

Residual-driven Fuzzy C-Means Clustering for Image Segmentation

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

Abstract—In this paper, we elaborate residual-driven Fuzzy C-Means (FCM) for image segmentation, which is the first approach that realizes accurate residual (noise/outliers) estimation and makes noise-free image participate in clustering. We propose a residual-driven FCM framework by integrating into FCM a residual-related regularization term derived from the distribution characteristic of different types of noises. Built on this framework, a weighted ℓ_2 -norm regularization term is presented by weighting mixed noise distribution, thus resulting in a universal residual-driven FCM algorithm in presence of mixed or unknown noises. Besides, with the constraint of spatial information, the residual estimation becomes more reliable than that only considering an observed image itself. Supporting experiments on synthetic, medical, and real-world images are conducted. The results demonstrate the superior effectiveness and efficiency of the proposed algorithm over its peers.

Index Terms—Fuzzy C-Means, mixed or unknown noise, residual-driven, weighted regularization, image segmentation.

I. INTRODUCTION

As an important approach to data analysis and processing, fuzzy clustering has been widely applied to a number of visible domains such as pattern recognition [1], [2], data mining [3], granular computing [4], and image processing [5]. One of the most popular fuzzy clustering methods is a Fuzzy C-Means (FCM) algorithm [6]–[8]. It plays a significant role in image segmentation; yet it only works well for noise-free images. In real-world applications, images are often contaminated by different types of noises, especially mixed or unknown noises, produced in the process of image acquisition and transmission. Therefore, to make FCM robust to noise, FCM is refined resulting in many modified versions

in two main means, i.e., introducing spatial information into its objective function [9]–[14] and substituting its Euclidean distance with a kernel distance (function) [15]–[22]. Even though such versions improve its robustness to some extent, they often fail to account for high computing overhead of clustering. To balance the effectiveness and efficiency of clustering, researchers have recently attempted to develop FCM with the aid of mathematical technologies such as Kullback-Leibler divergence [23], [24], sparse regularization [25], [26], morphological reconstruction [24], [27]–[29] and gray level histograms [30], [31], as well as pre-processing and post-processing steps like image pixel filtering [32], membership filtering [30] and label filtering [26], [32], [33]. To sum up, the existing studies make evident efforts to improve its robustness mainly by means of noise removal in each iteration or before and after clustering. However, they fail to take accurate noise estimation into account and apply it to improve FCM.

Generally speaking, noise can be modeled as the residual between an observed image and its ideal value (noise-free image). Clearly, its accurate estimation is beneficial for image segmentation as noise-free image instead of observed one can then be used in clustering. Most of FCM-related algorithms suppress the impact of such residual on FCM by virtue of spatial information. So far, there are no studies focusing on developing FCM variants based on an in-depth analysis and accurate estimation of the residual. To the best of our knowledge, there is only one attempt [34] to improve FCM by revealing the sparsity of the residual. To be specific, since a large proportion of image pixels have small or zero noise/outliers, ℓ_1 -norm regularization can be used to characterize the sparsity of the residual, thus forming deviation-sparse FCM (DSFCM). When spatial information is used, it upgrades to its augmented version, named as DSFCM_N. Their residual estimation is realized by using a soft thresholding operation. In essence, such estimation is only applicable to impulse noise having sparsity. Therefore, neither of them can achieve highly accurate residual estimation in presence of different noises.

Motivated by [34], to address a wide range of noise types, we elaborate residual-driven FCM (RFCM) for image segmentation, which furthers FCM's performance. We first design an RFCM framework, as shown in Fig. 1(b), by introducing a regularization term on residual as a part of the objective function of FCM. This term makes residual accurately estimated. It is determined by a noise distribution, e.g., an ℓ_2 -norm regularization term corresponds to Gaussian noise and an ℓ_1 -norm one suits impulse noise.

In real-world applications, since images are often corrupted by mixed or unknown noises, a specific noise distribution is

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).

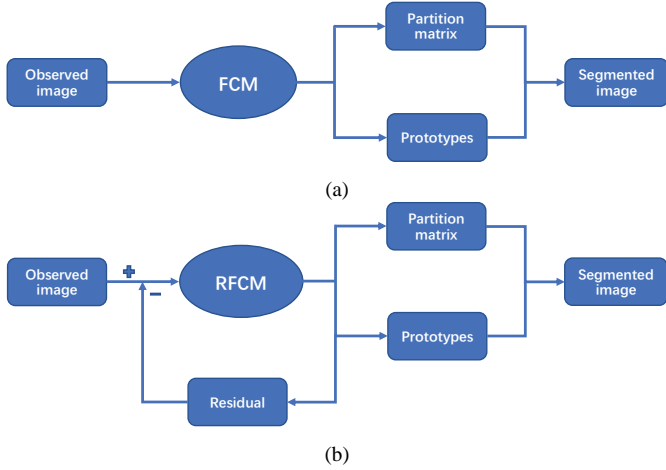


Fig. 1. A comparison between the frameworks of FCM and RFCM. (a) FCM; and (b) RFCM.

difficult to be obtained. To deal with this issue, by analyzing the distribution of a wide range of mixed noises, especially a mixture of Poisson, Gaussian and impulse noises, we present a weighted ℓ_2 -norm regularization term in which each residual is assigned a weight, thus resulting in an augmented version namely WRFCM for image segmentation with mixed or unknown noises. To obtain better noise suppression, we also consider spatial information of image pixels in WRFCM since it is naturally encountered in image segmentation. In addition, we design a two-step iterative algorithm to minimize the objective function of WRFCM. The first step is to employ the Lagrangian multiplier method to optimize the partition matrix, prototypes and residual when fixing the assigned weights. The second step is to update the weights by using the calculated residual. Finally, based on the optimal partition matrix and prototypes, a segmented image is obtained.

This study makes fourfold contributions to advance FCM for image segmentation:

- For the first time, we propose an RFCM framework for image segmentation by introducing a regularization term derived from a noise distribution into FCM. It relies on accurate residual estimation to greatly improve FCM's performance, which is absent from existing FCMs.
- Built on the RFCM framework, WRFCM is presented by weighting mixed noise distribution and incorporating spatial information. The use of spatial information makes resulting residual estimation more reliable. It is regarded as a universal RFCM algorithm for coping with mixed or unknown noises.
- We design a two-step iterative algorithm to realize WRFCM. Since only ℓ_2 vector norm is involved, it is fast by virtue of a Lagrangian multiplier method.
- WRFCM is validated to produce state-of-the-art performance on synthetic, medical and real-world images from four benchmark databases.

The originality of this work comes with a realization of accurate residual estimation from observed images, which benefits FCM's performance enhancement. It achieves more accurate residual estimation than DSFCM and DSFCM_N

introduced [34] since it is modeled by an analysis of noise distribution characteristic replacing noise sparsity. In essence, the proposed algorithm is an unsupervised method. Compared with commonly used supervised methods such as convolutional neural networks (CNNs) [35]–[39] and dictionary learning [40], [41], it realizes the residual estimation precisely by virtue of a regularization term rather than using any image samples to train a residual estimation model. Hence, it needs low computing overhead and can be experimentally executed by using a low-end CPU rather than a high-end GPU, which means that its practicality is high. In addition, it is free of the aid of mathematical techniques and achieves the superior performance over some recently proposed comprehensive FCMs. Therefore, we conclude that WRFCM is a fast and robust FCM algorithm. Finally, its minimization problem involves an ℓ_2 vector norm only. Thus it can be easily solved by using a well-known Lagrangian multiplier method.

Section II details the conventional FCM and the proposed methodology. Section III reports experimental results. Conclusions and some open issues are given in Section IV.

II. FCM AND PROPOSED METHODOLOGY

A. Fuzzy C-Means (FCM)

Given is a set $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^L : j = 1, 2, \dots, K\}$, where \mathbf{x}_j contains L channels, i.e., $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jL})^T$ with T being the transpose of a vector. FCM is applied to cluster \mathbf{X} by minimizing:

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

where $\mathbf{U} = [u_{ij}]_{c \times K}$ is a partition matrix under a constraint $\sum_{i=1}^c u_{ij} = 1$ for $j = 1, 2, \dots, K$, $\mathbf{V} = \{\mathbf{v}_i : i = 1, 2, \dots, c\}$ is a prototype set, $\|\cdot\|$ stands for the Euclidean distance, and m denotes a fuzzification exponent ($m > 1$).

An alternating iteration scheme [8] is used to minimize (1). Each iteration is realized as follows:

$$u_{ij}^{(t+1)} = \frac{(\|\mathbf{x}_j - \mathbf{v}_i^{(t)}\|^2)^{-1/(m-1)}}{\sum_{q=1}^c (\|\mathbf{x}_j - \mathbf{v}_q^{(t)}\|^2)^{-1/(m-1)}}$$

$$v_{il}^{(t+1)} = \frac{\sum_{j=1}^K (u_{ij}^{(t+1)})^m x_{jl}}{\sum_{j=1}^K (u_{ij}^{(t+1)})^m}$$

Here, $t = 0, 1, 2, \dots$ is an iterative step and $l = 1, 2, \dots, L$. By presetting a threshold ε , the procedure stops when $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$.

B. Noise Model

Consider an observed image \mathbf{X} with K pixels. It is denoted as $\mathbf{X} = \{\mathbf{x}_j : j = 1, 2, \dots, K\}$, where $\mathbf{x}_j = \{x_{jl} : l = 1, 2, \dots, L\}$. When $L = 1$, \mathbf{X} represents a gray image. For $L = 3$, \mathbf{X} is a Red-Green-Blue color image. Since there is

noise in an observed image, \mathbf{X} can be modeled as a sum of a noise-free image $\widetilde{\mathbf{X}}$ and noise \mathbf{R} :

$$\mathbf{X} = \widetilde{\mathbf{X}} + \mathbf{R} \quad (2)$$

Mathematically speaking, $\widetilde{\mathbf{X}} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_K\}$ is an ideal value of \mathbf{X} and thus is unknown. $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_K\}$ is viewed as the residual between \mathbf{X} and $\widetilde{\mathbf{X}}$. Its accurate estimation can make $\widetilde{\mathbf{X}}$ instead of \mathbf{X} participate in clustering so as to improve FCM's robustness. Hence, it is a necessary step to formulate a noise model before constructing an FCM model. In image processing, the models of single noise such as Gaussian, Poisson and impulse noise are widely used. In this work, in order to construct robust FCM, we mostly consider mixed or unknown noise since it is often encountered in real-world applications. Its specific model is unfortunately hard to be formulated. Therefore, a common solution is to assume the type of mixed noise in advance. In universal image processing, two kinds of mixed noises are the most common, namely mixed Poisson-Gaussian noise and mixed Gaussian and impulse noise. Beyond them, we focus on a mixture of a wide range of noises, i.e., a mixture of Poisson, Gaussian, and impulse noise. We investigate an FCM-related model based on the analysis of the mixed noise model and extend it to image segmentation with mixed or unknown noises.

Formally speaking, a noise-free image $\widetilde{\mathbf{X}}$ is defined in a domain $\Omega = \{1, 2, \dots, K\}$. It is first corrupted by Poisson noise, thus resulting in $\overline{\mathbf{X}} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_K\}$ that obeys a Poisson distribution, or, $\overline{\mathbf{X}} \sim \mathbf{P}(\widetilde{\mathbf{X}})$. Then additive zero-mean white Gaussian noise $\mathbf{R}' = \{\mathbf{r}'_1, \mathbf{r}'_2, \dots, \mathbf{r}'_K\}$ with standard deviation σ is added. Finally, impulse noise $\mathbf{R}'' = \{\mathbf{r}''_1, \mathbf{r}''_2, \dots, \mathbf{r}''_K\}$ with a given probability $p \in (0, 1)$ is imposed. Hence, for $j \in \Omega$, an arbitrary element in observed image \mathbf{X} is expressed as

$$x_j = \begin{cases} \bar{x}_j + \mathbf{r}'_j & j \in \Omega_1 \\ \mathbf{r}''_j & j \in \Omega_2 := \Omega \setminus \Omega_1 \end{cases} \quad (3)$$

where Ω_1 is called an observable region including mixed Poisson-Gaussian noise and Ω_2 denotes the region consisting of the missing information of $\overline{\mathbf{X}} + \mathbf{R}'$ and is assumed to be unknown with each element being drawn from the whole region Ω by Bernoulli trial with p . In image segmentation, mixed noise model (3) is for the first time presented.

C. Residual-driven FCM

Since there exists an unknown noise intensity in an observed image, the segmentation accuracy of FCM is greatly impacted without properly handling it. It is natural to understand that taking a noise-free image (the ideal value of an observed image) as data to be clustered can achieve better segmentation effects. In other words, if noise (residual) can be accurately estimated, the segmentation effects of FCM should be greatly improved. To do so, we introduce a regularization term on residual into the objective function of FCM. Consequently, an RFCM framework is first presented:

$$J(\mathbf{U}, \mathbf{V}, \mathbf{R}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2 + \beta \cdot \Gamma(\mathbf{R}) \quad (4)$$

where $\beta = \{\beta_l : l = 1, 2, \dots, L\}$ is a parameter set, which controls the impact of regularization term $\Gamma(\mathbf{R})$ on FCM. We rewrite \mathbf{R} as $\{\mathbf{R}_l : l = 1, 2, \dots, L\}$ with $\mathbf{R}_l = (r_{1l}, r_{2l}, \dots, r_{Kl})^T$, which indicates that \mathbf{R} has L channels and each of them contains K pixels. In this work, $L = 1$ (gray) or 3 (Red-Green-Blue). From a channel perspective, we have:

$$\beta \cdot \Gamma(\mathbf{R}) = \sum_{l=1}^L \beta_l \Gamma(\mathbf{R}_l) \quad (5)$$

The regularization term $\Gamma(\mathbf{R})$ guarantees that the solution accords with the degradation process of the minimization of (4). It is determined by a specified noise distribution. For example, when considering the Gaussian noise estimation, we use an ℓ_2 -norm regularization term:

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_2}^2 = \sum_{j=1}^K |r_{jl}|^2$$

where $\|\cdot\|_{\ell_2}$ stands for an ℓ_2 vector norm and $|\cdot|$ denotes an absolute value operation. In the presence of the impulse noise, we choose an ℓ_1 -norm regularization term:

$$\Gamma(\mathbf{R}_l) = \|\mathbf{R}_l\|_{\ell_1} = \sum_{j=1}^K |r_{jl}|$$

where $\|\cdot\|_{\ell_1}$ denotes an ℓ_1 vector norm. For the Poisson noise, we take the Csiszár's I-divergence [42] of \mathbf{R} from \mathbf{X} as a regularization term, i.e.,

$$\Gamma(\mathbf{R}_l) = \sum_{j=1}^K ((x_{jl} - r_{jl}) - x_{jl} \ln(x_{jl} - r_{jl}))$$

For a common single noise, i.e., Gaussian, Poisson, and impulse noise, the above regularization terms lead to a maximum a posteriori (MAP) solution to such noise estimations. In real-world applications, images are generally contaminated by mixed or unknown noises rather than a single noise. The regularization terms for single noise estimation become inapplicable since the distribution of mixed or unknown noise is difficult to be modeled mathematically. Therefore, one of the main purposes of this work is to design a universal regularization term for mixed or unknown noise estimation.

D. Analysis of Mixed Noise Distribution

To reveal the essence of mixed noise distributions, we here consider generic and representative mixed noise, i.e., a mixture of Poisson, Gaussian, and impulse noise. Let us take an example to exhibit its distribution. Here, we impose Gaussian noise ($\sigma = 10$) and a mixture of Poisson, Gaussian ($\sigma = 10$) and random-valued impulse noise ($p = 20\%$) on image 'Lena' with size 512×512 , respectively. We show original and two observed images in Fig. 2.

As Fig. 2(b) shows, Gaussian noise is overall orderly. As a common sense, Poisson distribution is a Gaussian-like one under the condition of enough samples. Therefore, due to impulse noise, mixed noise is disorderly and unsystematic as shown in Fig. 2(c). In Fig. 3, we portray the distributions of Gaussian and mixed noise, respectively.

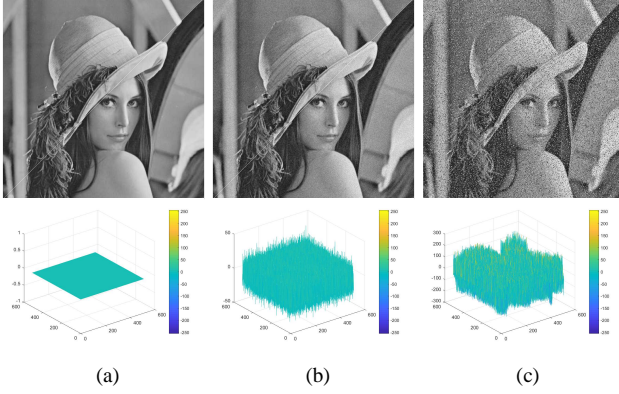


Fig. 2. Noise-free image and two observed ones corrupted by Gaussian and mixed noise, respectively. The first row: (a) noise-free image; (b) observed image with Gaussian noise; and (c) observed image with mixed noise. The second row portrays noise included in three images.

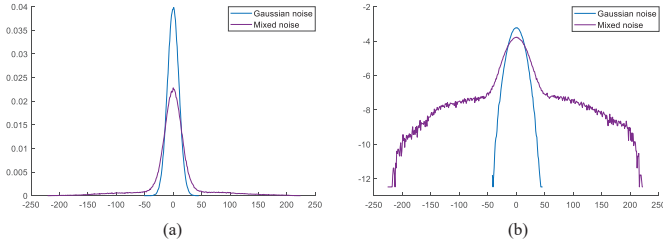


Fig. 3. Distributions of Gaussian and mixed noise in different domains. (a) linear domain; and (b) logarithmic domain.

Fig. 3(a) shows noise distribution in a linear domain. To illustrate a heavy tail intuitively, we present it in a logarithmic domain as shown in Fig. 3(b). Clearly, Poisson noise leads to a Gaussian-like distribution. Nevertheless, impulse noise gives rise to a more irregular distribution with a heavy tail. Therefore, neither ℓ_1 norm nor ℓ_2 norm can precisely characterize the residual \mathbf{R} in the sense of the MAP estimation.

E. Residual-driven FCM with Weighted ℓ_2 -norm Regularization

Intuitively, if the regularization term can be modified so as to make mixed noise distribution more Gaussian-like, we can still use ℓ_2 norm to characterize residual \mathbf{R} . It means that mixed noise can be more accurately estimated. Therefore, we adopt robust estimation techniques [43], [44] to weaken the heavy tail, which makes mixed noise distribution more regular. In the sequel, we assign a proper weight w_{jl} to each residual r_{jl} , which forms a weighted residual $w_{jl}r_{jl}$ that almost obeys a Gaussian distribution. Given Fig. 4, we use an example for showing the effect of weighting.

Fig. 4(a) shows the distribution of r_{jl} and the fitting Gaussian function based on the variance of r_{jl} . Fig. 4(b) gives the distribution of $w_{jl}r_{jl}$ and the fitting Gaussian function based on the variance of $w_{jl}r_{jl}$. Clearly, the distribution of $w_{jl}r_{jl}$ in Fig. 4(b) is more Gaussian-like than that in Fig. 4(a), which means that ℓ_2 -norm regularization can work on weighted residual $w_{jl}r_{jl}$ for an MAP-like solution of \mathbf{R} .

By analyzing Fig. 4, for $l = 1, 2, \dots, L$, we propose a weighted ℓ_2 -norm regularization term for mixed or unknown

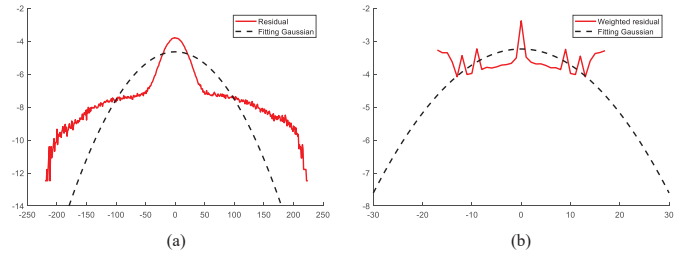


Fig. 4. Distributions of residual r_{jl} and weighted residual $w_{jl}r_{jl}$, as well as the fitting Gaussian function in the logarithmic domain.

noise estimation:

$$\Gamma(\mathbf{R}_l) = \|\mathbf{W}_l \circ \mathbf{R}_l\|_{\ell_2}^2 = \sum_{j=1}^K |w_{jl}r_{jl}|^2 \quad (6)$$

where \circ performs element-by-element multiplication of $\mathbf{R}_l = (r_{1l}, r_{2l}, \dots, r_{Kl})^T$ and $\mathbf{W}_l = (w_{1l}, w_{2l}, \dots, w_{Kl})^T$. For $l = 1, 2, \dots, L$, \mathbf{W}_l makes up a weight matrix $\mathbf{W} = [w_{jl}]_{K \times L}$. Each element w_{jl} is assigned to location (j, l) . Since it is inversely proportional to residual r_{jl} , it can be automatically determined. In this work, we adopt the following expression:

$$w_{jl} = e^{-\xi r_{jl}^2} \quad (7)$$

where ξ is a positive parameter, which aims to control the decreasing rate of w_{jl} .

By substituting (6) into (4) combined with (5), we present RFCM with weighted ℓ_2 -norm regularization (WRFCM) for image segmentation:

$$\begin{aligned} J(\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{W}) &= \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2 + \sum_{l=1}^L \beta_l \|\mathbf{W}_l \circ \mathbf{R}_l\|_{\ell_2}^2 \\ &= \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2 + \sum_{l=1}^L \beta_l \sum_{j=1}^K |w_{jl}r_{jl}|^2 \end{aligned} \quad (8)$$

When coping with image segmentation problems, since each image pixel is closely related to its neighbors, using spatial information has a positive impact on FCM as shown in [9], [34]. If there exists a small distance between a target pixel and its neighbors, they most likely belong to a same cluster. Therefore, we introduce spatial information into (8), thus resulting in our final objective function:

$$\begin{aligned} J(\mathbf{U}, \mathbf{V}, \mathbf{R}, \mathbf{W}) &= \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}} \\ &\quad + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|w_{nl}r_{nl}|^2}{1 + d_{nj}} \end{aligned} \quad (9)$$

Its minimization is completed subject to

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

In (9), an image pixel is sometimes loosely represented by its corresponding index even though this is not ambiguous. Thus, n is a neighbor pixel of j and d_{nj} represents the Euclidean distance between n and j . \mathcal{N}_j stands for a local window centralized at j including j and its size is denoted as $|\mathcal{N}_j|$.

F. Minimization Algorithm

Minimizing (9) involves four unknowns, i.e., U , V , R and W . According to (7), W is automatically determined by R . Hence, we can design a two-step iterative algorithm to minimize (9), which fixes W first to solve U , V and R , then uses R to update W . The main task in each iteration is to solve the minimization problem in terms of U , V and R when fixing W . Assume that W is given. We can apply a Lagrangian multiplier method to minimize (9). The Lagrangian function is expressed as

$$\mathcal{L}_\Lambda(U, V, R; W) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m \left(\sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|w_{nl} r_{nl}|^2}{1 + d_{nj}} + \sum_{j=1}^K \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right) \quad (10)$$

where $\Lambda = \{\lambda_j : j = 1, 2, \dots, K\}$ is a set of Lagrangian multipliers. The two-step iterative algorithm for minimizing (9) is realized in Algorithm 1.

Algorithm 1 Two-step iterative algorithm

Given a threshold ε , input $W^{(0)}$. For $t = 0, 1, \dots$, iterate:

Step 1: Find minimizers $U^{(t+1)}$, $V^{(t+1)}$, and $R^{(t+1)}$:

$$(U^{(t+1)}, V^{(t+1)}, R^{(t+1)}) = \arg \min_{U, V, R} \mathcal{L}_\Lambda(U, V, R; W^{(t)}) \quad (11)$$

Step 2: Update the weight matrix $W^{(t+1)}$

If $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$, stop; else update t such that $0 \leq t \uparrow < +\infty$

The minimization problem (11) can be divided into the following three subproblems:

$$\begin{cases} U^{(t+1)} = \arg \min_U \mathcal{L}_\Lambda(U, V^{(t)}, R^{(t)}; W^{(t)}) \\ V^{(t+1)} = \arg \min_V \mathcal{L}_\Lambda(U^{(t+1)}, V, R^{(t)}; W^{(t)}) \\ R^{(t+1)} = \arg \min_R \mathcal{L}_\Lambda(U^{(t+1)}, V^{(t+1)}, R; W^{(t)}) \end{cases} \quad (12)$$

Each subproblem in (12) has a closed-form solution. We use an alternative optimization scheme similar to the one used in FCM to optimize U and V . The following result is needed to obtain the iterative updates of U and V .

Theorem II.1. Consider the first two subproblems of (12). By applying the Lagrangian multiplier method to solve them, the iterative solutions are presented as

$$u_{ij}^{(t+1)} = \frac{\left(\sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n^{(t)} - \mathbf{v}_i^{(t)}\|^2}{1 + d_{nj}} \right)^{-1/(m-1)}}{\sum_{q=1}^c \left(\sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n^{(t)} - \mathbf{v}_q^{(t)}\|^2}{1 + d_{nj}} \right)^{-1/(m-1)}} \quad (13)$$

$$\mathbf{v}_i^{(t+1)} = \frac{\sum_{j=1}^K \left(\left(u_{ij}^{(t+1)} \right)^m \sum_{n \in \mathcal{N}_j} \frac{\mathbf{x}_n - \mathbf{r}_n^{(t)}}{1 + d_{nj}} \right)}{\sum_{j=1}^K \left(\left(u_{ij}^{(t+1)} \right)^m \sum_{n \in \mathcal{N}_j} \frac{1}{1 + d_{nj}} \right)} \quad (14)$$

Proof: See Appendix. ■

In the last subproblem of (12), both \mathbf{r}_j and \mathbf{r}_n appear simultaneously. Since \mathbf{r}_n is dependent on \mathbf{r}_j , it should not be considered as a constant vector. In other words, n is one of neighbors of j while j is one of neighbors of n symmetrically. Thus, $n \in \mathcal{N}_j$ is equivalent to $j \in \mathcal{N}_n$. Then we have:

$$\sum_{j=1}^K u_{ij}^m \left(f(\mathbf{r}_j) + \sum_{\substack{n \in \mathcal{N}_j \\ n \neq j}} f(\mathbf{r}_n) \right) = \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} u_{in}^m (f(\mathbf{r}_j)) \quad (15)$$

where f represents a function in terms of \mathbf{r}_j or \mathbf{r}_n . By (15), we rewrite (9) as

$$J(U, V, R, W) = \sum_{i=1}^c \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{u_{in}^m \|\mathbf{x}_j - \mathbf{r}_j - \mathbf{v}_i\|^2}{1 + d_{nj}} + \sum_{l=1}^L \beta_l \sum_{j=1}^K \sum_{n \in \mathcal{N}_j} \frac{|w_{jl} r_{jl}|^2}{1 + d_{nj}} \quad (16)$$

According to the two-step iterative algorithm, we assume that W in (16) is fixed in advance. When U and V are updated, the last subproblem of (12) is separable and can be decomposed into $K \times L$ subproblems:

$$r_{jl}^{(t+1)} = \arg \min_{r_{jl}} \sum_{i=1}^c \left(\sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m \|\mathbf{x}_{jl} - \mathbf{r}_{jl} - \mathbf{v}_i^{(t+1)}\|^2}{1 + d_{nj}} \right) + \sum_{n \in \mathcal{N}_j} \frac{\beta_l |w_{jl}^{(t)} r_{jl}|^2}{1 + d_{nj}} \quad (17)$$

By zeroing the gradient of the energy function in (17) in terms of r_{jl} , the iterative solution to (17) is expressed as

$$r_{jl}^{(t+1)} = \frac{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m (\mathbf{x}_{jl} - \mathbf{v}_i^{(t+1)})}{1 + d_{nj}}}{\sum_{i=1}^c \sum_{n \in \mathcal{N}_j} \frac{(u_{in}^{(t+1)})^m}{1 + d_{nj}} + \sum_{n \in \mathcal{N}_j} \frac{\beta_l (w_{jl}^{(t)})^2}{1 + d_{nj}}} \quad (18)$$

To show the impact of weighted ℓ_2 -norm regularization on FCM, we show an example, as shown in Fig. 5. We impose a mixture of Poisson, Gaussian, and impulse noise ($\sigma = 30$, $p = 20\%$) on a noise-free image in Fig. 5(a). We set c to 4. The settings of ξ and β are discussed in the later section.

As shown in Fig. 5, the noise estimation of DSFCM_N in Fig. 5(g) is far from the true one in Fig. 5(f). However, WRFCM achieves a better noise estimation result as shown in Fig. 5(h). In addition, it has better performance for noise-suppression and feature-preserving than DSFCM_N, which can be visually observed from Fig. 5(c) and (d).

Algorithm 1 is terminated when $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$. Based on optimal U and V , a segmented image \hat{X} is obtained. WRFCM for minimizing (9) is realized in Algorithm 2.

G. Convergence and Robustness Analysis

In WRFCM, we set $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$ as the termination condition. In order to analyze the convergence and robustness of WRFCM, we take Fig. 5 as a case study. We set $\varepsilon = 1 \times 10^{-8}$. For convergence analysis, we draw the curves of

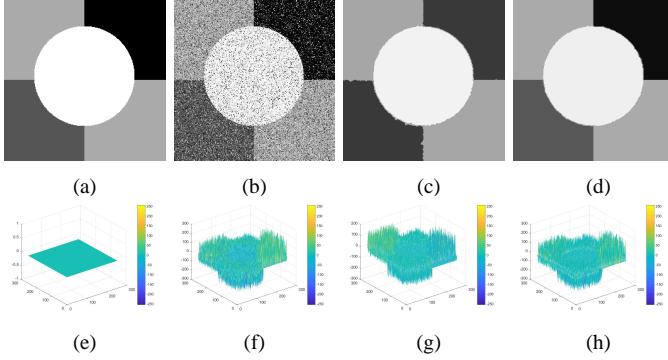


Fig. 5. Noise estimation comparison between DSFCM_N and WRFCM. (a) noise-free image; (b) observed image; (c) segmented image of DSFCM_N; (d) segmented image of WRFCM; (e) noise in the noise-free image; (f) noise in the observed image; (g) noise estimation of DSFCM_N; and (h) noise estimation of WRFCM.

Algorithm 2 Residual-driven FCM with weighted ℓ_2 -norm regularization (WRFCM)

Input: Observed image \mathbf{X} , fuzzification exponent m , number of clusters c , and threshold ε .

Output: Segmented image $\hat{\mathbf{X}}$.

- 1: Initialize $\mathbf{W}^{(0)}$ as a matrix with all elements of 1 and generate randomly prototypes $\mathbf{V}^{(0)}$
- 2: $t \leftarrow 0$
- 3: **repeat**
- 4: Calculate the partition matrix $\mathbf{U}^{(t+1)}$ via (13)
- 5: Calculate the prototypes $\mathbf{V}^{(t+1)}$ via (14)
- 6: Calculate the residual $\mathbf{R}^{(t+1)}$ via (18)
- 7: Update the weight matrix $\mathbf{W}^{(t+1)}$ via (7)
- 8: $t \leftarrow t + 1$
- 9: **until** $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$
- 10: **return** \mathbf{U} , \mathbf{V} , \mathbf{R} and \mathbf{W}
- 11: Generate the segmented image $\hat{\mathbf{X}}$ based on \mathbf{U} and \mathbf{V}

$\theta = \|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\|$ and J versus iteration step t , respectively. For robustness analysis, we draw the curve of $\tau = \|\mathbf{R}^{t+1} - \hat{\mathbf{R}}\|$ versus iteration step t , where $\hat{\mathbf{R}}$ represents the measured residual (noise) reserved in Fig. 5(b). The results are presented in Fig. 6.

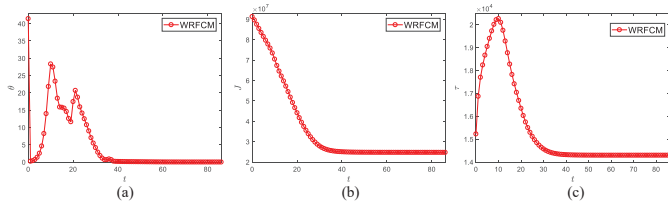


Fig. 6. Convergence and robustness of WRFCM. (a) θ ; (b) J ; and (c) τ versus t .

As Fig. 6(a) indicates, since the prototypes are randomly initialized, the convergence of WRFCM oscillates slightly at the beginning. Nevertheless, it reaches steady state after a few iterations. Even though θ exhibits an oscillating process, the objective function value J keeps decreasing until the iteration stops. Affected by the oscillating process of θ , τ takes on a pattern of increasing and then decreasing and eventually stabilizes. The finding indicates that WRFCM can obtain optimal

residual estimation as t increases. To sum up, WRFCM has outstanding convergence and robustness since the weight ℓ_2 -norm regularization makes mixed noise distribution estimated accurately so that the residual is gradually separated from observed data as iterations proceed.

III. EXPERIMENTAL STUDIES

In this section, to show the performance, efficiency and robustness of WRFCM, we provide numerical experiments on synthetic, medical, and other real-world images. To highlight the superiority and improvement of WRFCM over conventional FCM, we also compare it with seven FCM variants, i.e., FCM_S1 [10], FCM_S2 [10], FLICM [13], KWFLICM [15], FRFCM [30], WFCM [22], and DSFCM_N [34]. They are the most representative ones in the field. At the last of this section, to further verify WRFCM's strong robustness, we compare WRFCM with two competing approaches unrelated to FCM, i.e., PFE [45] and AMR_SC [46]. For a fair comparison, we note that all experiments are implemented in Matlab on a laptop with Intel(R) Core(TM) i5-8250U CPU of (1.60 GHz) and 8.0 GB RAM.

A. Evaluation Indicators

To quantitatively evaluate the performance of WRFCM, we adopt three objective evaluation indicators, i.e., segmentation accuracy (SA) [15], Matthews correlation coefficient (MCC) [47], and Sorensen-Dice similarity (SDS) [48], [49]. Note that a single one cannot fully reflect true segmentation results. SA is defined as

$$SA = \sum_{i=1}^c |S_i \cap G_i| / K$$

where S_i and G_i are the i -th cluster in a segmented image and its ground truth, respectively. $|\cdot|$ denotes the cardinality of a set. MCC is computed as

$$MCC = \frac{T_P \cdot T_N - F_P \cdot F_N}{\sqrt{(T_P + F_P) \cdot (T_P + F_N) \cdot (T_N + F_P) \cdot (T_N + F_N)}}$$

where T_P , F_P , T_N , and F_N are the numbers of true positive, false positive, true negative, and false negative, respectively. SDS is formulated as

$$SDS = \frac{2T_P}{2T_P + F_P + F_N}$$

B. Dataset Descriptions

Tested images except for synthetic ones come from four publicly available databases including a medical one and three real-world ones. The details are outlined as follows:

- 1) BrianWeb¹: 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 provides golden standard segmentation.
- 2) Berkeley Segmentation Data Set (BSDS)² [50]: This database contains 200 training, 100 validation and 200

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

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

testing images. Golden standard segmentation is annotated by different subjects for each image of size 321×481 or 481×321 .

- 3) Microsoft Research Cambridge Object Recognition Image Database (MSRC)³: This database contains 591 images and 23 object classes. Golden standard segmentation is provided.
- 4) NASA Earth Observation Database (NEO)⁴: This database continually provides information collected by NASA satellites about Earth's ocean, atmosphere, and land surfaces. Due to bit errors appearing in satellite measurements, sampled images of size 1440×720 contain unknown noise. Thus, their ground truths are unknown.
- 5) PASCAL Visual Object Classes 2012 (VOC2012)⁵: This dataset contains 20 object classes. The training/validation data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations. Each image is of size 500×375 or 375×500 .

C. Parameter Settings

Prior to numerical simulations, we report the parameter settings of WRFCM and comparative algorithms. Since AM-R_SC and PFE are not related to FCM, we follow their parameter settings introduced in their original articles [45], [46]. In the following, we focus on all FCM-related algorithms. Since spatial information is used in all algorithms, a filtering window of size $\omega = 3 \times 3$ is selected for FCM_S1, FCM_S2, KWFLICM, FRFCM, and WFCM and a local window of size $|\mathcal{N}_j| = 3 \times 3$ is chosen for FLICM, KWFLICM, DSFCM_N, and WRFCM. We set $m = 2$ and $\varepsilon = 1 \times 10^{-6}$ across all algorithms. The setting of c is presented in each experiment.

Except m , ε and c , FLICM and KWFLICM are free of all parameters. However, the remaining algorithms involve different parameters. In FCM_S1 and FCM_S2, α is set to 3.8, which controls the impact of spatial information on FCM by following [10]. In FRFCM, an observed image is taken as a mask image. A marker image is produced by a 3×3 structuring element. WFCM requires one parameter $\mu \in [0.55, 0.65]$ only, which constrains the neighbor term. For DSFCM_N, λ is set based on the standard deviation of each channel of image data.

As to WRFCM, it requires two parameters, i.e., ξ in (7) and β in (9). By analyzing mixed noise distributions, ξ is experimentally set to 0.0008. The weighting results are portrayed in Fig. 4. Since the standard deviation of image data is related to noise levels to some extent [34], we can set β in virtue of the standard deviation of each channel. Based on massive experiments, $\beta = \{\beta_l : l = 1, 2, \dots, L\}$ is recommended to be chosen as follows:

$$\beta_l = \frac{\phi \cdot \delta_l}{100} \text{ for } \phi \in [5, 10]$$

where δ_l is the standard deviation of the l -th channel of \mathbf{X} . If ϕ is set, β_l is computed. Therefore, β is equivalently replaced by ϕ . In Fig. 7, we cover an example to show the setting of ϕ associated with Fig. 8.

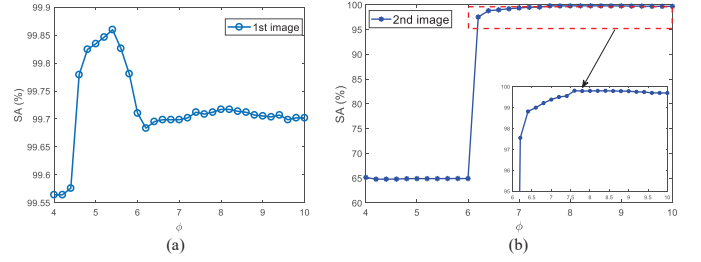


Fig. 7. SA values versus ϕ .

As Fig. 7(a) indicates, when coping with the first image, the SA value reaches its maximum gradually as the value of ϕ increases. Afterwards, it decreases rapidly and tends to be stable. As shown in Fig. 7(b), for the second image, after the SA value reaches its maximum, it has no apparent changes, implying that the segmentation performance is rather stable. In conclusion, for image segmentation, WRFCM can produce better and better performance as parameter ϕ increases from a small value.

D. Experimental Results and Analysis

In this subsection, we show the comparison results between WRFCM and seven FCM-related algorithms. We test a set of synthetic, medical, and real-world images. They are borrowed from four public datasets, i.e., BrainWeb, BSDS, MSRC, and NEO.

1) Results on Synthetic Images: In the first experiment, we representatively choose two synthetic images of size 256×256 . A mixture of Poisson, Gaussian, and impulse noises is considered for all cases. To be specific, Poisson noise is first added. Then we add Gaussian noise with $\sigma = 30$. Finally, the random-valued impulse noise with $p = 20\%$ is added since it is more difficult to detect than salt and pepper impulse noise. For two images, we set $c = 4$. The segmentation results are given in Fig. 8 and Table I. The best values are in bold.

As Fig. 8 indicates, FCM_S1, FCM_S2 and FLICM achieve poor results in presence of such a high level of mixed noise. Compared with them, KWFLICM, FRFCM and WFCM suppress the vast majority of mixed noise. Yet they cannot completely remove it. DSFCM_N visually outperforms other peers mentioned above. However, it generates several topology changes such as merging and splitting. By taking the second synthetic image as a case, we find that DSFCM_N produces some unclear contours and shadows. Superiorly to seven peers, WRFCM not only removes all the noise but also preserves more image features.

Table I shows the segmentation results of all algorithms quantitatively. It assembles the values of all three indicators. Clearly, WRFCM achieves better SA, SDS, and MCC results for all images than other peers. In particular, its SA value comes up to 99.859% for the first synthetic image. Among its seven peers, KWFLICM obtains generally better results. In the light of Fig. 8 and Table I, we conclude that WRFCM performs better than its peers.

2) Results on Medical Images: Next, we representatively segment two medical images from BrianWeb. They are represented as two slices in the axial plane with 70 and 80, which

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

⁴<http://neo.sci.gsfc.nasa.gov/>

⁵<http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html>

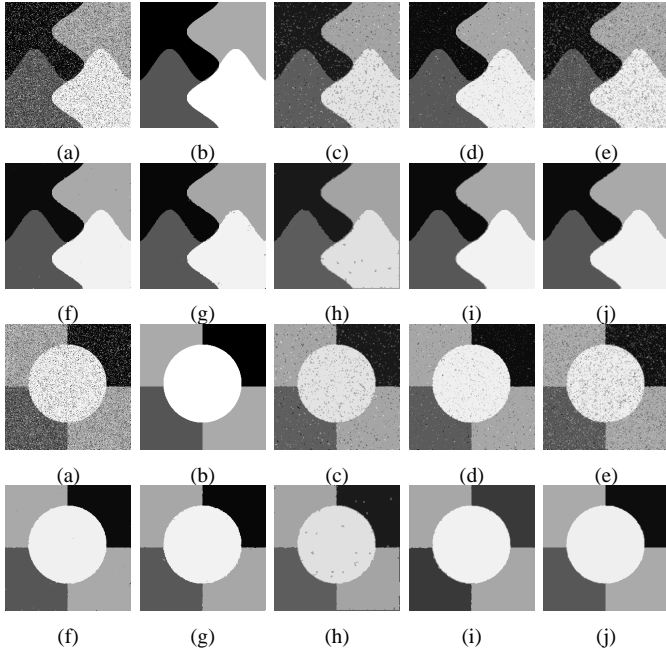


Fig. 8. Visual results for segmenting synthetic images ($\phi_1 = 5.58$ and $\phi_2 = 7.45$). From (a) to (j): noisy images, ground truth, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

TABLE I
SEGMENTATION PERFORMANCE (%) ON SYNTHETIC IMAGES

Algorithm	First synthetic image			Second synthetic image		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	92.902	98.187	96.362	92.625	98.414	95.528
FCM_S2	96.157	98.999	97.991	96.292	99.127	97.520
FLICM	85.081	90.145	95.082	85.667	95.894	88.576
KWFLICM	99.706	99.858	99.715	99.730	99.904	99.725
FRFCM	99.652	99.920	99.839	99.675	99.895	99.698
WFCM	97.827	99.325	98.652	98.079	99.363	98.197
DSFCM_N	98.954	99.545	99.086	99.226	99.757	99.303
WRFCM	99.859	99.937	99.843	99.802	99.958	99.792

are generated by T1 modality with slice thickness of 1mm resolution, 9% noise and 20% intensity non-uniformity. Here, we set $c = 4$ for all cases. The comparisons between WRFCM and its peers are shown in Fig. 9 and Table II. The best values are in bold.

TABLE II
SEGMENTATION PERFORMANCE (%) ON MEDICAL IMAGES IN BRIANWEB

Algorithm	First medical image			Second medical image		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	75.756	97.852	96.225	75.026	98.109	96.656
FCM_S2	75.769	98.119	96.664	74.970	98.176	96.765
FLICM	74.998	98.070	96.568	74.185	98.122	96.660
KWFLICM	74.840	98.259	96.878	73.839	97.860	96.190
FRFCM	75.853	97.620	95.775	75.514	97.660	95.830
WFCM	75.507	97.124	94.957	74.471	97.213	95.045
DSFCM_N	76.400	92.325	86.262	75.288	91.574	85.095
WRFCM	82.317	98.966	98.147	82.141	98.298	96.970

By viewing the marked red squares in Fig. 9, we find that FCM_S1, FCM_S2, FLICM, KWFLICM and DSFCM_N are vulnerable to noise and intensity non-uniformity. They give rise to the change of topological shapes to some extent. Unlike them, FRFCM and WFCM achieve sufficient noise removal.

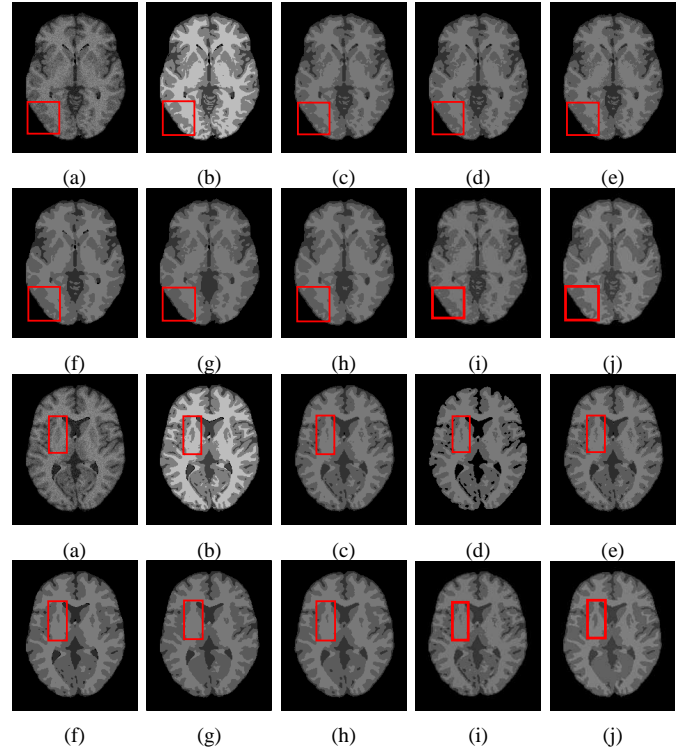


Fig. 9. Visual results for segmenting medical images ($\phi = 5.35$). From (a) to (j): noisy images, ground truth, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

However, they produce overly smooth contours. Compared with its seven peers, WRFCM can not only suppress noise adequately but also acquire accurate contours. Moreover, it yields the visual result closer to ground truth than its peers. As Table II shows, WRFCM obtains optimal SA, SDS and MCC results for all medical images. As a conclusion, it outperforms its peers visually and quantitatively.

3) **Results on Real-world Images:** In order to demonstrate the practicality of WRFCM for other image segmentation, we typically choose two sets of real-world images in the last experiment. The first set contains five representative images from BSDS and MSRC. There usually exist some outliers, noise or intensity inhomogeneity in each image. For all tested images, we set $c = 2$. The results of all algorithms are shown in Fig. 10 and Table III.

Fig. 10 visually shows the comparison between WRFCM and seven peers while Table III gives the quantitative comparison. Apparently, WRFCM achieves better segmentation results than its peers. FCM_S1, FCM_S2, FLICM, KWFLICM and DSFCM_N obtain unsatisfactory results on all tested images. Superiorly to them, FRFCM and WFCM preserve more contours and feature details. From a quantitative point of view, WRFCM acquires optimal SA, SDS, and MCC values much more than its peers. Note that it merely gets a slightly smaller SDS value than FRFCM and WFCM for the first and second images, respectively.

The second set contains images from NEO. Here, we select two typical images. Each of them represents an example for a specific scene. We produce the ground truth of each scene by randomly shooting it for 50 times within the time span 2000–

TABLE III
SEGMENTATION PERFORMANCE (%) ON REAL-WORLD IMAGES IN BSDS AND MSRC

Algorithm	Fig. 10 column 1			Fig. 10 column 2			Fig. 10 column 3			Fig. 10 column 4			Fig. 10 column 5		
	SA	SDS	MCC	SA	SDS	MCC	SA	SDS	MCC	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	86.384	89.687	69.705	50.997	66.045	2.724	67.289	72.570	32.232	80.688	88.159	49.369	78.717	47.696	48.874
FCM_S2	86.138	79.701	69.208	51.433	12.089	2.951	67.105	59.523	31.941	80.657	47.557	49.256	78.365	86.449	47.881
FLICM	86.476	89.771	69.882	55.292	70.055	2.403	89.233	91.167	78.117	80.771	47.826	49.729	80.617	54.490	54.029
KWFLICM	87.119	90.278	71.283	48.252	63.432	1.554	64.617	66.081	30.820	80.484	46.723	48.777	77.963	44.791	46.755
FRFCM	97.701	98.235	94.941	99.690	97.436	97.273	99.380	99.467	98.732	83.974	89.927	58.683	96.985	97.861	92.987
WFCM	98.442	97.755	96.563	99.688	99.834	97.268	99.295	99.160	98.555	84.480	62.664	60.043	96.445	93.943	91.719
DSFCM_N	93.116	90.279	84.987	50.688	11.093	0.638	92.101	90.791	83.922	50.858	60.181	0.506	95.412	92.319	89.179
WRFCM	98.732	98.162	97.201	99.746	97.906	97.771	99.442	99.520	98.857	99.826	99.888	99.074	99.869	99.789	99.694

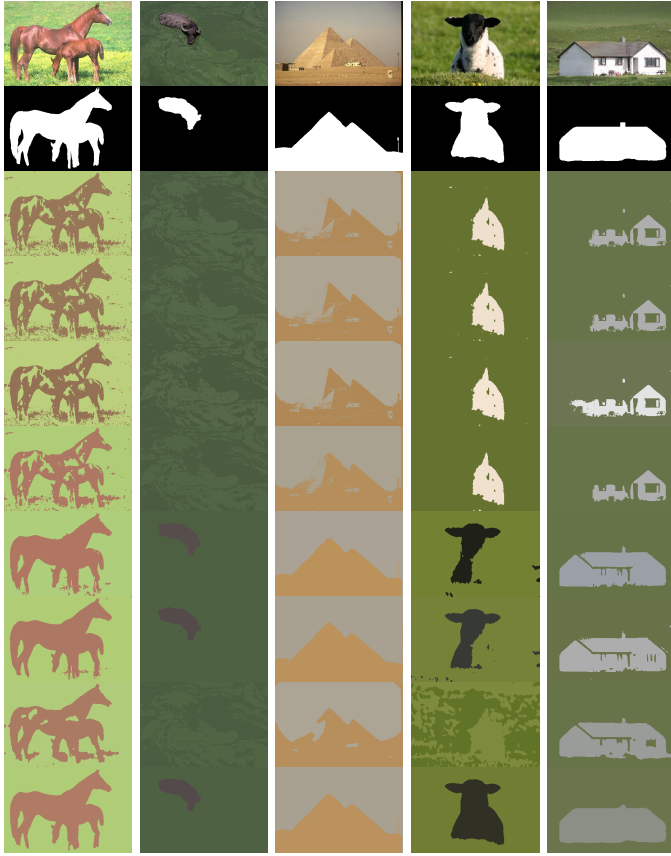


Fig. 10. Segmentation results on five real-world images in BSDS and MSRC ($\phi_1 = 6.05$, $\phi_2 = 10.00$, $\phi_3 = 9.89$, $\phi_4 = 9.98$, and $\phi_5 = 9.50$). From top to bottom: observed images, ground truth, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

2019. The visual results of all algorithms are shown in Figs. 11 and 12. The corresponding SA, SDS, and MCC values are given in Table IV.

Fig. 11 shows the segmentation results on sea ice and snow extent. The colors represent the land and ocean covered by snow and ice per week (here is February 7–14, 2015). We set $c = 4$. Fig. 12 gives the segmentation results on chlorophyll concentration. The colors represent where and how much phytoplankton is growing over a span of days. We choose $c = 2$. As a whole, by seeing Figs. 11 and 12, as well as Table IV, FCM_S1, FCM_S2, FLICM, KWFLICM, and WFCM are sensitive to unknown noise. FRFCM and DSFCM_N produce overly smooth results. Especially, they generate incorrect clusters when segmenting the first image in

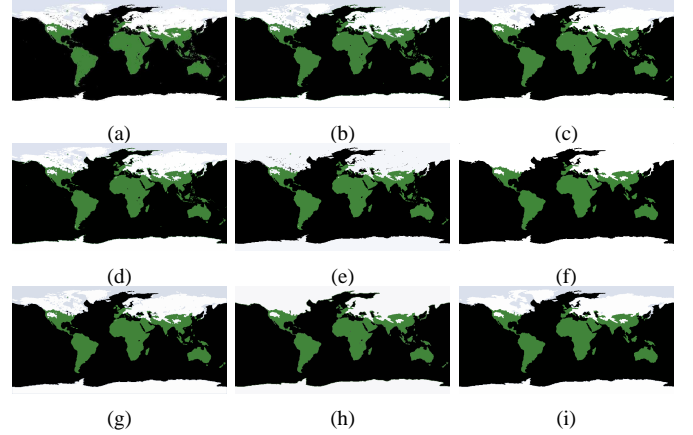


Fig. 11. Segmentation results on the first real-world image in NEO ($\phi = 6.10$). From (a) to (i): observed image and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

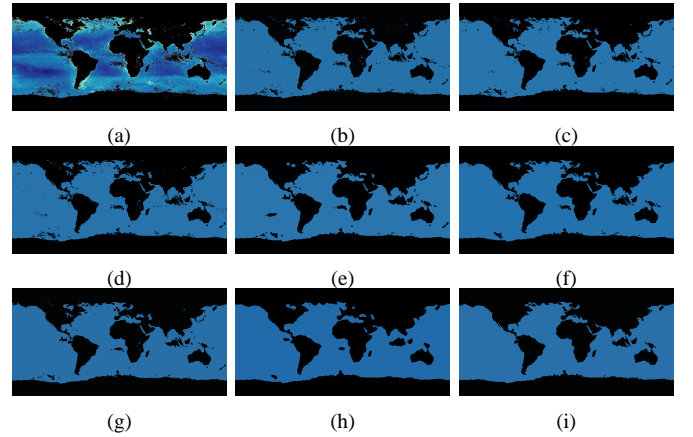


Fig. 12. Segmentation results on the second real-world image in NEO ($\phi = 9.98$). From (a) to (i): observed image and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

in NEO. Superiorly to its seven peers, WRFCM cannot only suppress unknown noise well but also retain image contours well. In particular, it makes up the shortcoming that other peers forge several topology changes in the form of black patches when coping with the second image in NEO.

4) **Performance Improvement:** Besides segmentation results reported for all algorithms, we also present the performance improvement of WRFCM over seven comparative algorithms in Table V. Clearly, for all types of images, the average SA, SDS and MCC improvements of WRFCM over other peers are within the value span 0.113%–27.836%,

TABLE IV
SEGMENTATION PERFORMANCE (%) ON REAL-WORLD IMAGES IN NEO

Algorithm	Fig. 11			Fig. 12		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	90.065	97.060	95.106	80.214	92.590	90.329
FCM_S2	93.801	97.723	95.563	81.054	92.066	90.023
FLICM	90.234	97.056	95.781	81.582	92.352	90.236
KWFLICM	85.902	80.109	76.329	95.001	96.364	95.633
FRFCM	81.319	80.616	78.220	96.369	97.309	96.215
WFCM	95.882	98.854	97.293	97.342	97.430	97.178
DSFCM_N	80.131	81.618	79.597	96.639	97.936	96.436
WRFCM	99.080	99.149	98.512	98.881	98.797	97.582

0.040%–41.989%, and 0.049%–58.681%, respectively.

TABLE V
AVERAGE PERFORMANCE IMPROVEMENTS (%) OF WRFCM OVER COMPARATIVE ALGORITHMS

Algorithm	Synthetic images			Medical images		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	7.067	1.647	3.873	6.838	0.651	1.118
FCM_S2	3.606	0.884	2.062	6.859	0.484	0.844
FLICM	14.456	6.928	7.989	7.637	0.536	0.945
KWFLICM	0.113	0.067	0.098	7.889	0.572	1.024
FRFCM	0.167	0.040	0.049	6.545	0.992	1.756
WFCM	1.878	0.603	1.393	7.240	1.463	2.558
DSFCM_N	0.741	0.297	0.623	6.385	6.682	11.880

Algorithm	Real-world images in BSDS and MSRC			Real-world images in NEO		
	SA	SDS	MCC	SA	SDS	MCC
FCM_S1	26.708	26.221	57.938	13.841	4.148	5.329
FCM_S2	26.783	41.989	58.271	11.553	4.078	5.254
FLICM	21.045	28.391	47.687	13.073	4.269	5.038
KWFLICM	27.836	36.791	58.681	8.529	10.736	12.066
FRFCM	3.977	2.468	9.996	10.137	10.010	10.829
WFCM	3.853	8.381	9.690	2.369	0.831	0.811
DSFCM_N	23.088	30.120	46.673	10.596	9.196	10.030

5) *Overhead Analysis*: In the previous subsections, the segmentation performance of WRFCM is presented. Next, we provide the comparison of computing overheads between WRFCM and seven comparative algorithms in order to show its practicality. We note that K is the number of image pixels, c is the number of prototypes, t is the iteration count, ω represents the size of a filtering window, $|\mathcal{N}_j|$ denotes the size of a local window centralized at pixel j , and μ is the number of pixel levels in an image. Generally, $\mu \ll K$. Generally speaking, since FCM-related algorithms are unsupervised and their solutions are easily updated iteratively, they have low storage complexities of $O(K)$. In addition, we summarize the computational complexity of all algorithms in Table VI.

TABLE VI
COMPUTATIONAL COMPLEXITY OF ALL ALGORITHMS

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

As Table VI shows, FRFCM has lower computational complexity than its peers due to $\mu \ll K$. Except WFCM,

the computational complexity of other algorithms is basically $O(K)$, the computational complexity of WRFCM is not high. To compare the practicability between WRFCM and its peers, we present the execution times of all algorithms for segmenting synthetic, medical, and real-world images in Table VII.

TABLE VII
COMPARISON OF EXECUTION TIME (IN SECONDS) OF ALL ALGORITHMS

Algorithm	Fig. 8 (Average)	Fig. 9 (Average)	Fig. 10 (Average)	Fig. 11	Fig. 12
FCM_S1	4.284	3.971	5.075	5.535	4.644
FCM_S2	3.674	3.362	5.346	5.288	4.182
FLICM	50.879	96.709	278.694	224.195	40.727
KWFLICM	67.838	128.946	371.592	298.926	54.303
FRFCM	0.263	0.247	1.162	5.815	1.786
WFCM	3.772	2.685	5.419	6.644	7.104
DSFCM_N	7.758	5.786	8.987	36.648	18.897
WRFCM	6.455	2.819	4.893	4.977	2.761

As Table VII shows, for gray and color image segmentation, the computational efficiencies of FLICM and KWFLICM are far lower than those of others. In contrast, since gray level histograms are considered, FRFCM takes the least execution time among all algorithms. Due to the computation of a neighbor term in advance, FCM_S1 and FCM_S2 are more time-saving than most of other algorithms. Even though WFCM and DSFCM_N need more computing overheads than FRFCM, they are still very efficient. For color image segmentation, the execution time of DSFCM_N increases dramatically. Compared with most of seven comparative algorithms, WRFCM shows higher computational efficiency. In most cases, it only runs slower than FRFCM. However, the shortcoming can be offset by its better segmentation performance. In a quantitative study, for each image, WRFCM takes 0.011 and 2.642 seconds longer than FCM_S2 and FRFCM, respectively. However, it saves 0.321, 133.860, 179.940, 0.744, and 11.234 seconds over FCM_S1, FLICM, KWFLICM, FRFCM, WFCM, and DSFCM_N, respectively.

E. Comparison with Non-FCM Methods

In this subsection, we compare WRFCM with two non-FCM methods, i.e., PFE [45] and AMR_SC [46]. We list the comparison results on VOC2012 in Fig. 13 and Table VIII. The computing times of the WRFCM, AMR_SC, and PFE are summarized in Table IX. The results indicate that WRFCM achieves better effectiveness and efficiency than AMR_SC and PFE.

IV. CONCLUSIONS AND FUTURE WORK

For the first time, a residual-driven FCM (RFCM) framework is proposed for image segmentation, which advances FCM research. It realizes favorable noise estimation in virtue of a residual-related regularization term coming with an analysis of noise distribution. On the basis of the framework, RFCM with weighted ℓ_2 -norm regularization (WRFCM) is presented for coping with image segmentation with mixed or unknown noises. Spatial information is also considered in WRFCM for making residual estimation more reliable. A two-step iterative algorithm is presented to implement WRFCM. Experiments

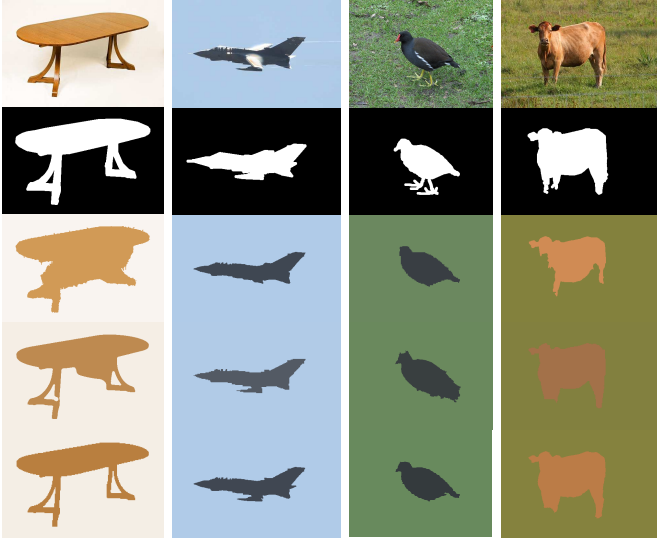


Fig. 13. Segmentation results on four real-world images in VOC2012 ($\phi_1 = 5.65$, $\phi_2 = 6.75$, $\phi_3 = 5.50$, $\phi_4 = 8.35$). From top to bottom: observed images, ground truth, and results of AMR_SC, PFE, and WRFCM.

TABLE VIII
SEGMENTATION PERFORMANCE (%) ON REAL-WORLD IMAGES IN VOC2012

Algorithm	Fig. 13 column 1			Fig. 13 column 2		
	SA	SDS	MCC	SA	SDS	MCC
AMR_SC	90.561	90.259	84.867	89.216	90.060	92.849
PFE	95.992	95.211	94.826	96.077	97.748	95.513
WRFCM	98.946	98.651	97.353	97.167	98.848	96.072

Algorithm	Fig. 13 column 3			Fig. 13 column 4		
	SA	SDS	MCC	SA	SDS	MCC
AMR_SC	85.095	88.615	80.878	91.148	92.414	90.156
PFE	90.358	94.516	90.923	93.273	94.950	93.743
WRFCM	92.756	96.132	93.330	95.436	96.869	97.012

reported for four benchmark databases demonstrate that it outperforms existing FCM variants and non-FCM methods. Moreover, differing from popular residual-learning methods, it is unsupervised and exhibits a high speed of clustering.

There are some open issues worth pursuing. First, since a tight wavelet frame transform [51]–[53] provides redundant representations of images, it can be used to manipulate and analyze image features and noise well. Therefore, it can be taken as a kernel function so as to produce an improved FCM algorithm, i.e., wavelet kernel-based FCM. Second, can the proposed algorithm be applied to a wide range of non-flat domains such as remote sensing [54], ecological systems [55], and transportation networks [56]? How can the number of clusters be selected automatically? Can more recently proposed minimization algorithms [57]–[60] be applied to the

TABLE IX
COMPARISON OF EXECUTION TIME (IN SECONDS) BETWEEN WRFCM AND TWO NON-FCM METHODS

Algorithm	Fig. 13 column 1	Fig. 13 column 2	Fig. 13 column 3	Fig. 13 column 4
AMR_SC	6.639	6.819	6.901	7.223
PFE	206.025	234.796	217.801	238.473
WRFCM	5.509	6.097	5.425	5.828

optimization problem involved in this work? Answering them needs more research efforts.

APPENDIX PROOF OF THEOREM II.1

Consider the first two subproblems of (12). The Lagrangian function (10) is reformulated as

$$\mathcal{L}_\Lambda(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{j=1}^K u_{ij}^m D_{ij} + \sum_{j=1}^K \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right), \quad (19)$$

$$\text{where } D_{ij} = \sum_{n \in \mathcal{N}_j} \frac{\|\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i\|^2}{1 + d_{nj}}.$$

By fixing \mathbf{V} , we minimize (19) in terms of \mathbf{U} . By zeroing the gradient of (19) in terms of \mathbf{U} , one has

$$\frac{\partial \mathcal{L}_\Lambda}{\partial u_{ij}} = m D_{ij} u_{ij}^{m-1} + \lambda_j = 0.$$

Thus, u_{ij} is expressed as

$$u_{ij} = \left(\frac{-\lambda_j}{m} \right)^{1/(m-1)} D_{ij}^{-1/(m-1)}. \quad (20)$$

Due to the constraint $\sum_{i=1}^c u_{ij} = 1$, one has

$$1 = \sum_{q=1}^c u_{qj} = \sum_{q=1}^c \left(\left(\frac{-\lambda_j}{m} \right)^{1/(m-1)} D_{qj}^{-1/(m-1)} \right) = \left(\frac{-\lambda_j}{m} \right)^{1/(m-1)} \sum_{q=1}^c D_{qj}^{-1/(m-1)}.$$

In the sequel, one can get

$$\left(\frac{-\lambda_j}{m} \right)^{1/(m-1)} = 1 / \sum_{q=1}^c D_{qj}^{-1/(m-1)}. \quad (21)$$

Substituting (21) into (20), the optimal u_{ij} is acquired:

$$u_{ij} = \frac{D_{ij}^{-1/(m-1)}}{\sum_{q=1}^c D_{qj}^{-1/(m-1)}}.$$

By fixing \mathbf{U} , we minimize (19) in terms of \mathbf{V} . By zeroing the gradient of (19) in terms of \mathbf{V} , one has

$$\frac{\partial \mathcal{L}_\Lambda}{\partial \mathbf{v}_i} = -2 \cdot \sum_{j=1}^K \left(u_{ij}^m \sum_{n \in \mathcal{N}_j} \frac{(\mathbf{x}_n - \mathbf{r}_n - \mathbf{v}_i)}{1 + d_{nj}} \right) = 0.$$

The intermediate process is presented as

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

The optimal \mathbf{v}_i is computed:

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

REFERENCES

- [1] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition. I," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 29, no. 6, pp. 778–785, Dec. 1999.
- [2] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition. II," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 29, no. 6, pp. 786–801, Dec. 1999.
- [3] C. Subbalakshmi, G. Ramakrishna, and S. K. M. Rao, "Evaluation of data mining strategies using fuzzy clustering in dynamic environment," in *Proc. Int. Conf. Adv. Comput., Netw., Informat.*, 2016, pp. 529–536.
- [4] 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.
- [5] M. Yambal and H. Gupta, "Image segmentation using fuzzy c means clustering: A survey," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 7, pp. 2927–2929, Jul. 2013.
- [6] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybernet.*, vol. 3, no. 3, pp. 32–57, 1973.
- [7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [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] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] X. Zhu, W. Pedrycz, and Z. W. Li, "Fuzzy clustering with nonlinearly transformed data," *Appl. Soft Comput.*, vol. 61, pp. 364–376, Dec. 2017.
- [22] 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.
- [23] R. R. Ghareeb, 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.
- [24] C. Wang, W. Pedrycz, Z. Li, and M. Zhou, "Kullback-Leibler divergence-based Fuzzy C-Means clustering incorporating morphological reconstruction and wavelet frames for image segmentation," arXiv preprint arXiv:2002.09479, 2020.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.*, to be published, doi: 10.1109/TFUZZ.2019.2930030.
- [32] C. Wang, W. Pedrycz, Z. Li, M. Zhou, and J. Zhao, "Residual-sparse Fuzzy C-Means clustering incorporating morphological reconstruction and wavelet frames," arXiv preprint arXiv:2002.08418, 2020.
- [33] 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.
- [34] 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.
- [35] A. Fakhry, T. Zeng, and S. Ji, "Residual deconvolutional networks for brain electron microscopy image segmentation," *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 447–456, Feb. 2017.
- [36] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
- [37] F. Kokkinos and S. Lefkimmiatis, "Iterative residual CNNs for burst photography applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5929–5938.
- [38] D. Ren, W. Zuo, D. Zhang, L. Zhang, and M. H. Yang, "Simultaneous fidelity and regularization learning for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2926357.
- [39] Y. Zhang, X. Li, M. Lin, B. Chiu, and M. Zhao, "Deep-recursive residual network for image semantic segmentation," *Neural Comput. Applic.*, to be published, doi: 10.1007/s00521-020-04738-5.
- [40] J. Jiang, L. Zhang, and J. Yang, "Mixed noise removal by weighted encoding with sparse nonlocal regularization," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2651–2662, Jun. 2014.
- [41] P. Zhou, C. Lu, J. Feng, Z. Lin and S. Yan, "Tensor low-rank representation for data recovery and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2954874.
- [42] T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by Poisson noise," *J. Math. Imag. Vis.*, vol. 27, no. 3, pp. 257–263, Apr. 2007.
- [43] P. J. Huber, "Robust regression: Asymptotics, conjectures and Monte Carlo," *Ann. Stat.*, vol. 1, no. 5, pp. 799–821, 1973.
- [44] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [45] C. Fang, Z. Liao, and Y. Yu, "Piecewise flat embedding for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1470–1485, Jun. 2019.
- [46] 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.
- [47] D. N. H. Thanh, D. Sergey, V. B. S. Prasath, and N. H. Hai, "Blood vessels segmentation method for retinal fundus images based on adaptive principal curvature and image derivative operators," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W12, pp. 211–218, May 2019. doi: 10.5194/isprs-archives-XLII-2-W12-211-2019
- [48] D. N. H. Thanh, U. Erkan, V. B. S. Prasath, V. Kumar, and N. N. Hien, "A skin lesion segmentation method for dermoscopic images based on adaptive thresholding with normalization of color models," in *Proc. IEEE 6th Int. Conf. Electr. Electron. Eng.*, Apr. 2019, pp. 116–120.

- [49] 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.
- [50] P. Arbelaez, 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] 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.
- [52] 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.
- [53] 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.
- [54] 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.
- [55] 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.
- [56] 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.
- [57] X. Luo, M. Zhou, Z. Wang, Y. Xia, and Q. Zhu, "An effective scheme for QoS estimation via alternating direction method-based matrix factorization," *IEEE Trans. Serv. Comput.*, vol. 12, no. 4, pp. 503–518, July–Aug. 2019.
- [58] X. Luo, M. Zhou, Y. Xia, Q. Zhu, A. C. Ammari, and A. Alabdulwahab, "Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 524–537, Mar. 2016.
- [59] X. Luo, H. Wu, H. Yuan, and M. Zhou, "Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1798–1809, May 2020.
- [60] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 601–614, Feb. 2019.



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.S.c. 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.

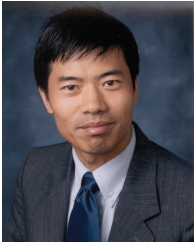
Dr. Pedrycz was a recipient of the IEEE Canada Computer Engineering Medal, the Cajastur Prize for Soft Computing from the European Centre for Soft Computing, the Killam Prize, and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is intensively involved in editorial activities. 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, workflow modeling and analysis, 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).