

# Residual-sparse Fuzzy C-Means Clustering Incorporating Morphological Reconstruction and Wavelet Frame

Cong Wang, Witold Pedrycz, *Fellow, IEEE*, ZhiWu Li, *Fellow, IEEE*, MengChu Zhou, *Fellow, IEEE*, and Jun Zhao, *Member, IEEE*

**Abstract**—In this paper, we develop a residual-sparse Fuzzy C-Means (FCM) algorithm for image segmentation, which furthers FCM's robustness by realizing the favorable estimation of the residual (e.g., unknown noise) between an observed image and its ideal version (noise-free image). To achieve a sound trade-off between detail preservation and noise suppression, morphological reconstruction is used to filter the observed image. By combining the observed and filtered images, a weighted sum image is generated. Tight wavelet frame decomposition is used to transform the weighted sum image into its corresponding feature set. Taking such feature set as data for clustering, we impose an  $\ell_0$  regularization term on residual to FCM's objective function, thus resulting in residual-sparse FCM, where spatial information is introduced for improving its robustness and making residual estimation more reliable. To further enhance segmentation accuracy of the proposed FCM, we employ morphological reconstruction to smoothen the labels generated by clustering. Finally, based on the prototypes and smoothed labels, a segmented image is reconstructed by using tight wavelet frame reconstruction. Experimental results reported for synthetic, medical, and real-world images show that the proposed algorithm is effective and efficient, and outperforms its peers.

**Index Terms**—Fuzzy C-Means; residual-sparse; wavelet frame; morphological reconstruction; image segmentation.

## I. INTRODUCTION

SINCE its inception, a Fuzzy C-Means (FCM) algorithm [1] has achieved much attention, and been applied to a wide range of research fields, such as granular computing [2],

pattern recognition [3] and image analysis [4]. However, the conventional FCM exhibits a substantial flaw as it is not robust to observed (noisy) images. To improve its robustness, its modified versions have been put forward by mainly introducing spatial information into its objective function [5]–[9] and substituting the Euclidean distance by kernel distances (functions) [10]–[17]. As the first improvement, some classic FCM-related algorithms, such as FCM\_S [5], FCM\_S1 [6], FCM\_S2 [6], EnFCM [7] and FGFCM [8], have been proposed. Especially, Krinidis et al. [9] report a fuzzy local information C-means algorithm (FLICM) with the assistance of a fuzzy factor, which brings a simplified parameter setting. It yields better segmentation performance than previous algorithms. Nevertheless, only non-robust Euclidean distance is adopted in it, which is not effective for coping with the spatial information of images. In order to enhance its robustness, the second improvement has been investigated by using kernel distances. The essence of kernel distances (functions) is to transform the original data space into a new one. By making full use of superior properties of the new space, image data can be analyzed and manipulated easily. As a result, the use of kernel distances gives rise to such well-known FCM-related algorithms as KWFLICM [10], ARKFCM [11], KGFCM [12], NDFCM [13] and NWFCM [14]. In particular, Wang et al. [15] propose a wavelet frame-based FCM algorithm (WFCM) for addressing image segmentation problems defined in regular Euclidean and irregular domains. By taking tight wavelet frames as a kernel function, image data characteristics are fully analyzed.

Recently, some comprehensive FCM algorithms have been presented [18]–[21], which involve various techniques. For instance, Gharieb et al. [18] introduce a FCM framework by using Kullback-Leibler divergence to control the membership distance between a pixel and its neighbors. However, their algorithm is time-consuming and its segmentation effects can be further improved. Lei et al. [19] report a new algorithm, namely FRFCM, by augmenting morphological reconstruction and membership filtering. It is fast thanks to using gray level histograms. However, its performance is sometimes unstable. More recently, Gu et al. [20] introduce a Fuzzy Double C-Means algorithm (FDCM) incorporating sparse representation. It deals with two datasets simultaneously. The one is the basic feature set coming from an observed image, and the other is the feature set learned from a sparse self-representation model. Although FDCM is robust to noise, its computational efficiency is low.

This work was supported in part by the Doctoral Students' Short Term Study Abroad Scholarship Fund of Xidian University, in part by the National Natural Science Foundation of China under Grant Nos. 61873342, 61672400, 62076189, in part by the Recruitment Program of Global Experts, and in part by the Science and Technology Development Fund, MSAR, under Grant No. 0012/2019/A1. (*Corresponding author: ZhiWu Li and MengChu Zhou.*)

C. Wang is with the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China (e-mail: wangc0705@stu.xidian.edu.cn).

W. Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China, and also with the Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: wpedrycz@ualberta.ca).

Z. Li is with the School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China, and also with the Institute of Systems Engineering, Macau University of Science and Technology, Macau, China (e-mail: zhwli@xidian.edu.cn).

M. Zhou is with the Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China and also with the Helen and John C. Hartmann Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: zhou@njit.edu).

J. Zhao is with School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: junzhao@ntu.edu.sg).

In fact, there usually exist some outliers, noise or intensity inhomogeneity in an observed image, which are produced by interferences during image acquisition and transmission. Such objects are usually modeled as the residual between the observed image and its ideal value (e.g. noise-free image). Intuitively, using the ideal value may benefit segmentation effects produced by the FCM. By briefly reviewing the literature, it has been found that most of the existing FCM-related algorithms utilize spatial information in observed images to suppress the residual present in fuzzy clustering. Yet no studies focus on in-depth analysis and transforming of data components to develop an FCM algorithm. In other words, the residual is often not introduced into the objective function of FCM by separating it from the observed image. Moreover, a large proportion of image data has small or zero number of outliers, noise or intensity inhomogeneity. Therefore, the residual is very sparse in some transformed domain. To take the sparsity of the residual into consideration, Zhang et al. [22] attempt to impose  $\ell_1$  regularization on the residual, thus resulting in two alternative clustering algorithms, namely DSFCM and DSFCM\_N. Once spatial information has been encountered, DSFCM upgrades to DSFCM\_N that makes full use of spatial information. However, both algorithms fail to fully analyze the sparsity in the residual. As a result, their segmentation performance remains to be improved.

Motivated by [22], we propose an  $\ell_0$  regularization-based FCM algorithm incorporating a morphological reconstruction (MR) operation [23], [24] and a tight wavelet frame transform [25], [26]. The framework of the proposed algorithm is given in Fig. 1. It consists of the following steps:

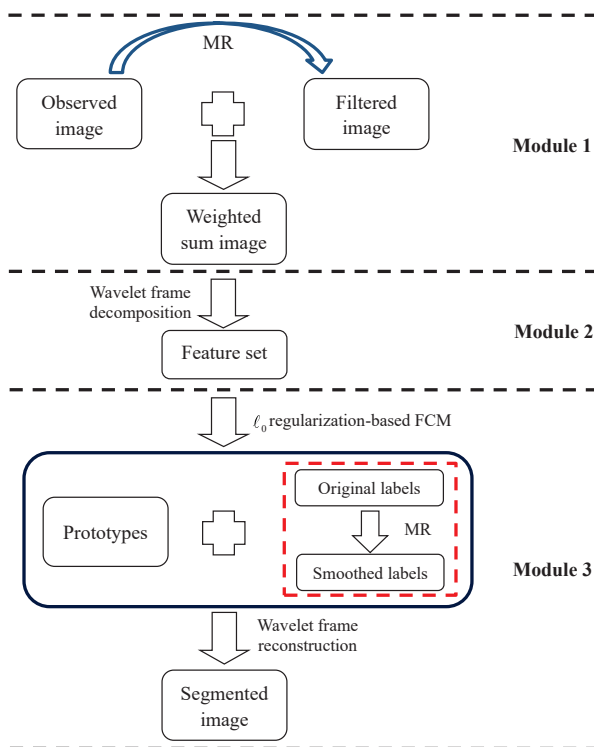


Fig. 1. The framework of the proposed algorithm.

As shown in Fig. 1, to achieve a good trade-off between

detail preservation and noise suppression, MR is used to filter an observed image. By combining the observed and filtered images, a weighted sum image is generated, which contains less noise than the observed image but more features than the filtered image. To acquire sparse representations of the weighted sum image, a tight wavelet frame system is employed to decompose an image, thus resulting in the formation of its feature set. Considering the feature set as data to be clustered, we present an improved FCM algorithm by imposing an  $\ell_0$  regularization term on residual between the feature set and its ideal value, which implies that the ideal value estimated from the observed value participates in clustering in the light of the sparsity of the residual. Moreover, due to its capacity of noise suppression, the spatial information of image pixels is also considered into the objective function of FCM since it is naturally encountered in image segmentation. To further enhance the segmentation performance of the improved algorithm, MR is employed to smoothen the generated labels in clustering. Finally, by combining the prototypes obtained by the improved FCM algorithm and the smoothed labels, a segmented image is reconstructed by using a tight wavelet frame reconstruction operation.

In conclusion, this study makes fourfold contributions to improve FCM's performance for image segmentation:

- As a pre-processing step, it uses MR to filter an observed image. The distribution characteristic of image pixels is made more favorable to fuzzy clustering. The speed of clustering is also improved.
- It employs tight wavelet frames to form a feature space, which overcomes the drawback of the direct use of image pixels. In this space, underlying image details can be fully analyzed and manipulated. In essence, this study proposes a novel kernel-based FCM algorithm by taking tight wavelet frames as a kernel function.
- It proposes residual-sparse FCM by introducing an  $\ell_0$  regularization term on residual into FCM. The proposed FCM makes the residual accurately-estimated by using its sparsity. To guarantee the close relationship between a target pixel and its neighbors, spatial information is also considered into the proposed FCM.
- We complete a label smoothing step by prudently using MR. Consequently, FCM's performance is further improved.

Overall, the originality of this work is to propose a comprehensive FCM algorithm by precisely estimating the residual and the assistance of various techniques. The essence of the proposed algorithm is a kernel-based FCM method with the aid of tight wavelet frames, which has better ability to identify features and noise in images. As a pre-processing step, MR removes a large portion of noise and preserves main features, thus making a weighted sum image more favorable to clustering. Moreover, since  $\ell_0$  regularization exhibits good properties related to the sparsity of the residual, the estimation of the residual is precisely realized. That is to say, the proposed algorithm conducts image segmentation by indirectly using the ideal value of an observed image. It is clear that  $\ell_0$  regularization exhibits the greatest impact on

FCM's robustness improvement. However, it requires more iterations. Morphological reconstruction and wavelet frames further improve segmentation performance to some extent. Additionally, they make the number of FCM's iterations greatly decreased. Although they are not as influential as the  $\ell_0$  regularization itself, they are indispensable for improving the FCM's segmentation performance.

This paper is organized as follows. The FCM algorithm and a tight wavelet frame transform are briefly recalled in Section II. Section III details the proposed algorithm. Section IV reports experimental results for a set of images. Conclusions are drawn in Section V.

## II. PRELIMINARIES

### A. Fuzzy C-Means (FCM) Algorithm

Given are data  $\mathbf{X} = \{\mathbf{x}_j : j = 1, 2, \dots, K\} \subset \mathbb{R}^K$ . Since each object  $\mathbf{x}_j$  has  $L$  attributes (channels), it is an  $L$ -dimensional vector  $(x_{j1}, x_{j2}, \dots, x_{jL})^T$ . FCM splits  $\mathbf{X}$  into  $c$  subsets by minimizing the following objective function

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|^2, \quad (1)$$

subject to

$$\sum_{i=1}^c u_{ij} = 1, \quad \forall j \in \{1, 2, \dots, K\},$$

where  $\mathbf{U} = [u_{ij}]_{c \times K}$  with  $0 \leq u_{ij} \leq 1$  is a partition matrix,  $\mathbf{V} = \{\mathbf{v}_i\}_{i=1,2,\dots,c}$  is a set of  $c$  prototypes,  $\|\cdot\|$  denotes the Euclidean distance, and  $m$  is a fuzzification exponent ( $m > 1$ ).

The FCM algorithm is an alternating iteration scheme aimed to minimize (1). Each iteration can be realized as follows [27]:

$$u_{ij}^{(t+1)} = \frac{(\|\mathbf{x}_j - \mathbf{v}_i^{(t)}\|^2)^{-\frac{1}{m-1}}}{\sum_{q=1}^c (\|\mathbf{x}_j - \mathbf{v}_q^{(t)}\|^2)^{-\frac{1}{m-1}}}, \quad v_{il}^{(t+1)} = \frac{\sum_{j=1}^K (u_{ij}^{(t+1)})^m x_{jl}}{\sum_{j=1}^K (u_{ij}^{(t+1)})^m}.$$

Here,  $t$  is an index of the  $t$ -th iteration. By presetting a positive threshold  $\varepsilon$ , the iterative process is terminated when  $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$ .

### B. Tight Wavelet Frame Transform

Since tight wavelet frames can provide redundant representations of image data and exhibit substantial ability for feature/texture extraction, they have been successfully applied to various research areas, such as image segmentation [15], [28], image denoising [29]–[31], image restoration [25], [32], and mesh surface reconstruction [33], [34]. For simplicity, we present the main idea of a tight wavelet frame transform concisely. Its more details can be found in [25], [26]. Generally speaking, it consists of two operators, i.e., decomposition  $\mathcal{W}$  and reconstruction  $\mathcal{W}^T$ . By presetting a set of filters (masks), some finite sub-filtering operators are generated, i.e.,  $\mathcal{W}_0, \mathcal{W}_1, \mathcal{W}_2, \dots$ , which make up decomposition  $\mathcal{W}$ . More specifically,  $\mathcal{W}_0$  is a low-pass filtering operator and  $\mathcal{W}_1, \mathcal{W}_2, \dots$  are high-pass filtering operators. By unitary extension principle [35], reconstruction  $\mathcal{W}^T$  is available. Therefore, we have  $\mathcal{W}^T \mathcal{W} = \mathcal{I}$ , where  $\mathcal{I}$  is an identity matrix [26].

## III. PROPOSED METHODOLOGY

### A. Image Filtering via MR

To reduce the impact of noise/outliers on FCM as much as possible, a pre-processing step is performed by applying a filtering operation. It also takes full advantage of spatial information in image pixels, thus optimizing image pixel distribution for making it more favorable to clustering. That is to say, the pre-processing step makes the number of iterations of FCM decreased. Being superior to many sophisticated filtering operations, such as curvature filtering [36] and bitonic filtering [37], an MR operation exhibits the sound noise-immunity and retention capacity of image details. In light of MR's superiority, we employ it to filter an observed image in advance. Formally speaking, the residual  $\eta$  between an observed image  $g$  and its ideal value  $\tilde{g}$  is expressed as

$$\eta = g - \tilde{g}.$$

To reduce  $\eta$  from  $g$ , MR is used. It consists of two basic operators, i.e., dilation and erosion reconstructions [21], [38]. Formally speaking, given a mask image  $g$  and a mark image  $f$ , we denote the dilation reconstruction as:

$$\mathcal{R}_g^{\mathcal{D}}(f) = \mathcal{D}_g^{(t)}(f),$$

where  $\mathcal{D}_g^{(t)}(f)$  is expressed as

$$\mathcal{D}_g^{(t)}(f) = \begin{cases} \mathcal{D}(f) \dot{\wedge} g, & t = 1 \\ \mathcal{D}(\mathcal{D}^{(t-1)}(f)) \dot{\wedge} g, & t = 2, 3, \dots \end{cases}$$

where  $f \leq g$ ,  $\dot{\wedge}$  represents a point-wise minimum operation, and  $\mathcal{D}$  is the dilation of  $f$  with a flat structuring element of size  $\omega$ . The erosion reconstruction is denoted as:

$$\mathcal{R}_g^{\mathcal{E}}(f) = \mathcal{E}_g^{(t)}(f),$$

where  $\mathcal{E}_g^{(t)}(f)$  is formulated as

$$\mathcal{E}_g^{(t)}(f) = \begin{cases} \mathcal{E}(f) \dot{\vee} g, & t = 1 \\ \mathcal{E}(\mathcal{E}^{(t-1)}(f)) \dot{\vee} g, & t = 2, 3, \dots \end{cases}$$

where  $f \geq g$ ,  $\dot{\vee}$  stands for a point-wise maximum operation, and  $\mathcal{E}$  is the erosion of  $f$  with a flat structuring element of size  $\omega$ .

Based on  $\mathcal{R}_g^{\mathcal{D}}$  and  $\mathcal{R}_g^{\mathcal{E}}$ , the morphological closing reconstruction of  $g$  can be modeled as

$$\mathcal{R}^{\mathcal{C}}(g) = \mathcal{R}_{\mathcal{R}_g^{\mathcal{D}}(\mathcal{E}(g))}^{\mathcal{E}}(\mathcal{D}(\mathcal{R}_g^{\mathcal{D}}(\mathcal{E}(g)))), \quad (2)$$

In the sequel, we denote the filtered image as  $\bar{g} = \mathcal{R}^{\mathcal{C}}(g)$ . To exhibit MR's performance, we refer to Fig. 2 as an example. Here, a square of  $\omega = 3 \times 3$  is taken as the structuring element.

In Fig. 2, the second row stands for the corresponding gray level histograms of five images presented in the first row. Obviously, the original image includes four gray levels, i.e., 0, 85, 170, and 255, while its gray level histogram has four obvious peaks. The observed image is corrupted by mixed Gaussian and impulse noise (standard deviation = 10 and density = 10%). Its histogram has only two apparent peaks, i.e., 0 and 255. As Figs. 2(c)–(e) indicate, MR clusters image pixels into 4 groups while both curvature and bitonic filters produce some inhomogeneous peaks and cannot fully remove

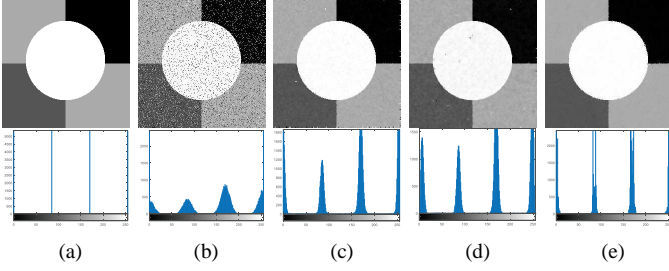


Fig. 2. Comparison of different filtering methods. (a) Original image; (b) observed image; (c) filtered result using a curvature filter; (d) filtered result using a bitonic filter; and (e) filtered result using MR.

noise. In summary, MR is superior to usual filters since it can effectively remove noise and retain image details.

In this work, by combining  $g$  and  $\bar{g}$ , we define a weighted sum image  $\hat{g}$  as

$$\hat{g} = \frac{g + \alpha \bar{g}}{1 + \alpha}, \quad (3)$$

where  $\alpha$  is a parameter assuming positive values for controlling the impact of the filtered image  $\bar{g}$ . Generally speaking,  $\hat{g}$  contains less noise than  $g$  and more features than  $\bar{g}$ . In addition, to express the composition of  $\hat{g}$  explicitly, we rewrite  $\hat{g}$  as

$$\hat{g} = \eta' + g', \quad (4)$$

where  $g'$  stands for the ideal value of  $\hat{g}$  and  $\eta'$  is the residual between  $\hat{g}$  and  $g'$ .

### B. Feature Extraction via Wavelet Frames

Due to sound feature extraction capacity of tight wavelet frames, we employ them to transform original data space to a new feature (kernel) space. Built on this new space, image features and noise can be sufficiently analyzed and manipulated. To be specific, after weighted sum image  $\hat{g}$  has been obtained, we generate its feature set by using a tight wavelet frame system. By employing a set of filters (masks), we construct  $M + 1$  sub-filtering operators  $\mathcal{W}_0, \mathcal{W}_1, \dots$ , and  $\mathcal{W}_M$ . They make up a tight wavelet frame decomposition operator  $\mathcal{W}$ . More specifically,  $\mathcal{W}_0$  is a low-pass filtering operator, and the remaining are high-pass filtering ones. As a result, we use  $\mathcal{W}$  to form the feature set associated with the weighted sum image, i.e.,

$$\mathbf{X} = \mathcal{W}\hat{g} = \{\mathcal{W}_0\hat{g}, \mathcal{W}_1\hat{g}, \dots, \mathcal{W}_M\hat{g}\}. \quad (5)$$

For image processing,  $\mathcal{W}_0\hat{g}$  is a wavelet coefficient that represents low-frequency information; while the rest represents high-frequency ones. Note that the selection of filters is described in Section IV-C. To better understand the redundancy and feature extraction capacity of wavelet frames, we show an example (see Fig. 3) in the presence of a piecewise linear B-spline tight frame system (described in Section IV-C). The level of tight wavelet frames is set to 1.

From Fig. 3, we see that the tight wavelet frame decomposition gives redundant representations of an original image, thus forming its feature set. Fig. 3(a) shows low-frequency information. High-frequency information is presented in Figs.

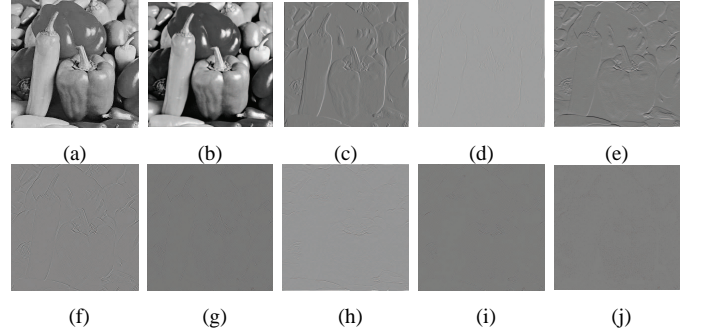


Fig. 3. Illustration for redundancy and feature extraction. (a) Original image and (b)–(j) wavelet coefficients.

3(b)–(j). In this case, each image pixel contains nine underlying attributes (channels). Formally speaking, for an image with  $K$  pixels, the size of its feature set  $\mathbf{X}$  is  $9 \times K$ . As a result, for image segmentation, the data dimensionality to be used in clustering is expanded in comparison with the direct use of image pixels.

### C. $\ell_0$ Regularization-based FCM

To achieve better segmentation effects, it is necessary to take the ideal value of an observed image as data for clustering, which means that the residual between them can be considered into clustering. As a result, we can augment FCM by introducing a sparse regularization term on the residual into its objective function. Before presenting the proposed algorithm, we respectively reformulate feature sets associated with  $\hat{g}$ ,  $g'$  and  $\eta'$  in (4) as

$$\begin{aligned} \mathbf{X} &= \mathcal{W}\hat{g} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}, \\ \widetilde{\mathbf{X}} &= \mathcal{W}g' = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_K\}, \\ \mathbf{R} &= \mathcal{W}\eta' = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K\}. \end{aligned}$$

Here, we have

$$\widetilde{\mathbf{X}} = \mathbf{X} - \mathbf{R}. \quad (6)$$

In addition,  $\mathbf{R}$  can also be rewritten as  $\{\mathbf{R}_l : l = 1, 2, \dots, L\}$ , which means that each component  $\mathbf{R}_l$  contains  $L$  channels. Then we introduce an  $\ell_p$  regularization term on  $\mathbf{R}$  into the objective function of FCM. In the sequel, the augmented objective function is expressed as

$$J(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2 + \sum_{l=1}^L \beta_l \|\mathbf{R}_l\|_{\ell_p}^p, \quad (7)$$

where  $p \geq 0$ ,  $\beta = \{\beta_l : l = 1, 2, \dots, L\}$  is a parameter vector that controls the impact of  $\ell_p$  regularization on FCM, and

$$\|\mathbf{R}_l\|_{\ell_p}^p = \begin{cases} \sum_{j=1}^K |r_{jl}|^p, & p > 0 \\ \sum_{j=1}^K |r_{jl}|_0, & p = 0 \end{cases} \quad (8)$$

with

$$|r_{jl}|_0 = \begin{cases} 1, & r_{jl} \neq 0 \\ 0, & r_{jl} = 0 \end{cases}. \quad (9)$$

In (8),  $\|\cdot\|_{\ell_p}$  is an  $\ell_p$  vector norm. Especially,  $\|\cdot\|_{\ell_0}$  denotes the  $\ell_0$  vector norm, and  $\|\mathbf{R}_l\|_{\ell_0}$  represents the number of nonzero

entries in  $\mathbf{R}_l$ . When  $p \in [0, 1)$ , the minimization of (7) is a nonconvex problem. As to  $p \in \{4/3, 3/2, 2\}$ , the closed-form solution to the minimization of (7) is derived in [39] and [40]. In particular, for  $p = 1$ , the closed-form solution is given in [22] by using a general soft-thresholding operator.

Generally speaking, a large proportion of image data has a small or zero number of outliers, noise or intensity inhomogeneity. Therefore,  $\mathbf{R}$  tends to be very sparse. To take the sparsity of  $\mathbf{R}$  into consideration, in this work we focus on the case  $p = 0$ . The main difference between this study and the above cases is in the form of the norms used for  $\mathbf{R}$ . Even though the use of  $\ell_0$  norm gives rise to the difficulty for designing effective algorithms to solve the underlying optimization problems, it is beneficial to cope with a variety of cases [41]. Therefore, (7) is fixed as the following expression:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2 + \sum_{l=1}^L \beta_l \sum_{j=1}^K |r_{jl}|_0. \quad (10)$$

In addition, the use of spatial information is beneficial to improve FCM's robustness. If the distance between an image pixel and its neighbors is small, there exists a large possibility that they belong to the same cluster. To further improve segmentation performance, we introduce spatial information into the objective function of FCM. Prior to performing the modified objective function, we need to complete an analysis of spatial information. To make spatial information easily understood, refer to Fig. 4.

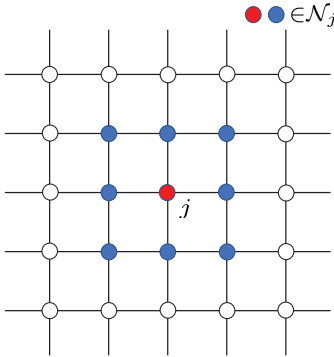


Fig. 4. Illustration of spatial information of pixel  $j$ .

In a concise way, a pixel or feature is sometimes loosely represented by its corresponding index while this is not ambiguous. In Fig. 4, we show an arbitrary pixel  $j$  and its spatial information with a local window.  $\mathcal{N}_j$  stands for a local window centralized in  $j$  including  $j$ . We let  $|\mathcal{N}_j|$  be the cardinality of  $\mathcal{N}_j$ , which represents the size of a local window  $\mathcal{N}_j$ . In Fig. 4, we have  $|\mathcal{N}_j| = 3 \times 3$ , which represents that  $\mathcal{N}_j$  contains pixel  $j$  and its eight neighbors.

According to (10),  $u_{ij}$  depends on the distance  $\|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|$ . Thus, it is optimized by considering spatial information of  $\mathbf{x}_j$  and  $\mathbf{r}_j$ . Motivated by [9], we express the spatial information of  $\mathbf{x}_j$  and  $\mathbf{r}_j$  as

$$\sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}}, \quad \sum_{n \in \mathcal{N}_j} \frac{|r_{nl}|_0}{1 + d_{nj}}, \quad (11)$$

where  $n$  is a neighbor pixel of  $j$ ,  $d_{nj}$  represents the Euclidean distance between  $n$  and  $j$ , and the factor  $1/(d_{nj} + 1)$  reflects

the spatial structure information. As a result, by substituting (11) into (10), the modified objective function can be defined as

$$J(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \left( \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|r_{nl}|_0}{1 + d_{nj}}, \quad (12)$$

subject to

$$\sum_{i=1}^c u_{ij} = 1, \quad \forall j \in \{1, 2, \dots, K\}.$$

In the sequel, the Lagrangian multiplier method is applied to minimize (12). The augmented Lagrangian function is:

$$\mathcal{L}_\Lambda(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \left( \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|r_{nl}|_0}{1 + d_{nj}} + \sum_{j=1}^K \lambda_j \left( \sum_{i=1}^c u_{ij} - 1 \right), \quad (13)$$

where  $\Lambda = \{\lambda_j : j = 1, 2, \dots, K\}$  stands for a set of Lagrangian multipliers. The solution to the minimization of (12) can be produced in an iterative manner by handling the following three subproblems:

$$\begin{cases} \mathbf{U}^{(t+1)} = \arg \min_{\mathbf{U}} \mathcal{L}_\Lambda(\mathbf{U}, \mathbf{V}^{(t)}, \mathbf{R}^{(t)}) \\ \mathbf{V}^{(t+1)} = \arg \min_{\mathbf{V}} \mathcal{L}_\Lambda(\mathbf{U}^{(t+1)}, \mathbf{V}, \mathbf{R}^{(t)}) \\ \mathbf{R}^{(t+1)} = \arg \min_{\mathbf{R}} \mathcal{L}_\Lambda(\mathbf{U}^{(t+1)}, \mathbf{V}^{(t+1)}, \mathbf{R}) \end{cases}. \quad (14)$$

Each of the subproblems of (14) has a closed-form solution. We adopt an alternative optimization scheme similar to that used in the FCM algorithm to conduct the optimization of the partition matrix  $\mathbf{U}$  and prototypes  $\mathbf{V}$ . The iterative updates of  $\mathbf{U}$  and  $\mathbf{V}$  are easily given as

$$u_{ij}^{(t+1)} = \frac{\left( \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n^{(t)} - \mathbf{v}_i^{(t)}\|^2}{1 + d_{nj}} \right)^{-\frac{1}{m-1}}}{\sum_{q=1}^c \left( \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n^{(t)} - \mathbf{v}_q^{(t)}\|^2}{1 + d_{nj}} \right)^{-\frac{1}{m-1}}}, \quad (15)$$

$$\mathbf{v}_i^{(t+1)} = \frac{\sum_{j=1}^K \left( \left( u_{ij}^{(t+1)} \right)^m \sum_{n \in \mathcal{N}_j} \frac{\mathbf{x}_n - \mathbf{r}_n^{(t)}}{1 + d_{nj}} \right)}{\sum_{j=1}^K \left( \left( u_{ij}^{(t+1)} \right)^m \sum_{n \in \mathcal{N}_j} \frac{1}{1 + d_{nj}} \right)}. \quad (16)$$

When optimizing  $\mathbf{R}$ , it is obvious that both  $\mathbf{r}_j$  and  $\mathbf{r}_n$  are in (12). Since  $\mathbf{r}_j$  is not independent from  $\mathbf{r}_n$ ,  $\mathbf{r}_n$  cannot be treated as a constant vector. If  $n$  is one of neighbors of  $j$ ,  $j$  is also one of neighbors of  $n$  symmetrically. In the sequel,  $n \in \mathcal{N}_j$  is equivalent to  $j \in \mathcal{N}_n$ . Then we have

$$\sum_{j=1}^K u_{ij}^m \left( f(\mathbf{r}_j) + \sum_{\substack{n \in \mathcal{N}_j \\ n \neq j}} f(\mathbf{r}_n) \right) = \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} u_{in}^m (f(\mathbf{r}_j)), \quad (17)$$

where  $f$  stands for a function in terms of  $\mathbf{r}_j$  or  $\mathbf{r}_n$ . According to (17), (12) is rewritten as

$$J(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2}{1 + d_{nj}} + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|r_{jl}|_0}{1 + d_{nj}}. \quad (18)$$

Based on (18), once  $\mathbf{U}$  and  $\mathbf{V}$  have been updated, the third subproblem of (14) is separable and the optimization of  $\mathbf{R}$  can be decomposed into  $K \times L$  subproblems as follows:

$$r_{jl}^{(t+1)} = \arg \min_{r_{jl}} \sum_{i=1}^c \left( \sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m \|x_{jl} - r_{jl} - v_{il}^{(t+1)}\|^2}{1 + d_{nj}} \right) + \sum_{n \in \mathcal{N}_j} \frac{\beta_l |r_{jl}|_0}{1 + d_{nj}}. \quad (19)$$

The following result is needed to obtain the iterative update of residual  $r_{jl}$ .

**Theorem III.1.** *Consider the minimization problem (19). By applying a well-known hard-thresholding operator to solve (19), the iterative formula of residuals is obtained as follows:*

$$r_{jl}^{(t+1)} = \mathcal{H}_{\sum_{n \in \mathcal{N}_j} \frac{\beta_l}{1 + d_{nj}}} \left( \frac{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m (x_{jl} - v_{il}^{(t+1)})}{1 + d_{nj}}}{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m}{1 + d_{nj}}} \right), \quad (20)$$

where  $\mathcal{H}$  is a hard-thresholding operator defined as [42]

$$\mathcal{H}_\sigma(\xi) = \begin{cases} \xi, & \xi \geq \sqrt{\sigma} \\ 0, & \xi < \sqrt{\sigma} \end{cases}. \quad (21)$$

*Proof:* See Appendix. ■

#### D. Label Smoothing via MR

To further enhance the segmentation effects of the  $\ell_0$  regularization-based FCM algorithm, we use MR to smoothen the obtained labels of pixels. It is taken as a post-processing step to reduce the possibility of misclassification. We define the label of the  $j$ -th pixel as  $\phi_j = \text{label}(u_{ij})$ :

$$u_{ij} = \arg \max \{u_{1j}, u_{2j}, \dots, u_{cj}\},$$

where  $\text{label}$  denotes the location of maximum  $u_{ij}$ , i.e.,  $\phi_j = i$ . This means that the  $j$ -th pixel belongs to the  $i$ -th cluster. Thus, we can define the label set of  $K$  pixels as

$$\Phi = \{\phi_j : j = 1, 2, \dots, K\}. \quad (22)$$

In the sequel,  $\Phi$  is arranged into a matrix of the same size as  $\hat{g}$ , thus generating a label image  $\Phi_{im}$ . We employ MR to smoothen the obtained label image  $\Phi_{im}$  so as to generate a smoothed label image that is formulated as

$$\bar{\Phi}_{im} = \mathcal{R}^C(\Phi_{im}). \quad (23)$$

Based on the smoothed label image  $\bar{\Phi}_{im}$  and the obtained prototypes  $\mathbf{V}$ , the segmented feature set  $\hat{\mathbf{X}}$  is obtained. Then we use the wavelet frame reconstruction operator  $\mathcal{W}^T$  to reconstruct a segmented image  $g''$ :

$$g'' = \mathcal{W}^T(\hat{\mathbf{X}}). \quad (24)$$

By combining four components mentioned above, i.e., image filtering, feature extraction,  $\ell_0$  regularization, and label smoothing, the proposed algorithm for image segmentation is realized in **Algorithm 1**.

**Algorithm 1**  $\ell_0$  regularization-based FCM algorithm incorporating MR and wavelet frames (LRFCM)

**Input:** Observed image  $g$ , control parameter  $\alpha$ , fuzzification coefficient  $m$ , number of clusters  $c$ , and threshold  $\varepsilon$ .

**Output:** Segmented image  $g''$ .

- 1: Calculate the filtered image  $\bar{g}$  via (2)
- 2: Calculate the weighted sum image  $\hat{g}$  via (3)
- 3: Generate the feature set  $\mathbf{X}$  via (5)
- 4: Initialize randomly the prototypes  $\mathbf{V}^{(0)}$
- 5:  $t \leftarrow 0$
- 6: **repeat**
- 7:   Calculate the partition matrix  $\mathbf{U}^{(t+1)}$  via (15)
- 8:   Update the prototypes  $\mathbf{V}^{(t+1)}$  via (16)
- 9:   Update the residuals  $\mathbf{R}^{(t+1)}$  via (20)
- 10:    $t \leftarrow t + 1$
- 11: **until**  $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$
- 12: **return** partition matrix  $\mathbf{U}$ , prototypes  $\mathbf{V}$ , and residuals  $\mathbf{R}$
- 13: Generate the labels of image pixels via (22)
- 14: Smoothen the labels via (23)
- 15: Calculate the segmented image  $g''$  via (24)

## IV. EXPERIMENTAL STUDIES

In this section, we proceed with numerical experiments to investigate the effectiveness and efficiency of the proposed algorithm (namely LRFCM). Numerical results reported for synthetic, medical, and real-world images are provided. We compare LRFCM with the four classic algorithms studied in the literature, i.e., FCM\_S1 [6], FCM\_S2 [6], FGFCM [8], and FLICM [9], and six recently proposed algorithms including KWFLICM [10], ARKFCM [11], FRFCM [19], WFCM [15], SRFCM [21], and DSFCM\_N [22]. We also conduct ablation studies and analyze the impact of each component in LRFCM.

### A. Evaluation Indicator

To quantitatively evaluate comparison results of all algorithms, we adopt five performance indicators as follows:

- 1) Segmentation Accuracy (SA) [10]: It is used to assess the performance of segmenting images with known ground truth. The larger SA, the better segmentation effect.
- 2) Dice Coefficient (DC) [43]: It is used to measure the similarity between a segmented image and its ground truth. The higher DC, the better segmentation result.
- 3) Entropy-based Information (EI) [44]: It is used to evaluate the performance for segmenting images without ground truth. Its essence is to minimize the uniformity across all clusters by maximizing the uniformity of pixels within each segmented cluster. Hence, a better segmentation effect is associated with a smaller EI value.
- 4) Partition Coefficient (PC) [45]: It is a validity index to measure the partitions of a clustering algorithm. The higher PC is, the better clustering result is.
- 5) Partition Entropy (PE) [46]: Its use is similar to PC's. However, a better segmentation effect associates with a smaller PE value.



### B. Dataset Description

Except synthetic images, we test four publicly available databases to exhibit the performance of all algorithms. They are outlined as follows:

- 1) BrianWeb<sup>1</sup>: This is an online interface to a 3D MRI simulated brain database. The parameter settings are fixed to 3 modalities, 5 slice thicknesses, 6 levels of noise, and 3 levels of intensity non-uniformity. BrianWeb offers golden standard segmentation.
- 2) Berkeley Segmentation Data Set (BSDS)<sup>2</sup>: It contains 500 images, i.e., 200 training, 100 validation and 200 testing images. It gives golden standard segmentation.
- 3) Microsoft Research Cambridge Object Recognition Image Database (MSRC)<sup>3</sup>: It has 591 images and 23 object classes with golden standard segmentation.
- 4) NASA Earth Observation Database (NEO)<sup>4</sup>: It continually provides information collected by NASA satellites about Earth's ocean, atmosphere, and land surfaces. There exist different levels of unknown noise in sampled images of size  $1440 \times 720$ , which result from bit errors appearing in satellite measurements. Therefore, their ground truth is unknown.

### C. Parameter Setting

Prior to accomplishing all numerical experiments, we report parameter settings of all algorithms. For FLICM, KWFLICM, ARKFCM, and DSFCM\_N, the size of a local window centralized in pixel  $j$  is set to  $|\mathcal{N}_j| = 3 \times 3$ . For other peers, the size  $\omega$  of a filtering window is fairly set to  $3 \times 3$  since image filtering is considered in such comparative algorithms. We set fuzzification exponent  $m = 2$ , and threshold  $\varepsilon = 1 \times 10^{-6}$  across all algorithms. Moreover, the suitable number of clusters  $c$  is assumed to be known, and how to decide  $c$  is introduced in each experiment.

Except for usual parameters  $m$ ,  $\varepsilon$ , and  $c$ , there are no more parameters in FLICM, KWFLICM and ARKFCM. Nevertheless, there exist different parameters in the remaining algorithms. We have to clarify that such parameters are optimal in the experiment. For FCM\_S1 and FCM\_S2,  $\alpha$  is uniformly set to 4, which aims to constrain the neighbor term. In FGFCM, the spatial scale factor  $\lambda_s$  and gray-level scale factor  $\lambda_g$  are respectively set to 3 and 5. For FRFCM, according to [19], we select the observed image as the mask image, and generate the marker image with the aid of a square structuring element of size  $3 \times 3$ . Moreover, a median filter of size  $3 \times 3$  is applied to the membership filtering. As to WFCM, we experimentally select  $\mu \in [0.55, 0.65]$ , which is used to control the effect of spatial information. The level of tight wavelet frame transform is chosen as 1. In SRFCM, MR is applied to image pixel filtering and label smoothing based on the same setting as that in FRFCM. The 1-level wavelet frame transform is used.  $\beta$  is chosen from an interval  $[0, 0.06]$ , which

controls the impact of a sparse regularization term on FCM. For DSFCM\_N, the parameter vector  $\lambda$  is selected according to the standard deviation of each channel of image data.

Except a usual threshold  $\varepsilon$ , there are four common parameters, i.e.,  $m$ ,  $\omega$ ,  $\alpha$ , and  $|\mathcal{N}_j|$ , are involved in LRFCM, where  $m$  is a fuzzification exponent,  $\omega$  is a filtering window of MR,  $\alpha$  is a positive parameter for controlling the impact of the filtered image, and  $|\mathcal{N}_j|$  denotes the size of a local window centralized in pixel  $j$ . To better elaborate their selection, we cover an example in Table I, where the second image in Fig. 6 is segmented after 40% impulse noise has been added. As Table I shows, LRFCM's performance depends on these parameters significantly. When  $m = 2$  and  $\alpha = 4$ , both SA and iterations reach the best. When  $\omega$  and  $|\mathcal{N}_j|$  are both  $3 \times 3$ , in spite of more iterations, LRFCM's performance is the best. Therefore, in this work, we set  $m = 2$ ,  $\omega = 3 \times 3$ ,  $\alpha = 4$ , and  $|\mathcal{N}_j| = 3 \times 3$ .

TABLE I  
SELECTION OF COMMON PARAMETERS IN LRFCM

$m$	SA (%)	Iterations	$\omega$	SA (%)	Iterations
2	99.830	17	$1 \times 1$	99.241	25
3	99.827	18	$3 \times 3$	99.830	17
4	99.825	25	$5 \times 5$	99.825	14
5	90.420	65	$7 \times 7$	99.813	11
6	89.256	67	$9 \times 9$	99.738	9
$\alpha$	SA (%)	Iterations	$ \mathcal{N}_j $	SA (%)	Iterations
1	98.858	20	$1 \times 1$	99.179	23
2	99.395	18	$3 \times 3$	99.830	17
3	99.723	17	$5 \times 5$	99.765	15
4	99.830	17	$7 \times 7$	99.626	11
5	99.824	17	$9 \times 9$	99.570	10

As to a tight wavelet frame transform, we choose a piecewise linear B-spline tight frame system since it is commonly used for feature extraction and redundant representations of images [35], [47]. In this system, three one-dimensional filters are discretized as  $h_0 = [1/4, 1/2, 1/4]$ ,  $h_1 = [-1/4, 1/2, -1/4]$ , and  $h_2 = [\sqrt{2}/4, 0, -\sqrt{2}/4]$ . By employing such filters, we generate nine two-dimensional filters so as to construct nine corresponding sub-filtering operators  $\mathcal{W}_0, \mathcal{W}_1, \dots, \mathcal{W}_8$ . Thus,  $M$  in (5) is equal to 8. For  $j = 1, 2, \dots, K$ , each pixel  $x_j$  has nine channels, i.e.,  $L = 9$ . In addition, the level of tight wavelet frame transform is set to 1 since higher levels bring expensive calculations while segmentation performance is only slightly improved [15].

Except the above parameters, there only exists one parameter  $\beta$  in LRFCM. Since the main difference between LRFCM and DSFCM\_N is the norms used for residual  $R$ , the parameter setting for the sparse regularization term in DSFCM\_N could be referenced.  $\beta$  is associated with the standard deviation of each channel of image data since the standard deviation can reflect the noise level to some extent [22]. Therefore,  $\beta = \{\beta_l : l = 1, 2, \dots, L\}$  is expressed as

$$\beta_l = \zeta \cdot \delta_l,$$

where  $\zeta$  is a positive parameter and  $\delta_l$  is the standard deviation of the  $l$ -th channel of image data. After massive experiments,  $\zeta$  is recommended to be 70. Here, we report an example about the selection process of  $\zeta$ , as shown in Fig. 5. Here,

<sup>1</sup><http://www.bic.mni.mcgill.ca/brainweb/>

<sup>2</sup><https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html>

<sup>3</sup><http://research.microsoft.com/vision/cambridge/recognition/>

<sup>4</sup><http://neo.sci.gsfc.nasa.gov/>

we segment the second image shown in Fig. 6, which is contaminated by 40% impulse noise.

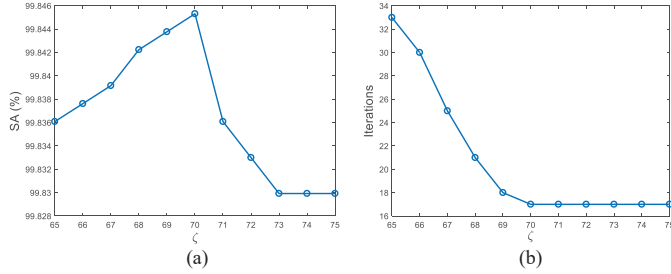


Fig. 5. Segmentation results with changes of  $\zeta$ . (a) SA and (b) iterations versus  $\zeta$ .

As Fig. 5 reveals, as  $\zeta$  increases, the SA value first increases and then decreases sharply. When  $\zeta = 70$ , the SA reaches its maximum value. Moreover, the iteration count of LRFCM is minimized. Therefore, we choose  $\zeta = 70$ .

#### D. Ablation Studies and Analysis

We provide ablation experiments to show the effects of four key components involved in LRFCM, i.e., image filtering, feature extraction,  $\ell_0$  regularization, and label smoothing. We conduct ablation experiments on BSDS. BSDS collects 500 images and pays students to hand-segment each one (usually each image has multiple hand-segmentations). In this study, we choose a typical hand-segmentation by following [48]. The numbers of clusters for 500 images are summarized in the Excel form titled as ‘BSDS500\_Performance’<sup>5</sup>. All experiments are implemented on the same experimental platform. The average results are presented in Table II. The symbol  $\times$  indicates that its corresponding component is absent while symbol  $\checkmark$  states the component is activated.

TABLE II  
INVESTIGATION OF EACH COMPONENT IN LRFCM

Image filtering	Feature extraction	$\ell_0$ regularization	Label smoothing	SA (%)	Iterations
$\times$	$\times$	$\times$	$\times$	72.815	28
$\checkmark$	$\times$	$\times$	$\times$	82.729	24
$\times$	$\checkmark$	$\times$	$\times$	78.375	20
$\times$	$\times$	$\checkmark$	$\times$	93.497	35
$\times$	$\times$	$\times$	$\checkmark$	73.537	28
$\checkmark$	$\checkmark$	$\checkmark$	$\times$	97.672	23
$\checkmark$	$\checkmark$	$\times$	$\checkmark$	87.754	18
$\checkmark$	$\times$	$\checkmark$	$\checkmark$	95.275	28
$\times$	$\checkmark$	$\checkmark$	$\checkmark$	93.428	25
$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	98.180	23

As shown in Table II, ten combinations of the four key components are tested. When all components are not present, the SA value is only 72.82%. If we consider each component alone, the average SA values are increased by 9.91%, 5.56%, 20.68%, and 0.72%, respectively. Hence,  $\ell_0$  regularization exhibits the greatest impact on the improvement of FCM. Nevertheless, it also results in more iterations. The label smoothing slightly improves SA values. It does not give rise to more iterations since it is used as a post-processing step in LRFCM. Moreover, MR and tight wavelet frames can not only

further segmentation performance, but also make iterations of LRFCM greatly decreased due to their characteristics for data optimization.

#### E. Results for Synthetic Images

In the first experiments, we test three synthetic images shown in Fig. 6. The first two images have ground truth, which contain four gray levels, i.e., 0, 85, 170, and 255. The last image is without ground truth. We impose different levels of Rician noise (with probability  $R_N$ ), Gaussian noise (with standard deviation  $G_N$ ), and impulse noise (with probability  $I_N$ ) on them, respectively. Note that salt and pepper impulse noise is only considered since it is one of the most common types of impulse noise. The numbers of clusters are set to 4, 4, and 2, respectively. The segmentation results of all algorithms are summarized in Figs. 7–9. To intuitively show the visual comparison of all algorithms, we cover three cases in Fig. 10.

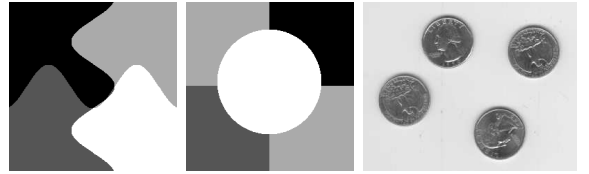


Fig. 6. Three synthetic images.

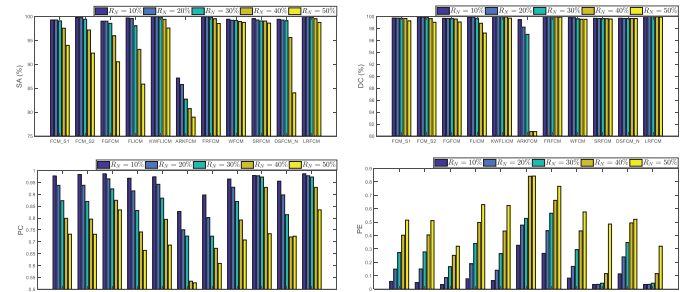


Fig. 7. Segmentation results on the first synthetic image corrupted by different levels of Rician noise.

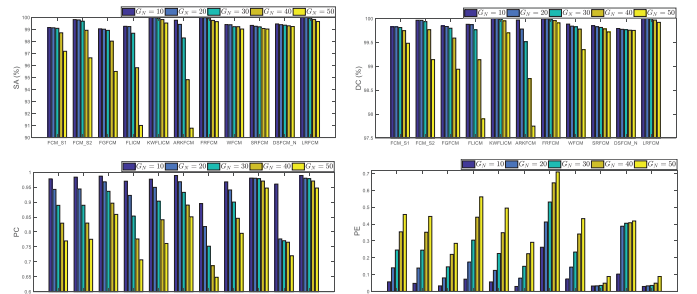


Fig. 8. Segmentation results on the second synthetic image contaminated by different levels of Gaussian noise.

As illustrated in Figs. 7–9, LRFCM generally performs better than its peers. From Fig. 10, we find that FCM\_S1, FCM\_S2, FGFCM, FLICM and ARKFCM have poor performance in suppressing high levels of Gaussian noise and

<sup>5</sup><https://github.com/jiaxsust>



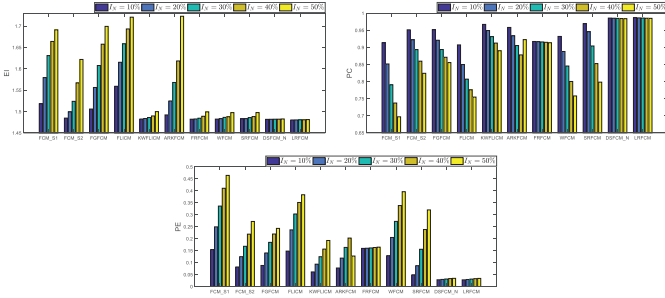


Fig. 9. Segmentation results on the third synthetic image corrupted by different levels of impulse noise.

Rician noise. Although KWFLICM, FRFCM, and DSFCM\_N can remove a large proportion of such noise, there is a small amount of noise in their segmentation results. On contrary, WFCM and SRFCM have a good capacity of noise suppression. However, they produce several topology changes such as merging and splitting. Superior to its ten peers, LRFCM is robust to such noise and retains more image features. As to impulse noise removal, except WFCM, SRFCM, and DSFCM\_N, other comparative algorithms cannot perform well. In particular, the segmentation results of FCM\_S1, FCM\_S2, FGFCM, FLICM and ARKFCM are far from being satisfactory. When focusing on the results of WFCM, SRFCM, and DSFCM\_N, we find that there are some unsmooth edges. Nevertheless, DSFCM\_N performs better than WFCM since it is modeled based on an analysis of noise sparsity. Compared with DSFCM\_N, LRFCM yields slightly better results since it achieves smoother edges.

#### F. Results for Medical Images

In the second experiments, we segment five medical images coming from BrianWeb, as shown in Fig. 11. They are generated by T1 modality with slice thickness of 1mm resolution, 9% noise and 20% intensity non-uniformity. Here, they are represented as five slices in the axial plane with the sequence of 70, 80, 90, 100, and 110. We set the numbers of clusters to 4. The segmentation results are illustrated in Table III and Fig. 12.

From Table III, we see that LRFCM achieves better results than other peers. In particular, the SA value of LRFCM comes up to 82.490% for the fourth medical image shown in Fig. 11. By focusing on the marked red square in Fig. 12, we easily find that FCM\_S1, FCM\_S2, FGFCM and ARKFCM are sensitive to noise. FLICM and KWFLICM are vulnerable to severe intensity inhomogeneity. FRFCM brings overly smooth results due to the use of gray level histograms. WFCM, SRFCM, and DSFCM\_N cause several contours to change. However, LRFCM acquires clear contours and suppresses noise adequately. Moreover, its segmentation result is closer to the ground truth than its peers'.

#### G. Results for Real-world Images

In the last experiments, we segment some real-world images. They are represented as Red-Green-Blue (RGB) color

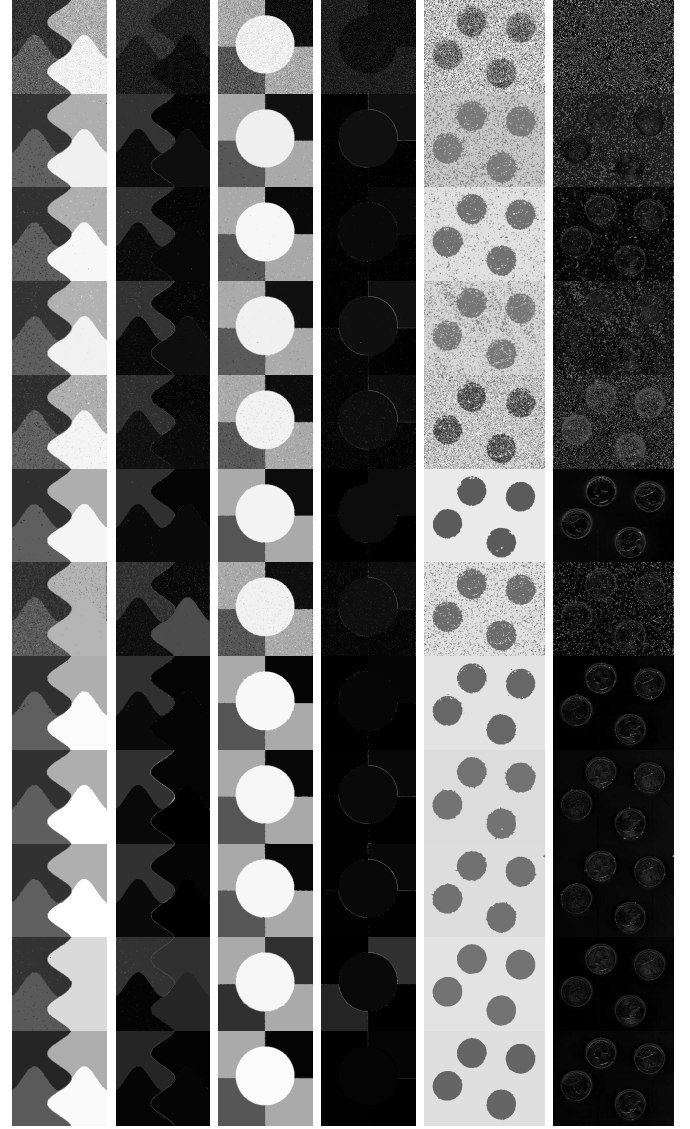


Fig. 10. Segmentation results for three synthetic images. The noise levels:  $R_N = 40\%$ ,  $G_N = 40$ , and  $I_N = 40\%$ . The odd-numbered columns: segmented images; and even-numbered columns: residual images. From top to bottom: noisy image and results of FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, FRFCM, WFCM, SRFCM, DSFCM\_N, and LRFCM.

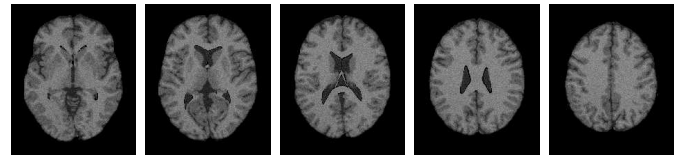


Fig. 11. Five medical images.

images. It is easy to extend LRFCM to color image segmentation. We apply the multivariate MR to color images [49]. Moreover, we conduct the tight wavelet frame transform in each channel of an RGB color image. The dimensionality of the obtained feature set is three times higher than that of a gray image. The remaining settings are similar to those in gray image segmentation. In the following, we test real-worlds coming from BSDS, MSRC, and NEO. Note that we first compute SA, DC, PC, and PE values of segmenting each

TABLE III  
SEGMENTATION PERFORMANCE ON MEDICAL IMAGES IN BRIANWEB

Algorithm	First medical image				Second medical image				Third medical image				Fourth medical image				Fifth medical image			
	SA (%)	DC (%)	PC	PE	SA (%)	DC (%)	PC	PE	SA (%)	DC (%)	PC	PE	SA (%)	DC (%)	PC	PE	SA (%)	DC (%)	PC	PE
FCM_S1	75.757	97.853	0.874	0.243	75.026	98.109	0.873	0.246	79.792	98.452	0.888	0.219	81.888	98.614	0.902	0.193	81.870	94.254	0.906	0.184
FCM_S2	75.770	98.120	0.879	0.233	74.970	98.176	0.880	0.232	79.886	98.458	0.894	0.207	82.073	98.625	0.908	0.182	81.788	98.224	0.910	0.175
FGFCM	75.553	97.279	0.896	0.199	74.940	97.096	0.894	0.204	79.255	97.590	0.907	0.180	81.625	97.773	0.919	0.158	81.552	94.458	0.921	0.152
FLICM	74.998	98.070	0.858	0.270	74.186	98.122	0.858	0.271	79.089	98.516	0.870	0.248	81.447	98.627	0.886	0.220	81.669	98.273	0.894	0.203
KWFLICM	74.840	98.260	0.856	0.287	73.840	97.860	0.855	0.289	79.561	98.454	0.869	0.262	81.888	98.482	0.885	0.230	81.371	98.297	0.898	0.205
ARKFCM	75.754	98.216	0.913	0.169	74.942	91.940	0.906	0.182	80.105	98.382	0.921	0.155	81.037	94.522	0.914	0.166	80.459	95.028	0.916	0.160
FRFCM	75.854	97.621	0.761	0.460	75.515	97.660	0.760	0.461	80.284	98.278	0.784	0.420	81.852	98.319	0.805	0.381	81.666	98.079	0.815	0.360
WFCM	75.507	97.124	0.826	0.328	74.471	97.214	0.827	0.327	79.316	97.846	0.848	0.291	81.358	97.547	0.865	0.260	81.452	95.247	0.868	0.251
SRFCM	75.683	97.172	0.887	0.177	74.850	97.155	0.889	0.173	79.609	97.988	0.904	0.151	81.979	98.202	0.916	0.131	81.511	94.282	0.919	0.126
DSFCM_N	76.401	92.326	0.840	0.305	75.288	91.575	0.848	0.292	79.861	97.679	0.862	0.267	81.832	93.305	0.873	0.247	81.750	94.303	0.884	0.226
LRFCM	80.758	98.323	0.928	0.132	77.513	98.208	0.929	0.124	81.548	98.614	0.930	0.149	82.490	98.824	0.927	0.123	82.483	98.344	0.930	0.118

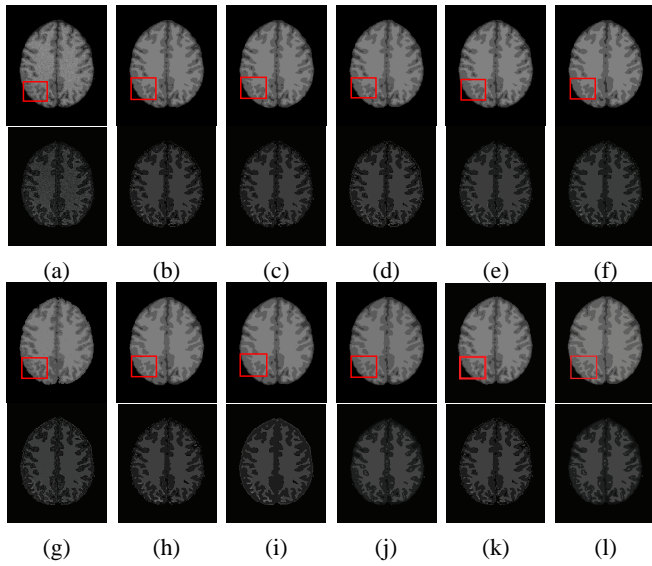


Fig. 12. Segmentation results for a medical image. The odd-numbered rows: segmented images; and even-numbered rows: residual images. From (a) to (l): noisy image and results of FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, FRFCM, WFCM, SRFCM, DSFCM\_N, and LRFCM.

image in BSDS and MSRC. Then we compute their average values for each dataset. The average segmentation results on BSDS and MSRC are gathered in Table IV. For illustrate LRFCM's superior performance intuitively, we show some visual comparison results with fixed clusters in Fig. 13.

TABLE IV  
AVERAGE SEGMENTATION PERFORMANCE ON REAL-WORLD IMAGES IN BSDS AND MSRC

Algorithm	BSDS				MSRC			
	SA (%)	DC (%)	PC	PE	SA (%)	DC (%)	PC	PE
FCM_S1	72.815	72.832	0.882	0.214	69.631	70.019	0.791	0.258
FCM_S2	72.940	73.064	0.883	0.210	70.184	70.341	0.789	0.253
FGFCM	73.089	73.180	0.885	0.201	69.758	70.045	0.793	0.222
FLICM	78.478	70.662	0.891	0.190	75.664	67.496	0.794	0.212
KWFLICM	71.687	62.262	0.884	0.213	68.536	57.536	0.781	0.249
ARKFCM	71.963	66.717	0.864	0.232	69.047	62.418	0.784	0.274
FRFCM	95.546	96.585	0.954	0.125	95.078	96.061	0.851	0.154
WFCM	95.670	96.672	0.956	0.124	94.754	96.078	0.847	0.137
SRFCM	96.023	96.829	0.961	0.108	95.515	95.538	0.858	0.139
DSFCM_N	76.435	68.933	0.891	0.190	74.071	65.651	0.802	0.222
LRFCM	98.180	98.644	0.964	0.067	97.047	97.917	0.867	0.075

According to Table IV, we observe that LRFCM exhibits the best segmentation effects among all algorithms. The similar finding can be acquired in Fig. 13. LRFCM can not only retain

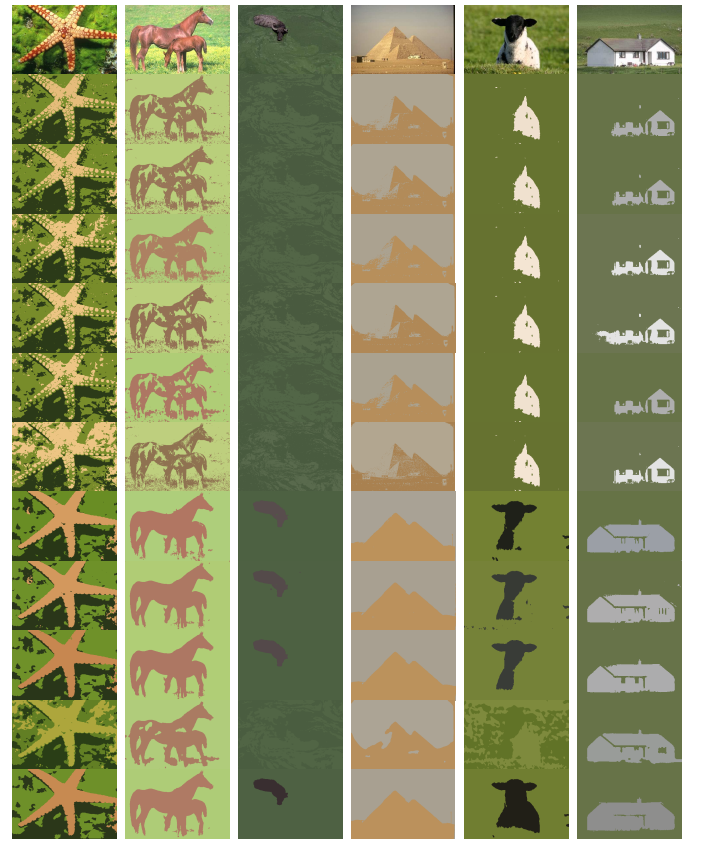


Fig. 13. Segmentation results for six real-world images in BSDS (column 1–4) and MSRC (column 5–6). From top to bottom: original images and results of FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, FRFCM, WFCM, SRFCM, DSFCM\_N, and LRFCM.

true contours but also suppress clutter in color images. In contrast, FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM and ARKFCM achieve unsatisfactory visual results. These six algorithms cannot preserve clear contours while losing important image details. Superior to them, FRFCM, WFCM, and SRFCM retain a large proportion of shapes. However, there are still a small amount of clutter in their segmentation results. In addition, DSFCM\_N has unstable segmentation performance. As shown in the penultimate row of Fig. 13, even though DSFCM\_N has slightly better ability to track the contours of the second image and last image, it cannot exhibit good segmentation effects for other images.

Besides real-world images in BSDS and MSRC, we also segment two real-world images from NEO. We clarify that each image represents a specific scene. Since there exists unknown noise in these images, their reference (original) images are missing. In the sequel, indicator EI cannot directly used. To address this issue, we randomly shoot each scene for 50 times within the time span 2000–2019, which generates the mean image. It is used as the noise-free (reference) image. The calculated EI, PC, and PE values are given in Table V. The corresponding visual results are illustrated in Figs. 14 and 15.

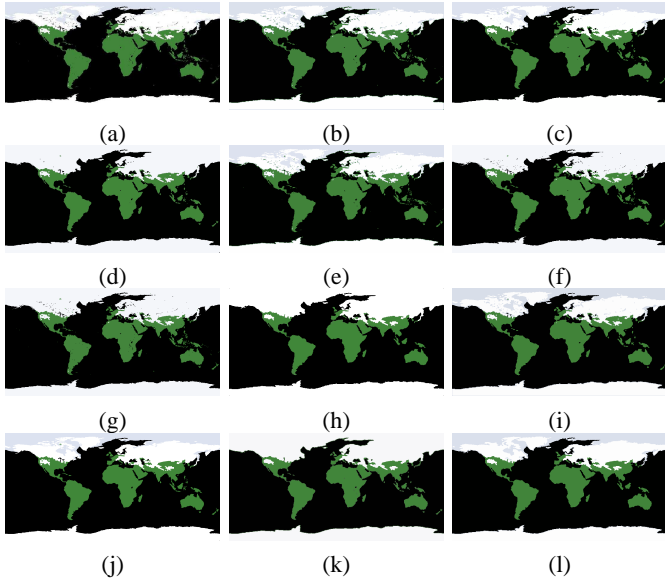


Fig. 14. Segmentation results for the first real-world image in NEO. From (a) to (l): noisy image and results of FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, FRFCM, WFCM, SRFCM, DSFCM\_N, and LRFCM.

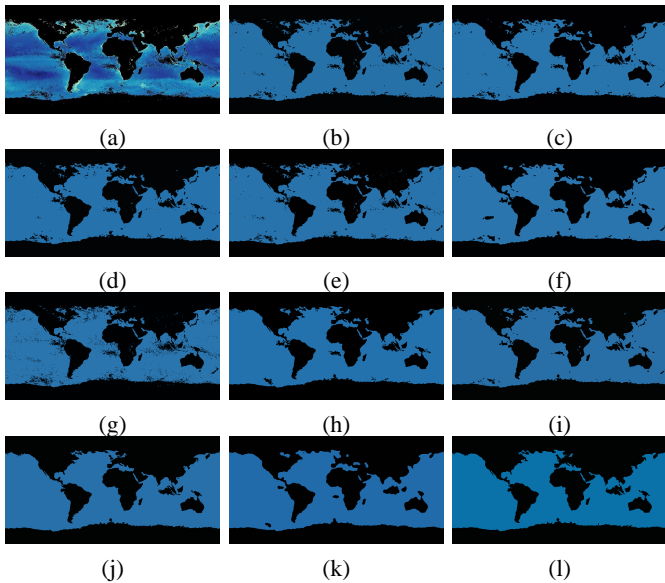


Fig. 15. Segmentation results for the second real-world image in NEO. From (a) to (l): noisy image and results of FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, FRFCM, WFCM, SRFCM, DSFCM\_N, and LRFCM.

Fig. 14 shows the results for segmenting a real-world image showing sea ice and snow extent. The colors represent where the land and ocean are covered by snow and ice per week (here

TABLE V  
SEGMENTATION PERFORMANCE ON REAL-WORLD IMAGES IN NEO

Algorithm	First image			Second image		
	EI	PC	PE	EI	PC	PE
FCM_S1	0.871	0.934	0.558	1.500	0.929	0.499
FCM_S2	0.852	0.942	0.518	1.502	0.928	0.496
FGFCM	0.884	0.938	0.583	1.511	0.924	0.511
FLICM	0.859	0.946	0.508	1.489	0.936	0.479
KWFLICM	0.831	0.954	0.418	1.499	0.932	0.483
ARKFCM	0.829	0.962	0.337	1.485	0.947	0.459
FRFCM	0.815	0.983	0.194	1.370	0.955	0.318
WFCM	0.817	0.981	0.248	1.395	0.950	0.322
SRFCM	0.810	0.991	0.172	1.310	0.963	0.262
DSFCM_N	0.873	0.931	0.574	1.506	0.921	0.503
LRFCM	0.804	0.995	0.099	1.285	0.969	0.214

is February 7–14, 2015). The number of clusters is set to 4. Fig. 15 shows the segmentation results on a real-world image showing chlorophyll concentration. The colors represent where and how much phytoplankton are growing over a span of days. The black areas show where the satellite could not measure phytoplankton. The number of clusters is set to 2. Obviously, most of algorithms, i.e., FCM\_S1, FCM\_S2, FGFCM, FLICM, KWFLICM, ARKFCM, and WFCM, cannot fully suppress unknown noise. Both FRFCM and DSFCM\_N have an aptitude for noise suppression. However, they produce incorrect clusters, as shown in Fig. 14. Even though SRFCM exhibits a good capacity of noise suppression, it forges several topology changes in the form of black patches, as shown in Fig. 15. Compared with its ten peers, LRFCM can remove unknown noise and preserve image contours as well. As Table V illustrates, LRFCM universally achieves the better EI, PC, and PE values than its peers. In particular, the EI values of LRFCM are universally smaller than the others', which indicates that LRFCM acquires better uniformity in segmented images. Note that the EI value of LRFCM is down to 0.804 for the first real-world image illustrated in Fig. 14. Based on the quantitative comparison results, we can conclude that LRFCM outperforms other FCM-related algorithms.

#### H. Computational Complexity and Time

In order to exhibit the segmentation efficiency of LRFCM, we compare its computing overhead with its peers'. The computational complexity of all algorithms are presented in Table VI. We again clarify that  $K$  is the number of image pixels,  $c$  is the number of prototypes,  $t$  is the iteration count,  $\omega$  represents the size of a filtering window,  $|\mathcal{N}_j|$  denotes the size of a local window centralized at pixel  $j$ , and  $\mu$  is the number of pixel levels in an image. Generally,  $\mu \ll K$ .

As Table VI shows, FGFCM and FRFCM have lower computational complexity than its peers due to  $\mu \ll K$ . Except WFCM, SRFCM, and LRFCM, the computational complexity of other algorithms is regarded as  $O(K)$ . Since  $O(K \log K)$  associated with a tight wavelet transform is close to  $O(K)$ , the computational complexity of WFCM, SRFCM, and LRFCM is not high. In addition, to compare the practicability between LRFCM and its peers, we exhibit the computing time (in seconds) of all algorithms for segmenting different images, as summarized in Table VII. To ensure a fair comparison, we note that all experiments are completed with MATLAB

TABLE VI  
COMPUTATIONAL COMPLEXITY OF ALL ALGORITHMS

Algorithm	Computational complexity
FCM_S1	$O(K \times \omega + K \times c \times t)$
FCM_S2	$O(K \times \omega + K \times c \times t)$
FGFCM	$O(K \times \omega + \mu \times c \times t)$
FLICM	$O(K \times c \times t \times  \mathcal{N}_j )$
KWFLICM	$O(K \times (\sqrt{\omega} + 1)^2 + K \times c \times t \times  \mathcal{N}_j )$
ARKFCM	$O(4 \times K \times \omega + K \times c \times t)$
FRFCM	$O(K \times \omega + \mu \times c \times t)$
WFCM	$O(K \times \omega + 9 \times K \times \log K + K \times c \times t)$
SRFCM	$O(K \times \omega + 9 \times K \times \log K + K \times c \times t)$
DSFCM_N	$O(K \times c \times t \times  \mathcal{N}_j )$
LRFCM	$O(K \times \omega + 9 \times K \times \log K + K \times c \times t \times  \mathcal{N}_j )$

running on a laptop with Intel(R) Core(TM) i5-8250U CPU of (1.60 GHz) and 8.0 GB RAM.

TABLE VII  
COMPUTING TIME (IN SECONDS) ON DIFFERENT IMAGES

Algorithm	Fig. 10 column 1	Fig. 10 column 3	Fig. 10 column 5	Fig. 12	Fig. 13 (Average)	Fig. 14	Fig. 15
FCM_S1	4.50	3.90	3.02	3.31	4.80	5.68	4.51
FCM_S2	3.14	4.35	3.17	3.82	5.08	5.10	4.10
FGFCM	2.31	2.39	2.47	2.14	3.66	3.87	2.74
FLICM	55.85	75.30	187.51	107.64	104.41	160.88	42.37
KWFLICM	82.76	103.59	237.42	125.53	160.82	298.44	53.95
ARKFCM	10.60	9.37	6.89	5.70	10.03	23.46	13.37
FRFCM	0.33	0.31	0.28	0.22	1.48	4.97	1.46
WFCM	3.50	3.52	3.57	3.88	6.46	6.21	6.49
SRFCM	3.17	2.62	2.92	3.36	3.73	5.73	3.69
DSFCM_N	16.89	13.86	11.79	10.44	16.32	36.32	18.64
LRFCM	13.47	11.74	9.02	8.88	13.41	32.08	15.02

As Table VII indicates, for gray image segmentation, all algorithms have acceptable computing overheads. Among them, FLICM and KWFLICM are time-consuming. In contrast, FGFCM and FRFCM take the least time due to the usage of gray level histograms. LRFCM runs faster than DSFCM\_N, FLICM and KWFLICM due to few iterations. When copying with color image segmentation, all algorithms incur more computational costs. Obviously, the computing time of FLICM and KWFLICM is much higher than those of other algorithms. FGFCM and FRFCM still exhibit high computational efficiency. Although LRFCM is slightly slower than comparative algorithms except DSFCM\_N, FLICM and KWFLICM, its good performance makes up for this shortcoming.

## V. CONCLUSIONS

In this work, we propose a comprehensive FCM-related algorithm for image segmentation by taking advantage of various mathematical tools. By preprocessing an observed image by using MR, its weighted sum image with good properties is first generated. With the use of tight wavelet frames, the feature set associated with the weighted sum image is taken as data for clustering, which is adaptive for the analysis of image data. More importantly, an  $\ell_0$  regularization-based FCM algorithm is proposed, which implies that the favorable estimation of the residual is obtained and the ideal value participates in clustering. In fact, the sparsity imposed on the residual is beneficial to acquire more suitable estimation, which is positive for the segmentation performance. As a post-processing step, MR is also applied to filter the obtained labels,

which implies that the performance of the  $\ell_0$  regularization-based FCM is improved. Finally, many supporting experiments are conducted to show that the proposed algorithm is superior to other FCM-related algorithms even though its running time is slightly larger than the time required by a couple of compared algorithms.

Although experimental results illustrate that the proposed algorithm is effective and practical, there exist some open issues worth pursuing. For example, can one apply the proposed algorithm to non-flat domains, such as remote sensing [50], computer networks, ecological systems [51], and transportation networks [52]? How to automatically select the numbers of clusters is another research topic?

## APPENDIX PROOF OF THEOREM III.1

Let us consider a minimization problem

$$r_{jl} = \arg \min_{r_{jl}} \Gamma = \sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m \|x_{jl} - r_{jl} - v_{il}\|^2}{1 + d_{nj}} + \sum_{n \in \mathcal{N}_j} \frac{\beta_l |r_{jl}|_0}{1 + d_{nj}}. \quad (25)$$

According to (9), we rewrite the energy function  $\Gamma$  as:

$$\Gamma = \begin{cases} \sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m \|x_{jl} - r_{jl} - v_{il}\|^2}{1 + d_{nj}} + \sum_{n \in \mathcal{N}_j} \frac{\beta_l}{1 + d_{nj}}, & r_{jl} \neq 0 \\ \sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m \|x_{jl} - v_{il}\|^2}{1 + d_{nj}}, & r_{jl} = 0 \end{cases}. \quad (26)$$

By introducing two variables  $a$  and  $b$ , we have

$$a = \frac{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m (x_{jl} - v_{il})}{1 + d_{nj}}}{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m}{1 + d_{nj}}}, \quad b = \sum_{n \in \mathcal{N}_j} \frac{\beta_l}{1 + d_{nj}}.$$

Thus, (26) can be rewritten as

$$\Gamma = \begin{cases} (r_{jl} - a)^2 + b, & r_{jl} \neq 0 \\ a^2, & r_{jl} = 0 \end{cases}. \quad (27)$$

For the first subproblem in (27), we have  $\Gamma_{\min} = b$  when  $r_{jl} = a$ . For the second subproblem in (27), we have  $\Gamma_{\min} = a^2$  when  $r_{jl} = 0$ . To sum up, for minimizing (27), we have

$$\Gamma_{\min} = \begin{cases} b, & a^2 \geq b \\ a^2, & a^2 < b \end{cases}, \quad r_{jl} = \begin{cases} a, & a^2 \geq b \\ 0, & a^2 < b \end{cases}. \quad (28)$$

Here,  $r_{jl}$  is reformulated as

$$r_{jl} = \begin{cases} a, & a \geq \sqrt{b} \\ 0, & a < \sqrt{b} \end{cases}. \quad (29)$$

By introducing a hard-thresholding operator  $\mathcal{H}$  expressed by (21), (29) is rewritten as

$$r_{jl} = \mathcal{H}_b(a).$$

## REFERENCES

- [1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- [2] X. Zhu, W. Pedrycz, and Z. Li, "Granular encoders and decoders: a study in processing information granules," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 5, pp. 1115–1126, Oct. 2017.
- [3] 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.



- [4] 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.
- [5] M. Ahmed, S. Yamany, N. Mohamed, 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, Aug. 2002.
- [6] 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. Part B Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.
- [7] L. Szilagyi, Z. Benyo, S. Szilagyi, and H. Adam, "MR brain image segmentation using an enhanced fuzzy C-means algorithm," in *Proc. 25th Annu. Int. Conf. IEEE EMBS*, Sep. 2003, pp. 724–726.
- [8] 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.
- [9] 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, Jan. 2010.
- [10] 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.
- [11] A. Elazab, C. Wang, F. Jia, J. Wu, G. Li, and Q. Hu, "Segmentation of brain tissues from magnetic resonance images using adaptively regularized kernel-based fuzzy-means clustering," *Comput. Math. Method. M.*, vol. 2015, pp. 1–12, Nov. 2015. [Online]. Available: <http://dx.doi.org/10.1155/2015/485495>.
- [12] 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, Jan. 2013.
- [13] F. Guo, X. Wang, and J. Shen, "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.
- [14] Z. Zhao, L. Cheng, and G. Cheng, "Neighbourhood weighted fuzzy c-means clustering algorithm for image segmentation," *IET Image Process.*, vol. 8, no. 3, pp. 150–161, Mar. 2014.
- [15] C. Wang, W. Pedrycz, J. Yang, M. Zhou, and Z. Li, "Wavelet frame-based Fuzzy C-Means clustering for segmenting images on graphs," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3938–3949, Sep. 2020.
- [16] K. P. Lin, "A novel evolutionary kernel intuitionistic fuzzy C-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1074–1087, Aug. 2014.
- [17] X. Zhu, W. Pedrycz, and Z. W. Li, "Fuzzy clustering with nonlinearly transformed data," *Appl. Soft Comput.*, vol. 61, pp. 364–376, Dec. 2017.
- [18] R. R. Gharieb, G. Gendy, A. Abdelfattah, and H. Selim, "Adaptive local data and membership based KL divergence incorporating C-means algorithm for fuzzy image segmentation," *Appl. Soft Comput.*, vol. 59, pp. 143–152, Oct. 2017.
- [19] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and K. N. Asoke, "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.
- [20] 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.
- [21] C. Wang, W. Pedrycz, M. Zhou, and Z. Li, "Sparse regularization-based Fuzzy C-Means clustering incorporating morphological grayscale reconstruction and wavelet frames," *IEEE Trans. Fuzzy Syst.*, to be published, doi: 10.1109/TFUZZ.2020.2985930.
- [22] 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.
- [23] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 12, pp. 1163–1173, Dec. 1996.
- [24] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.
- [25] J. F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: total variation, wavelet frames, and beyond," *J. Amer. Math. Soc.*, vol. 25, no. 4, pp. 1033–1089, May 2012.
- [26] B. Dong and Z. Shen, "MRA-based wavelet frames and applications," in *The Mathematics of Image Processing* (IAS Lecture Notes Series). Salt Lake City, UT, USA: Park City Mathematics Institute, 2010.
- [27] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy C-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2-3, pp. 191–203, 1984.
- [28] L. Liu, A. Yang, W. Zhou, X. Zhang, M. Fei, and X. Tu, "Robust dataset classification approach based on neighbor searching and kernel fuzzy c-means," *IEEE/CAA J. Autom. Sinica*, vol. 2, no. 3, pp. 235–247, Jul. 2015.
- [29] B. Dong, "Sparse representation on graphs by tight wavelet frames and applications," *Appl. Comput. Harmon. Anal.*, vol. 42, no. 3, pp. 452–479, May 2017.
- [30] C. Wang, Z. Yan, W. Pedrycz, M. Zhou, and Z. Li, "A weighted fidelity and regularization-based method for mixed or unknown noise removal from images on graphs," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 5229–5243, Dec. 2020.
- [31] C. Wang and J. Yang, "Poisson noise removal of images on graphs using tight wavelet frames," *Visual Comput.*, vol. 34, no. 10, pp. 1357–1369, Oct. 2018.
- [32] D. Ma and C. Wang, "Removal of mixed Gaussian and impulse noise using data-driven tight frames," *J. Eng. Sci. Technol. Rev.*, vol. 11, no. 2, pp. 26–31, Mar. 2018.
- [33] B. Dong, Q. T. Jiang, C. Q. Liu, and Z. Shen, "Multiscale representation of surfaces by tight wavelet frames with applications to denoising," *Appl. Comput. Harmon. Anal.*, vol. 41, no. 2, pp. 561–589, Sep. 2016.
- [34] J. Yang and C. Wang, "A wavelet frame approach for removal of mixed Gaussian and impulse noise on surfaces," *Inverse Probl. Imaging*, vol. 11, no. 5, pp. 783–798, Oct. 2017.
- [35] A. Ron and Z. Shen, "Affine systems in  $L_2(\mathbb{R}^d)$ : The analysis of the analysis operator," *J. Funct. Anal.*, vol. 148, no. 2, pp. 408–447, Aug. 1997.
- [36] Y. Gong and I. F. Sbalzarini, "Curvature filters efficiently reduce certain variational energies," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1786–1798, Apr. 2017.
- [37] G. Treece, "The bitonic filter: linear filtering in an edge-preserving morphological framework," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5199–5211, Nov. 2016.
- [38] J. Chen, C. Su, W. Grimson, J. Liu, and D. Shiue, "Object segmentation of database images by dual multiscale morphological reconstructions and retrieval applications," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 828–843, Feb. 2012.
- [39] C. Chau, P. Combettes, J. C. Pesquet, and V. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Probl.*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.
- [40] P. Combettes and V. Wajs, "Signal recovery by proximal forward backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [41] B. Dong, Y. Zhang, "An efficient algorithm for  $\ell_0$  minimization in wavelet frame based image restoration," *J. Sci. Comput.*, vol. 54, pp. 350–368, 2013.
- [42] D. L. Donoho and I. M. Johnstone, "Threshold selection for wavelet shrinkage of noisy data," in *Proc. 16th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, Nov. 1994, pp. A24–A25.
- [43] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Med. Imaging*, vol. 15, no. 29, pp. 1–29, Aug. 2015.
- [44] H. Zhang, J. Fritts, and S. Goldman, "An entropy-based objective evaluation method for image segmentation," in *Proc. SPIE, Storage Retrieval Methods Appl. Multimedia*, vol. 5307, Jan. 2004, pp. 38–49.
- [45] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 841–847, Aug. 1991.
- [46] O. Arbelaitz, I. Gurrutxaga, J. M. Perez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognit.*, vol. 46, no. 1, pp. 243–256, Jan. 2013.
- [47] J. Yang, G. Zhu, D. Tong, L. Lu, and Z. Shen, "B-spline tight frame based force matching method," *J. Comput. Phys.*, vol. 362, no. 1, pp. 208–219, Jun. 2018.
- [48] 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.
- [49] T. Lei, Y. Zhang, Y. Wang, S. Liu, and Z. Guo, "A conditionally invariant mathematical morphological framework for color images," *Inf. Sci.*, vol. 387, pp. 34–52, May 2017.
- [50] T. Xu, L. Jiao, and W. J. Emery, "SAR image content retrieval based on fuzzy similarity and relevance feedback," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1824–1842, May 2017.
- [51] C. Wang, J. Chen, Z. Li, E. Nasr, and A. M. El-Tamimi, "An indicator system for evaluating the development of land-sea coordination systems:

A case study of Lianyungang port,” *Ecol. Indic.*, vol. 98, pp. 112–120, Mar. 2019.

- [52] Y. Lv, Y. Chen, X. Zhang, Y. Duan, and N. Li, “Social media based transportation research: the state of the work and the networking,” *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 1, pp. 19–26, Jan. 2017.



**Cong Wang** received the B.S. degree in automation and the M.S. degree in mathematics from Hohai University, Nanjing, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in mechatronic engineering, Xidian University, Xi'an, China.

He was a Visiting Ph.D. Student in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He was also a Research Assistant at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He is currently a Visiting Ph.D. Student in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His current research interests include wavelet analysis and its applications, granular computing, as well as image processing.



**Witold Pedrycz** (Fellow, IEEE) received the M.Sc. degree in computer science and technology, the Ph.D. degree in computer engineering, and the D.Sci. degree in systems science from the Silesian University of Technology, Gliwice, Poland, in 1977, 1980, and 1984, respectively.

He is a Professor and the Canada Research Chair in Computational Intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. He is a foreign member of the Polish Academy of Sciences. He has authored 15 research monographs covering various aspects of computational intelligence, data mining, and software engineering. His current research interests include computational intelligence, fuzzy modeling, and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering. He has published numerous papers in the above areas.

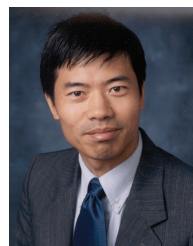
Dr. Pedrycz was a recipient of the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is intensively involved in editorial activities. He is an Editor-in-Chief of *Information Sciences*, an Editor-in-Chief of *WIREs Data Mining and Knowledge Discovery* (Wiley) and the *International Journal of Granular Computing* (Springer). He currently serves as a member of a number of editorial boards of other international journals. He is a fellow of the Royal Society of Canada.



**ZhiWu Li** (Fellow, IEEE) received the B.S. degree in mechanical engineering, the M.S. degree in automatic control, and the Ph.D. degree in manufacturing engineering from Xidian University, Xi'an, China, in 1989, 1992, and 1995, respectively.

He joined Xidian University in 1992. He is also currently with the Institute of Systems Engineering, Macau University of Science and Technology, Macau, China. He was a Visiting Professor with the University of Toronto, Toronto, ON, Canada, the Technion-Israel Institute of Technology, Haifa, Israel, the Martin-Luther University of Halle-Wittenburg, Halle, Germany, Conservatoire National des Arts et Métiers, Paris, France, and Meliksah Universitesi, Kayseri, Turkey. His current research interests include Petri net theory and application, supervisory control of discrete-event systems, system reconfiguration, game theory, and data and process mining.

Dr. Li was a recipient of an Alexander von Humboldt Research Grant, Alexander von Humboldt Foundation, Germany. He is listed in *Marquis Who's Who in the World*, 27th Edition, 2010. He serves as a Frequent Reviewer of 90+ international journals, including *Automatica* and a number of the *IEEE TRANSACTIONS* as well as many international conferences. He is the Founding Chair of Xi'an Chapter of IEEE Systems, Man, and Cybernetics Society. He is a member of Discrete-Event Systems Technical Committee of the IEEE Systems, Man, and Cybernetics Society and IFAC Technical Committee on Discrete-Event and Hybrid Systems, from 2011 to 2014.



**MengChu Zhou** (Fellow, IEEE) received his B.S. degree in Control Engineering from Nanjing University of Science and Technology, Nanjing, China, in 1983, M.S. degree in Automatic Control from Beijing Institute of Technology, Beijing, China, in 1986, and Ph. D. degree in Computer and Systems Engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990, respectively.

He joined New Jersey Institute of Technology (NJIT), Newark, NJ, in 1990, and is now a Distinguished Professor of Electrical and Computer Engineering. His research interests are in Petri nets, intelligent automation, Internet of Things, big data, web services, and intelligent transportation.

He has over 900 publications including 12 books, 600+ journal papers (450+ in *IEEE TRANSACTIONS*), 26 patents and 29 book-chapters. He is the founding Editor of IEEE Press Book Series on Systems Science and Engineering, Editor-in-Chief of *IEEE/CAA Journal of Automatica Sinica*, and Associate Editor of *IEEE Internet of Things Journal*, *IEEE Transactions on Intelligent Transportation Systems*, and *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. He is a recipient of Humboldt Research Award for US Senior Scientists from Alexander von Humboldt Foundation, Franklin V. Taylor Memorial Award and the Norbert Wiener Award from IEEE Systems, Man and Cybernetics Society, and Excellence in Research Prize and Medal from NJIT. He is a life member of Chinese Association for Science and Technology-USA and served as its President in 1999. He is a Fellow of International Federation of Automatic Control (IFAC), American Association for the Advancement of Science (AAAS) and Chinese Association of Automation (CAA).



**Jun Zhao** (Member, IEEE) is currently an Assistant Professor in the School of Computer Science and Engineering at Nanyang Technological University (NTU) in Singapore. He received a PhD degree in Electrical and Computer Engineering from Carnegie Mellon University (CMU), USA. Before joining NTU first as a postdoc with Xiaokui Xiao and then as a faculty member, he was a postdoc at Arizona State University as an Arizona Computing PostDoc Best Practices Fellow. His research interests include blockchains, security, and privacy with applications

to the Internet of Things and deep learning. One of his first-authored papers was shortlisted for the best student paper award in IEEE International Symposium on Information Theory (ISIT) 2014, a prestigious conference in information theory.