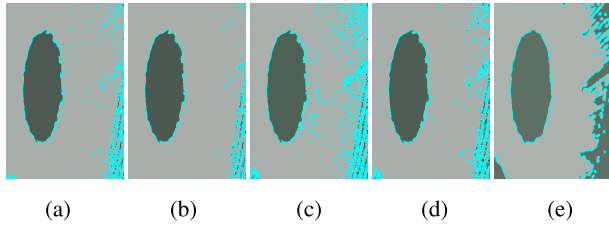


**FIGURE 5.** Comparison of segmentation results using different clustering algorithms with local spatial information. (a) FLICM. (b) KWFLICM. (c) NWFCM. (d) FRFCM. (e) DSFCM_N.

where $Q$ is the number of connected components in an image generated by the SSFCA. The $\kappa_q$ is the mapping result of $\alpha_q$, where $\kappa_q$ can be computed as follows:

$$\kappa_q = \sum_{\chi_p \leq \alpha_q} \xi_p. \tag{19}$$

We perform the normalization on $\kappa_q$ to obtain $\kappa_q'$. More details about the density balance algorithm can be seen in [51]. We present the detailed description of the CCF-ADB as follows:

**Input**: A connected-component image generated by the SSFCA.

Initialization: Normalized areas $\alpha_q$ for the input image, $K = 1000$, and $\varepsilon = 0.1$.

Step 1: Compute $\varphi_q$ using (18).
Step 2: Compute $\xi_p$ using (17).
Step 3: Compute $\kappa_q$ using (19).
Step 4: Implement the normalization on $\kappa_q$ to obtain $\kappa_q'$.
Step 5: Compute the value of cutoff area $M$ using the maximal interval in $\kappa_q'$.

Step 6: Merge regions whose areas are smaller than the value of $M$ using the minimum distance in connected-component images.

**Output**: A labeled image containing fewer regions than the input image.

According to the CCF-ADB, we compute decision-graph and improved decision-graph as shown in Fig. 7. By comparing Fig. 7(b) and (c), the ADB provides accurate maximum intervals and thus helps the CCF to improve segmentation results from the SSFCA.

## IV. EXPERIMENTS

To demonstrate the advantages of the proposed RSSFCA for image segmentation. We conducted experiments on synthetic images and two benchmark images. The first experiment demonstrates the superiority of RSSFCA on image with significant noise. The latter two experiments demonstrate the practicality and robustness of RSSFCA on different images from the Berkeley Segmentation Dataset (BSDS500) [50] and the Microsoft Research in Cambridge (MSRC) [52].

### A. PARAMETERS SETTING

In our experiments, PFCM, KWFLICM, NDFCM, FRFCM, Liu's algorithm, DSFCM_N, MSFCM and AWSFCM are considered as comparative algorithms. The fuzzy exponent is $m = 2$, convergence threshold is $\eta = 10^{-5}$, and maximum iteration number is $T = 50$. Most of comparative algorithms employ local spatial information except PFCM, MSFCM and RSSFCA, the size of neighboring window is $3 \times 3$. For PFCM, the relative importance of membership and typicality are both 1, and the exponent of typicality is selected as 2. In addition, both KWFLICM and MSFCM are free of other parameters. For the NDFCM, the values of spatial weighting factor, gray-level weight factor and controlled scale factor are set to 3, 5, and 3, respectively. Liu's algorithm uses the mean-shift to generate pre-segmentation results in which three parameters are necessary, i.e., spatial bandwidth $h_s = 10$, the range bandwidth $h_r = 10$, and the minimum output regions $M = 100$ following the original paper. We use default values of parameters for the FRFCM and DSFCM_N. For AWSFCM, we also referred to the paper and chose the number of $\alpha$-*planes* as 3, which can further reduce execution time. The proposed RSSFCA requires a regularization parameter; we set $\gamma = 0.2$. All test algorithms
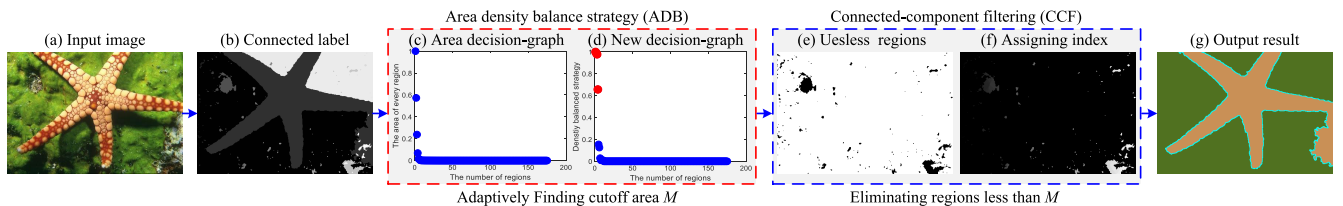


**FIGURE 6.** The framework of connected-component filtering based on adaptive density balance (CCF-ADB).
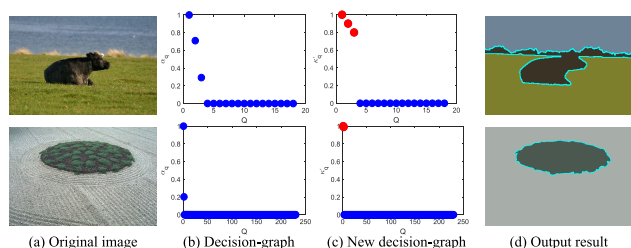
**FIGURE 7.** Improving segmentation results using the CCF-ADB.

are executed on a DELL desktop with Intel (R) Core (TM) CPU, i7-6700, 3.4GHz, 16GB RAM.

### B. SYNTHETIC IMAGE

To demonstrate the effect of the proposed RSSFCA on image segmentation. We firstly apply testing algorithms on a synthetic image corrupted by significant outliers. Fig. 8 shows comparative segmentation results. It can be seen that the synthetic image includes four clear objects and a lot of useless small objects as shown in Fig. 8(a). Our purpose is to segment four clear objects while suppressing useless small objects. Due to the fact that both PFCM and MSFCM ignore spatial information of images, their segmentation results contain more interference points as shown in Fig. 8(b) and (h). Focusing on Fig. 8(d), (f), and (i) generated by the NDFCM, FRFCM and AWSFCM, respectively, we find that these segmentation results still include many small objects that influence the final segmentation accuracy. Both KWFLICM and DSFCM_N provide misclassified results as shown in Fig. 8(c) and (g). Compared to

other algorithms, Liu's algorithm can achieve better noise suppression and the removal of useless small region as shown in Fig. 8(e), since it employs the mean-shift algorithm to obtain better adaptive neighboring spatial information.

Although Fig. 8 shows that better spatial information corresponds to better segmentation results, Liu's algorithm shows limited capability for the synthetic image with complex background. The proposed RSSFCA shows the best segmentation result since four objects are segmented accurately as shown in Fig. 8(j). In general, the RSSFCA not only suppresses noise effectively, but also merges useless small regions to obtain excellent segmentation result.

To estimate the performance of all testing algorithms on Fig. 8(a), we use two indices, i.e., the segmentation accuracy (SA) and the quantitative index score (S). The SA and S are computed as follows:

$$SA = \sum_{i=1}^{c} \frac{A_i \cap G_i}{n}, \qquad (20)$$

$$S = \sum_{i=1}^{c} \frac{A_i \cap G_i}{A_i \cup G_i}, \qquad (21)$$

where the $A_i$ represents a segmentation result, the $G_i$ denotes the corresponding Ground Truth, the $c$ is the number of clusters and $n$ is the total number of pixels of images. Ideal segmentation results correspond to high values of SA ( $= 1$) and S ( $= 1$). Table 1 shows values of SA and S for Fig. 8. In Table 1, both PFCM and MSFCM provide lower values of SA and S, which is consistent with the visual effect of Fig. 8(b) and (h). Both KWFLICM and DSFCM_N provide low values of S, which shows that these two algorithms cannot achieve accurate object segmentation for the synthetic
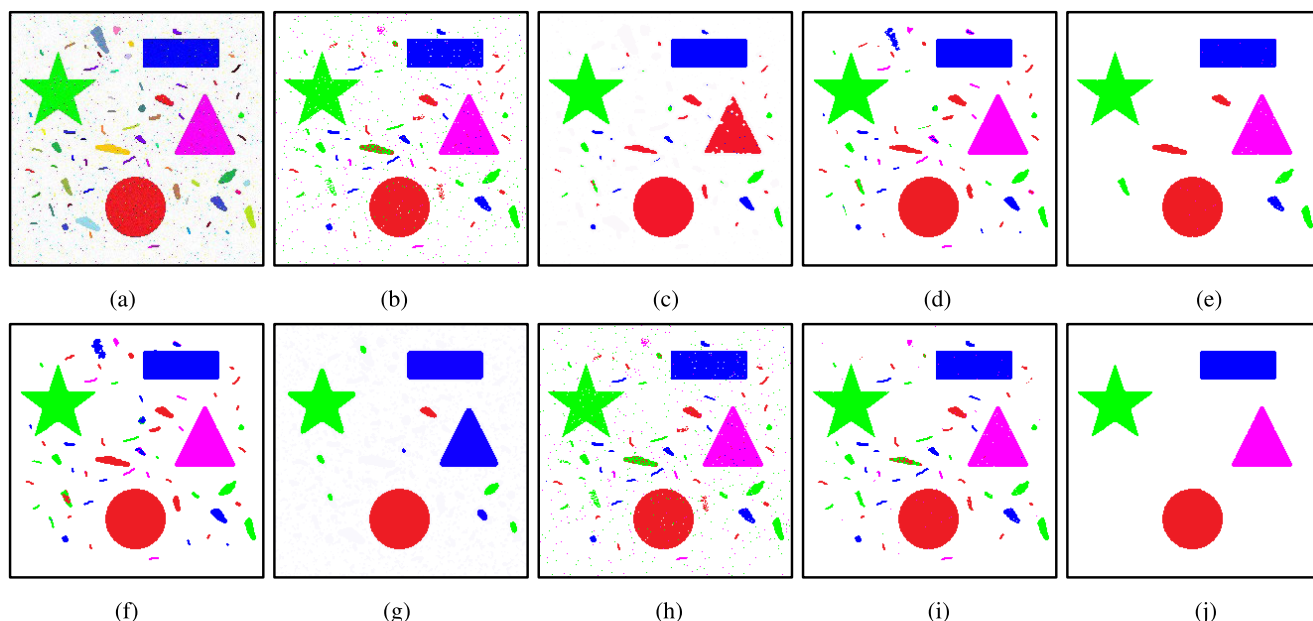


**FIGURE 8.** Comparison results of different algorithms on the synthetic image. The number of clusters is 5. (a) Testing image. (b) PFCM. (c) KWFLICM. (d) NDFCM. (e) Liu's algorithm. (f) FRFCM. (g) DSFCM_N. (h) MSFCM. (i) AWSFCM. (j) RSSFCA.

**TABLE 1.** Performance comparison of different algorithms on the synthetic image. The best values are in bold.

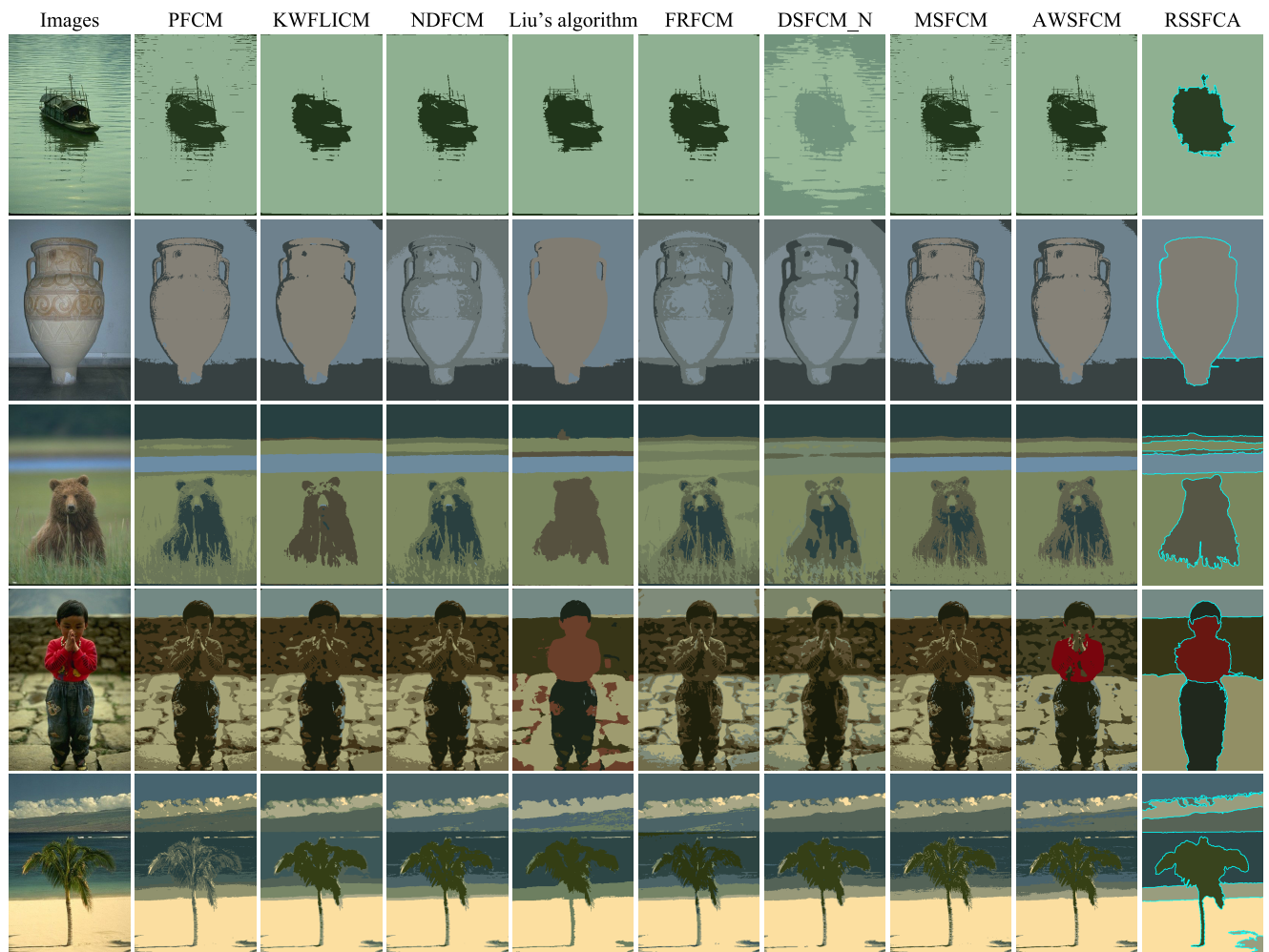| Indices | PFCM [17] | KWFLICM [33] | NDFCM [46] | Liu's algorithm [35] | FRFCM [47] | DSFCM_N [13] | MSFCM [20] | AWSFCM [34] | RSSFCA |
|---|---|---|---|---|---|---|---|---|---|
| SA | 0.9615 | 0.9521 | 0.9729 | 0.9857 | 0.9667 | 0.9588 | 0.9612 | 0.9735 | **0.9995** |
| S | 0.9258 | 0.5074 | 0.9472 | 0.9717 | 0.9356 | 0.5129 | 0.9252 | 0.9484 | **0.9991** |



**FIGURE 9.** Segmentation results using different algorithms on the BSDS500 (the first group of images).

image. The NDFCM, FRFCM and AWSFCM provide similar values of SA and S due to the employment of same neighboring windows. Liu's algorithm obtains higher values of SA and S than previous algorithms but lower values than the RSSFCA. Table 1 further demonstrates that the RSSFCA outperforms comparative algorithms since it integrates a regularization under Gaussian metric into its objective function and uses the CCF-ADB to optimize the final segmentation result.

### C. BENCHMARKS

In this section, we mainly validate the effectiveness of the proposed RSSFCA on BSDS500 [50] and MSRC [52]. The BSDS500 includes 500 natural images with size of $481 \times 321$

or $321 \times 481$, and each image corresponds to 4-9 manually generated ground truths with accurate pixel-wise labels. These ground truths are delineated manually by different human subjects. The MSRC collects 591 images with size of $320 \times 213$ or $213 \times 320$ and covers 23 object classes. In addition, all testing images have been transformed from color space RGB to CIELAB.

Figs. 9-10 show the comparison of segmentation results using different algorithms on BSDS500. According to Figs. 9-10, both PFCM and MSFCM generate poor segmentation results including too many small regions since they cannot overcome the sensitivity to intensity nonuniformity. Although both KWFLICM and DSFCM_N reduce the number of small regions by employing a neighboring window,

**FIGURE 10.** Segmentation results using different algorithms on the BSDS500 (the second group of images).

they easily cause mis-segmentation as shown in Figs. 9-10. The AWSFCM shows low robustness since it is only valid for the first test image as shown in Fig. 10. However, the utilization of local spatial information may cause the loss of image details, or even wrong clustering results for some images. Although both NDFCM and FRFCM obtain better segmentation results than KWFLICM, DSFCM_N and AWSFCM by using improved image filtering approaches, they still suffer from over-segmentation. Liu's algorithm further improves segmentation results by incorporating large region-level information into its objective function, which avoids the limitation of small and fixed neighboring windows. Compared to previous algorithms, the proposed RSSFCA generates better segmentation results due to its property of self-sparsity and the utilization of CCF-ADB.

Fig. 11 shows comparison of segmentation results on MSRC. It is clear that segmentation results in Fig. 11 show better visual effect than Figs. 9-10 due to simpler background of MSRC than that of BSDS500. As the same as Figs. 9-10, the proposed RSSFCA provides the best segmentation result, which further demonstrates the proposed RSSFCA outperforms comparative algorithms for image segmentation.

To evaluate segmentation performances of different algorithms, four performance measures [47], namely, probabilistic rand index (PRI), coving (CV), variation of information (VI), and global consistency error (GCE), are

used in our experiments. If a segmentation result is close to its corresponding Ground Truth, then the values of PRI and CV will be large while the values of VI and GCE will be small.

In our experiments, the number of clusters $c$ is set from 2 to 6 for each image in BSDS500, while it is set from 2 to 4 for each image in MSRC. We chose the group of measures that corresponds to the largest value of PRI as the final performance measures for each image. Tables 2-3 show the average values of PRI, CV, VI and GCE on all images in the BSDS500 or in MSRC, respectively. By comparing the values in Tables 2-3, PFCM provides low performance measures because it is sensitive to parameter tuning. Both Liu's algorithm and AWSFCM obtain better values of PRI, CV, VI, and GCE than those of other comparative algorithms. The proposed RSSFCA provides the best values of PRI, CV, VI, and GCE. By analyzing Figs. 9-11 together with Tables 2-3, the proposed RSSFCA achieves high-quality image segmentation on different benchmark images, which further demonstrates the effectiveness and robustness of RSSFCA.

### D. COMPUTATIONAL COMPLEXITY
Computational complexity of algorithms plays an important role in practical applications. Table 4 shows the comparison of computational complexity of different algorithms, where $n$ is the number of pixels of an image, $c$ is the number of clusters, $t$ is the number of iteration, $w$ is the size of local window, $l$ is the number of gray levels of image, and the value of $l$ is close to the value of $n$ for color images, $O(M(c))$ is the computational complexity of Newton's method for each iteration [49].

Table 4 shows the comparison of computational complexity of different algorithms. According to the Table 4, PFCM requires higher computational burden than MSFCM due to PFCM needing the parameter initialization by using FCM with $t'$ iterations. The KWFLICM, DSFCM_N, and AWSFCM require high computational complexity since these algorithms compute local spatial information of images in each iteration. In addition, $K'$ denotes the number of $\alpha$-*planes* for AWSFCM. Different from the previous five algorithms, as both NDFCM and FRFCM only compute the local spatial information of images once in the process of image segmentation, they have low computational complexity. Because the SSFCA involves two computational steps, its computational complexity is composed of $O(nct')$ and $O(n(M(c) + c)t)$. The RSSFCA includes SSFCA and CCF-ABD. As the time complexity of SSFCA is much larger than CCF-ABD, the time complexity of RSSFCA is close to the SSFCA. Therefore, RSSFCA avoids the heavily computational complexity because of linear iterative scheme.

### E. DISCUSSION
In this section, we mainly discuss the influence of regularization parameter $\gamma$ for the sparsity of membership matrices. Fig. 12 shows the comparison of membership matrices by tuning the value of $\gamma$ on the image "12003", where the number of clusters is $c = 2$. We obtain the sparsest membership
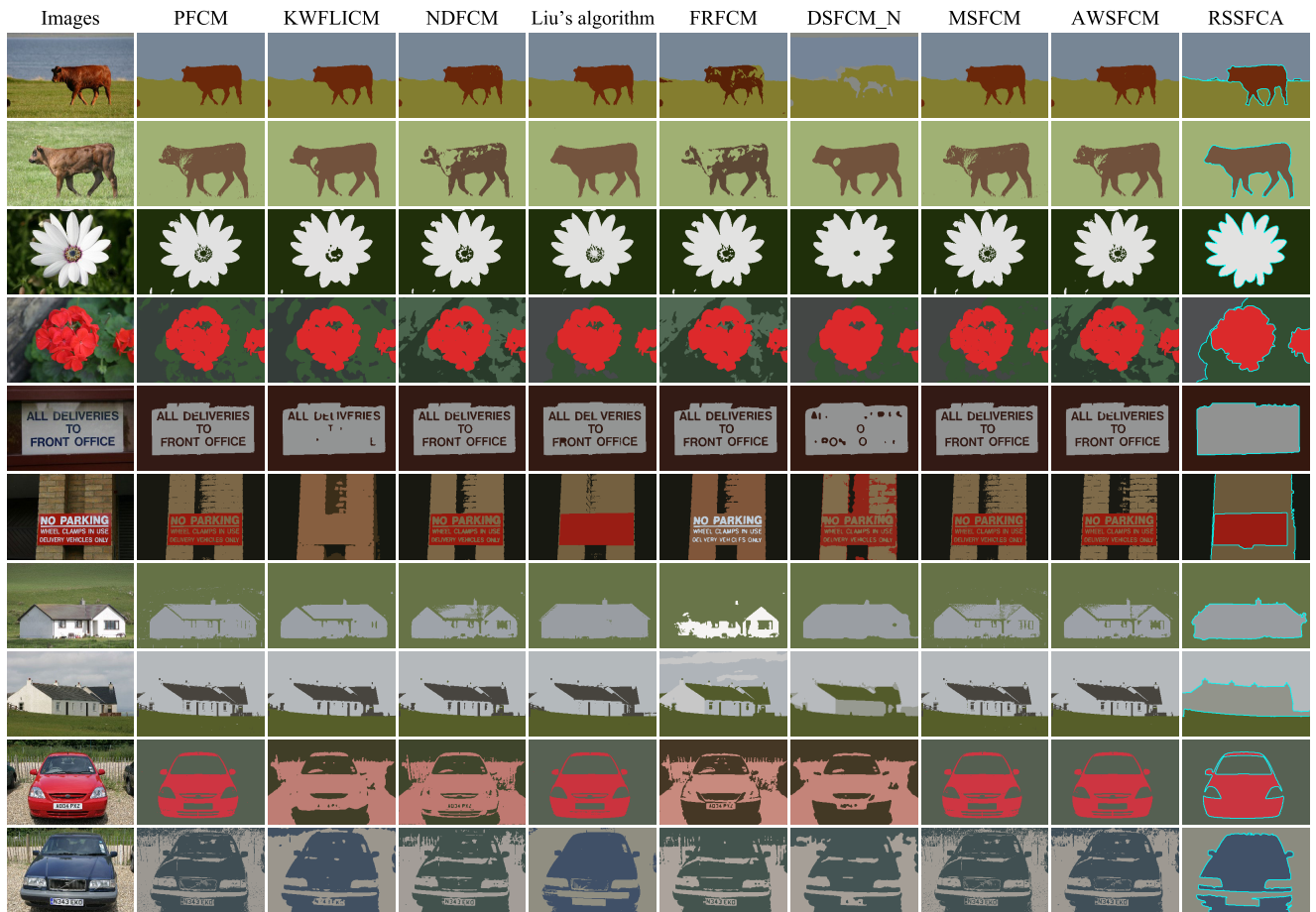
**FIGURE 11.** Segmentation results using different algorithms on the MSRC.

**TABLE 2.** Performance comparison of different algorithms on the BSDS500. The best values are in bold.

| Algorithms | PRI↑ | CV↑ | VI↓ | GCE↓ |
|---|---|---|---|---|
| PFCM [17] | 0.72 | 0.41 | 2.97 | 0.42 |
| KWFLICM [33] | 0.74 | 0.44 | 2.83 | 0.40 |
| NDFCM [46] | 0.75 | 0.44 | 2.78 | 0.39 |
| Liu's algorithm [35] | 0.76 | 0.47 | 2.58 | 0.36 |
| FRFCM [47] | 0.76 | 0.45 | 2.67 | 0.37 |
| DSFCM_N [13] | 0.74 | 0.42 | 2.90 | 0.41 |
| MSFCM [20] | 0.74 | 0.43 | 2.85 | 0.40 |
| AWSFCM [34] | 0.75 | 0.45 | 2.74 | 0.38 |
| RSFFCA | **0.78** | **0.52** | **2.12** | **0.28** |

**TABLE 3.** Performance comparison of different algorithms on the MSRC. The best values are in bold.

| Algorithms | PRI↑ | CV↑ | VI↓ | GCE↓ |
|---|---|---|---|---|
| PFCM [17] | 0.67 | 0.55 | 1.84 | 0.31 |
| KWFLICM [33] | 0.69 | 0.55 | 1.93 | 0.32 |
| NDFCM [46] | 0.69 | 0.55 | 1.90 | 0.32 |
| Liu's algorithm [35] | 0.71 | 0.54 | 1.77 | 0.34 |
| FRFCM [47] | 0.71 | 0.58 | 1.79 | 0.30 |
| DSFCM_N [13] | 0.69 | 0.54 | 1.91 | 0.32 |
| MSFCM [20] | 0.68 | 0.57 | 1.80 | 0.30 |
| AWSFCM [34] | 0.70 | 0.59 | 1.74 | 0.29 |
| RSFFCA | **0.75** | **0.64** | **1.51** | **0.24** |

matrix that is close to $k$-means when the value of $\gamma$ is close to zero. With the increase of the value of $\gamma$, more values of membership are close to 0.5. All values of membership equal to 0.5 when the value of $\gamma$ equals to 1.

To discuss the effectiveness of the parameter $\gamma$, two validity functions [53], i.e., the partition coefficient denoted by ($V_{pc}$) and the partition entropy denoted by ($V_{pe}$) are used for performance evaluation. The $V_{pc}$ and $V_{pe}$ are defined as:

$$V_{pc} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2}{n}, \qquad (22)$$

$$V_{pe} = -\frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} ln(u_{ij})}{n}, \qquad (23)$$

where both $V_{pc}$ and $V_{pe}$ are intuitive indicators used to measure the sparsity strength of membership matrices. If the fuzziness of membership matrices is weaker, the value of the $V_{pc}$ will be larger while the value of the $V_{pe}$ will be smaller, and vice versa. Fig. 13 shows the change of validity functions by varying the value of $\gamma$, which shows the performance change of the SSFCA. In Fig. 13, it is observed that both
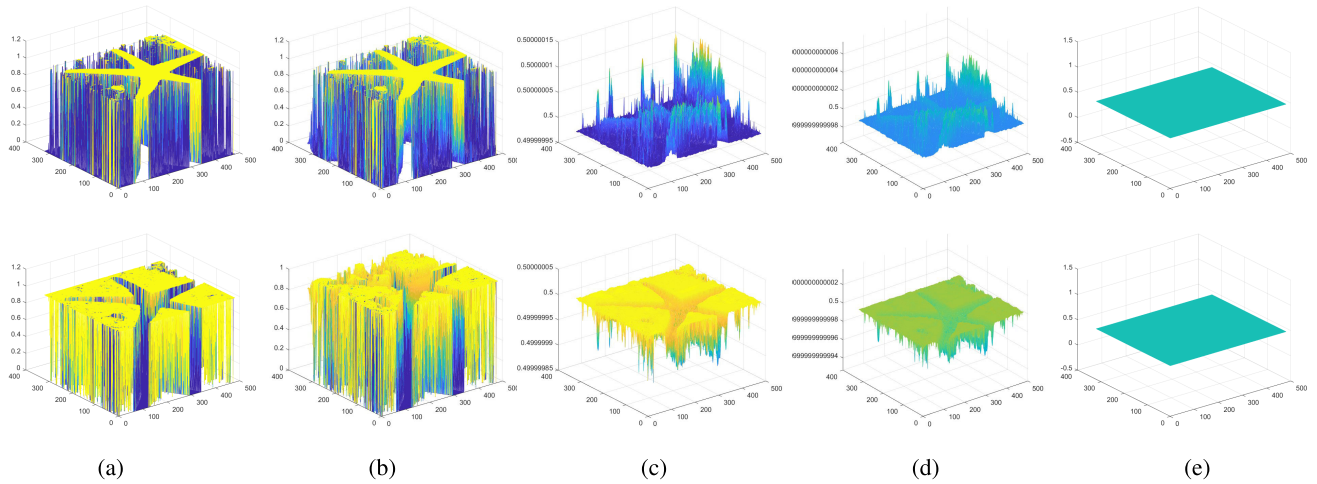
**FIGURE 12.** Comparison of membership matrices by tuning the value of $\gamma$ on the image "12003". Note that the membership matrix tends to sparse by decreasing the value of $\gamma$. (a) $\gamma = 0.2$. (b) $\gamma = 0.4$. (c) $\gamma = 0.6$. (d) $\gamma = 0.8$. (e) $\gamma = 1.0$.

**TABLE 4.** Computational complexity of different methods.

| Algorithms | Computational complexity |
|---|---|
| PFCM [17] | $O(nct' + nct)$ |
| KWFLICM [33] | $O(n(w+1)^2 + nw^2ct)$ |
| NDFCM [46] | $O(nw^2 + lct)$ |
| FRFCM [47] | $O(nw^2 + lct)$ |
| DSFCM_N [13] | $O(nw^2ct)$ |
| MSFCM [20] | $O(nct)$ |
| AWSFCM [34] | $O(K'(nw^2 + 2nc)t)$ |
| RSSFCA | $O(nct' + n(M(c) + c)t)$ |
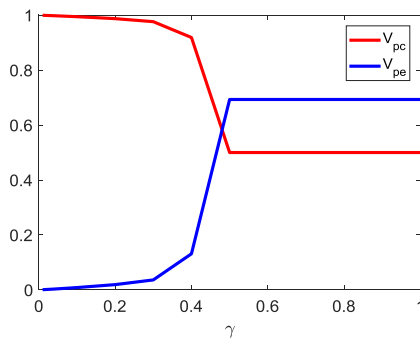


**FIGURE 13.** The change of validity functions by tuning the value of $\gamma$.

$V_{pe}$ and $V_{pc}$ are constant when $\gamma > 0.5$ because all values of membership are non-sparse and they are equivalent. The parameter $\gamma$ can control the degree of sparsity only when $\gamma \leq 0.5$. Therefore, we experientially set the value of $\gamma$ to 0.2.

## V. CONCLUSION

In this work, we have proposed a novel self-sparse fuzzy c-means clustering algorithm (RSSFCA) for image segmentation. The proposed RSSFCA has been used to address two drawbacks in current image segmentation algorithms based on fuzzy clustering. On the one hand, the RSSFCA incorporates a regularization term to its objective function to balance the sparsity of membership and fuzziness, and it thus achieves self-sparse fuzzy clustering. On the other hand, the RSSFCA employs the CCF-ADB to achieve effectively small region merging, which leads to excellent image segmentation results. Experiments on synthetic images and benchmark images demonstrate the proposed RSSFCA outperforms state-of-the-art algorithms in terms of better segmentation results and higher values of performance indices. Furthermore, we analyzed the influence of regularization parameter values on clustering results, which shows that the proposed RSSFCA can provide good clustering results when the value of $\gamma$ is less than 0.5.

## REFERENCES

[1] I. Gath and A. B. Geva, "Unsupervied optimal fuzzy clsutering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 773–780, Jul. 1989.

[2] T. Schultz and G. L. Kindlmann, "Open-box spectral clustering: Applications to medical image analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2100–2108, Dec. 2013.

[3] B. Tu, X. Zhang, X. Kang, J. Wang, and J. A. Benediktsson, "Spatial density peak clustering for hyperspectral image classification with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5085–5097, Jul. 2019.

[4] J. Warren, J. Lipkowitz, and V. Sokolov, "Clusters of driving behavior from observational smartphone data," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 171–180, Jun. 2019.

[5] C.-F. Juang, C.-M. Hsiao, and C.-H. Hsu, "Hierarchical cluster-based multispecies particle-swarm optimization for fuzzy-system optimization," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 1, pp. 14–26, Feb. 2010.

[6] M. Tang, D. Marin, I. B. Ayed, and Y. Boykov, "Kernel cuts: Kernel and spectral clustering meet regularization," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 447–511, Sep. 2019.

[7] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, May 1984.

[8] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.

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

[10] A. Abu and R. Diamant, "Enhanced fuzzy-based local information algorithm for sonar image segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 445–460, 2020.

[11] H.-J. Xing and M.-H. Ha, "Further improvements in feature-weighted fuzzy C-means," *Inf. Sci.*, vol. 267, pp. 1–15, May 2014.

[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] Y. Zhang, X. Bai, R. Fan, and Z. Wang, "Deviation-sparse fuzzy C-means with neighbor information constraint," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 185–199, Jan. 2019.

[14] D. Hallac, J. Leskovec, and S. Boyd, "Network lasso: Clustering and optimization in large graphs," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, 2015, pp. 387–396.

[15] L. Guo, L. Chen, X. Lu, and C. L. P. Chen, "Membership affinity lasso for fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 2, pp. 294–307, Feb. 2020.

[16] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Mar. 2015.

[17] 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.

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

[19] X. Bai, Y. Zhang, H. Liu, and Z. Chen, "Similarity measure-based possibilistic FCM with label information for brain MRI segmentation," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2618–2630, Jul. 2019.

[20] S. Zhou, D. Li, Z. Zhang, and R. Ping, "A new membership scaling fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, early access, May 12, 2020, doi: 10.1109/TFUZZ.2020.3003441.

[21] O. Linda and M. Manic, "General type-2 fuzzy C-means algorithm for uncertain fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 5, pp. 883–897, Oct. 2012.

[22] C. Hwang and F. C.-H. Rhee, "Uncertain fuzzy clustering: Interval type-2 fuzzy approach to C-means," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 107–120, Feb. 2007.

[23] S. Miyamoto and M. Mukaidono, "Fuzzy C-means as a regularization and maximum entropy approach," in *Proc. 7th Int. Fuzzy Syst. Assoc. World Congr.*, Prague, Czech Republic, 1997, pp. 86–92.

[24] M. Zarinbal, M. H. Fazel Zarandi, and I. B. Turksen, "Relative entropy fuzzy C-means clustering," *Inf. Sci.*, vol. 260, pp. 74–97, Apr. 2014.

[25] D. Kumar, R. K. Agrawal, and H. Verma, "Kernel intuitionistic fuzzy entropy clustering for MRI image segmentation," *Soft Comput.*, vol. 24, no. 6, pp. 4003–4026, Mar. 2020.

[26] S. D. Xenaki, K. D. Koutroumbas, and A. A. Rontogiannis, "Sparsity-aware possibilistic clustering algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1611–1626, Dec. 2016.

[27] K. D. Koutroumbas, S. D. Xenaki, and A. A. Rontogiannis, "On the convergence of the sparse possibilistic C-means algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 324–337, Feb. 2018.

[28] J. Xu, J. Han, K. Xiong, and F. Nie, "Robust and sparse fuzzy K-means clustering," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, New York, NY, USA, 2016, pp. 2224–2230.

[29] 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.

[30] 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.

[31] 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.

[32] Z. Zaixin, C. Lizhi, and C. Guangquan, "Neighbourhood weighted fuzzy C-means clustering algorithm for image segmentation," *IET Image Process.*, vol. 8, no. 3, pp. 150–161, Mar. 2014.

[33] 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.

[34] P. K. Mishro, S. Agrawal, R. Panda, and A. Abraham, "A novel type-2 fuzzy C-means clustering for brain MR image segmentation," *IEEE Trans. Cybern.*, early access, May 12, 2020, doi: 10.1109/TCYB.2020.2994235.

[35] G. Liu, Y. Zhang, and A. Wang, "Incorporating adaptive local information into fuzzy clustering for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3990–4000, Nov. 2015.

[36] J. Gu, L. Jiao, S. Yang, and F. Liu, "Fuzzy double C-means clustering based on sparse self-representation," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 612–626, Apr. 2018.

[37] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.

[38] T. Lei, P. Liu, X. Jia, X. Zhang, H. Meng, and A. K. Nandi, "Automatic fuzzy clustering framework for image segmentation," *IEEE Trans. Fuzzy Syst.*, early access, Jul. 23, 2019, doi: 10.1109/TFUZZ.2019.2930030.

[39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[40] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

[41] T. Lei, X. Jia, T. Liu, S. Liu, H. Meng, and A. K. Nandi, "Adaptive morphological reconstruction for seeded image segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5510–5523, Nov. 2019.

[42] L. Szilagyi, Z. Benyo, S. M. Szilagyii, and H. S. Adam, "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, Sep. 2003, pp. 724–726.

[43] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy C-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.

[44] F. Zhao, J. Fan, and H. Liu, "Optimal-selection-based suppressed fuzzy C-means clustering algorithm with self-tuning non local spatial information for image segmentation," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4083–4093, Jul. 2014.

[45] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *Int. J. Comput. Vis.*, vol. 76, no. 2, pp. 123–139, Feb. 2008.

[46] F.-F. Guo, J. Shen, and X.-X. Wang, "Adaptive fuzzy C-means algorithm based on local noise detecting for image segmentation," *IET Image Process.*, vol. 10, no. 4, pp. 272–279, Apr. 2016.

[47] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy C-means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027–3041, Oct. 2018.

[48] M. V. W. Zibetti, E. S. Helou, and D. R. Pipa, "Accelerating overrelaxed and monotone fast iterative shrinkage-thresholding algorithms with line search for sparse reconstructions," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3569–3578, Jul. 2017.

[49] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, Buenos Aires, Argentina, 2015, pp. 3569–3575.

[50] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[51] Y. Zhu, K. M. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recognit.*, vol. 60, pp. 983–997, Dec. 2016.

[52] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Graz, Austria, 2006, pp. 1–15.

[53] S. K. Adhikari, J. K. Sing, D. K. Basu, and M. Nasipuri, "Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images," *Appl. Soft Comput.*, vol. 34, pp. 758–769, Jun. 2015.

**XIAOHONG JIA** received the M.S. degree in signal and information processing from Lanzhou Jiaotong University, Lanzhou, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an, China.

His current research interests include image processing and pattern recognition.
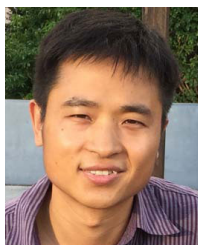
**TAO LEI** (Senior Member, IEEE) received the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2011.

From 2012 to 2014, he was a Postdoctoral Research Fellow with the School of Electronics and Information, Northwestern Polytechnical University. From 2015 to 2016, he was a Visiting Scholar with the Quantum Computation and Intelligent Systems Group, University of Technology Sydney, Sydney, NSW, Australia. He is currently a Professor with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology. He has authored or coauthored more than 80 research articles, including the IEEE Transactions on Image Processing, the IEEE Transactions on Fuzzy Systems, the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Transactions on Geoscience and Remote Sensing Letters, ICASSP, ICIP, and FG. His current research interests include image processing, pattern recognition, and machine learning.

**XIAOGANG DU** received the B.S., M.S., and Ph.D. degrees in computer science and technology from Lanzhou Jiaotong University, Lanzhou, China, in 2007, 2010, and 2017, respectively.

He is currently an Associate Professor with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, China. His research interests include image processing, computer vision, and machine learning.

**SHIGANG LIU** received the B.S. and M.S. degrees from Harbin Engineering University, Harbin, China, in 1997 and 2001, respectively, and the Ph.D. degree from Xidian University, Xi'an, China, in 2005.

From 2009 to 2010, he was a Visiting Scholar with the Department of Computing, The Hong Kong Polytechnic University. From 2015 to 2016, he was a Visiting Scholar with the Quantum Computation and Intelligent System Group, University of Technology Sydney, Sydney, NSW, Australia. He is currently a Professor with the School of Computer Science, Shaanxi Normal University, Xi'an. His research interests include pattern recognition and image processing.

**HONGYING MENG** (Senior Member, IEEE) received the Ph.D. degree in communication and electronic systems from Xi'an Jiaotong University, Xi'an, China.

He is currently a Senior Lecturer with the Department of Electronic and Computer Engineering, Brunel University London, U.K. He has authored over 130 publications, including the IEEE Transactions on Image Processing, the IEEE Transactions on Cybernetics, the IEEE Transactions on Fuzzy Systems, the IEEE Transactions on Automatic Control, the IEEE Transactions on Circuits and Systems for Video Technology, the IEEE Transactions on Biomedical Engineering, the IEEE Transactions on Cognitive and Developmental Systems, ICASSP, and CVPR. His research interests include digital signal processing, affective computing, machine learning, human–computer interaction, and computer vision. He is a Fellow of The Higher Education Academy (HEA) and a member of Engineering Professors Council, U.K. He is also an Associate Editor of the IEEE Transactions on Circuits and Systems for Videos Technology (TCSVT) and the IEEE Transactions on Cognitive and Developmental Systems (TCDS).

**ASOKE K. NANDI** (Fellow, IEEE) received the Ph.D. degree in physics from the University of Cambridge (Trinity College), Cambridge, U.K.

He held academic positions in several universities, including Oxford, U.K., Imperial College London, U.K., Strathclyde, U.K., and Liverpool, U.K., as well as a Finland Distinguished Professorship in Jyvaskyla, Finland. He is currently a Professor with the Department of Electronic and Computer Engineering, Brunel University London, U.K. He is a Distinguished Visiting Professor with Tongji University, China, and an Adjunct Professor with the University of Calgary, Canada. He has made many fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in big data, dealing with estimation, prediction, clustering, classification, and so on. He co-discovered the three fundamental particles known as $W^+$, $W^-$ and $Z^0$ (by the UA1 team at CERN), in 1983, providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics, in 1984, awarded the prize to his two team leaders for their decisive contributions. He has authored over 600 technical publications, including 250 journal articles as well as five books, entitled *Condition Monitoring with Vibrator Signals: Compressive Sampling and Learning Algorithms for Rotating Machines* (Wiley, 2019), *Automatic Modulation Classification: Principles, Algorithms and Applications* (Wiley, 2015), *Integrative Cluster Analysis in Bioinformatics* (Wiley, 2015), *Blind Estimation Using Higher-Order Statistics* (Springer, 1999), and *Automatic Modulation Recognition of Communications Signals* (Springer, 1996). The H-index of his publications is 77 (Google Scholar) and ERDOS number is two. His current research interests include signal processing and machine learning, with applications to communications, gene expression data, functional magnetic resonance data, and biomedical data.

Dr. Nandi is a Fellow of the Royal Academy of Engineering, U.K., and seven other institutions, including IET. He received the Institute of Electrical and Electronics Engineers (USA) Heinrich Hertz Award, in 2012, the Glory of Bengal Award for his outstanding achievements in scientific research, in 2010, the Water Arbitration Prize of the Institution of Mechanical Engineers, U.K., in 1999, and the Mountbatten Premium, Division Award of the Electronics and Communications Division, Institution of Electrical Engineers, U.K., in 1998. From 2018 to 2019, he was the IEEE EMBS Distinguished Lecturer.

. . .