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—Due to its inferior characteristics, an observed (noisy) image’s direct use gives rise to poor segmentation results. Intuitively, using its noise-free image can favorably impact image segmentation. Hence, the accurate estimation of the residual between observed and noise-free images is an important task. To do so, we elaborate on residual-driven Fuzzy C-Means (FCM) for image segmentation, which is the first approach that realizes accurate residual estimation and leads noise-free image to participate in clustering. We propose a residual-driven FCM framework by integrating into FCM a residual-related fidelity term derived from the distribution of different types of noise. Built on this framework, we present a weighted $\ell_2$-norm fidelity term by weighting mixed noise distribution, thus resulting in a universal residual-driven FCM algorithm in presence of mixed or unknown noise. 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 existing FCM-related algorithms.

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

1 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], [7], [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 noise, especially mixed or unknown noise, 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], [10], [11], [12], [13], [14] and substituting its Euclidean distance with a kernel distance (function) [15], [16], [17], [18], [19], [20], [21], [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], [28], [29] and gray level histograms [30], [31], as well as pre-processing and post-processing steps like image pixel filtering [32], membership filtering [33] and label filtering [26], [27], [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, $\ell_1$-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 equivalent to noise removal. Therefore, neither of them can achieve highly accurate residual estimation.

To address this issue, we elaborate on 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 fidelity 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 $\ell_2$-norm fidelity term corresponds...
to Gaussian noise and an \(\ell_1\)-norm one suits impulse noise. In real-world applications, since images are often corrupted by mixed or unknown noise, a specific noise distribution is difficult to be obtained. To deal with this issue, by analyzing the distribution of a wide range of mixed noise, especially a mixture of Poisson, Gaussian and impulse noise, we present a weighted \(\ell_2\)-norm fidelity term in which each residual is assigned a weight, thus resulting in an augmented version namely WRFCM for image segmentation with mixed or unknown noise. 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.

The originality of this work comes with a realization of accurate residual estimation from observed images, which benefits FCM’s performance enhancement. In essence, the proposed algorithm is an unsupervised method. Compared with commonly used supervised methods such as convolutional neural networks (CNNs) \([33], [36], [37], [38], [39], [40]\) and dictionary learning \([41], [42]\), it realizes the residual estimation precisely by virtue of a fidelity 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, being free of the aid of mathematical techniques, it achieves the superior performance over some recently proposed comprehensive FCMs. Therefore, we conclude that WRFCM is a fast and robust FCM algorithm. Finally, in a mathematical sense, its minimization problem involves an \(\ell_2\) vector norm only. Thus it can be easily solved by using a well-known Lagrangian multiplier method.

Section 2 reviews the state of the art relevant to this work. Section 3 details conventional FCM and the proposed methodology. Section 4 reports experimental results. Conclusions and some open issues are given in Section 5.

2 Related Work

In 1984, Bezdek et al. \([8]\) first proposed FCM. So far, it has evolved into the most popular fuzzy clustering algorithm. However, it cannot work well for segmenting observed (noisy) images. It has been improved by mostly considering spatial information \([9], [10], [11], [12], [13], [14]\), kernel distances (functions) \([15], [16], [17], [18], [19], [20], [21]\), \([22]\), and various mathematical techniques \([23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]\). In this paper, we mainly focus on the improvement of FCM with regard to its robustness to noise for image segmentation. Therefore, we introduce related work about it in this section.

2.1 FCM with Spatial Information

Over the past two decades, using spatial information to improve FCM’s robustness achieved remarkable successes, thus resulting in many improved versions \([9], [10], [11], [12], [13], [14]\). For instance, Ahmed et al. \([9]\) introduce a neighbor term into the objective function of FCM so as to improve its robustness by leaps and bounds, thus yielding FCM-S where S refers to “spatial information”. To further improve it, Chen and Zhang \([10]\) integrate mean and median filters into a neighbor term, thus resulting in two FCM-S variants labeled as FCM-S1 and FCM-S2. However, their computing overhead is very high. To lower it, Szilagyi et al. \([11]\) propose an enhanced FCM (EnFCM) where a weighted sum image is generated by the observed pixels and their neighborhoods. Based on it, Cai et al. \([12]\) substitute image pixels by gray level histograms, which gives rise to fast generalized FCM (FGFCM). Although it has a high computational efficiency, more parameters are required and tuned. Krinidis et al. \([13]\) come up with a fuzzy local information C-means algorithm (FLICM) for simplifying the parameter setting in FGFCM. Nevertheless, FLICM considers only non-robust Euclidean distance that is not applicable to arbitrary spatial information.
2.2 FCM with Kernel Distance

To address the serious shortcoming of FLICM [13], kernel distances (functions) are used to replace Euclidean distance in FCM. They realize the transformation from an original data space to a new one. As a result, a collection of kernel-based FCMs have been put forward [15], [16], [17], [18], [19], [20], [21], [22]. For example, Gong et al. [15] propose an improved version of FLICM, namely KWFLICM, which augments a tradeoff weighted fuzzy factor and a kernel metric into FCM. Even though it is generally robust to extensive noise, it is more time-consuming than most of existing FCMs. Zhao et al. [20] take a neighborhood weighted distance into account, thus presenting a novel algorithm called NWFCM. Although it runs faster than KWFLICM, its segmentation performance is worse. Moreover, it exhibits lower computational efficiency than other FCMs. More recently, Wang et al. [22] consider tight wavelet frames as a kernel function so as to present wavelet frame-based FCM (WFCM), which takes full advantage of the feature extraction capacity of tight wavelet frames. In spite of its rarely low computational cost, its segmentation effects can be further improved by using various mathematical techniques.

2.3 Comprehensive FCM

To keep a sound trade-off between performance and speed of clustering, comprehensive FCMs involving various mathematical techniques has been put forward [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. For instance, Gharieb et al. [23] present an FCM framework based on Kullback-Leibler (KL) divergence. It uses KL divergence to optimize the membership similarity between a pixel and its neighbors. Yet it has slow clustering speed. Gu et al. [25] report a fuzzy double C-Means algorithm (FDCM) through the utility of sparse representation, which addresses two datasets simultaneously, i.e., a basic feature set associated with an observed image and a feature set learned from a spare self-representation model. Overall, FDCM is robust and applicable to a wide range of image segmentation problems. However, its computational efficiency is not satisfactory. Lei et al. [30] present a fast and robust FCM algorithm (FRFCM) by using gray level histograms and morphological gray reconstruction. In spite of its fast clustering, its performance is sometimes unstable since morphological gray reconstruction may cause the loss of useful image features. More recently, Lei et al. [31] propose an automatic fuzzy clustering framework (AFCF) by incorporating threefold techniques, i.e., superpixel algorithms, density peak clustering and prior entropy. It overcomes two difficulties in existing algorithms [22], [25], [30]. One is to select the number of clusters automatically. The other one is to employ superpixel algorithms and the prior entropy to improve image segmentation performance. However, AFCF’s results are unstable.

In this work, the proposed algorithm differs from all algorithms mentioned above in the sense that we take a wide range of mixed noise estimations as the starting point and directly minimize the objective function of WRFCM formulated by using fidelity without dictionary learning and CNNs and archives outstanding performance in image segmentation tasks.

3 FCM AND PROPOSED METHODOLOGY

3.1 Fuzzy C-Means (FCM)

Given a set \( X = \{ x_j \in \mathbb{R}^L : j = 1, 2, \ldots, K \} \), where \( x_j \) contains \( L \) channels, i.e., \( x_j = (x_{j1}, x_{j2}, \ldots, x_{jL})^T \). FCM is applied to cluster \( X \) by minimizing:

\[
J(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{K} u_{ij}^m \| x_j - v_i \|^2
\]

where \( U = [u_{ij}]_{c \times K} \) is a partition matrix under a constraint \( \sum_{i=1}^{c} u_{ij} = 1 \) for \( j = 1, 2, \ldots, K \), \( V = \{ v_i : i = 1, 2, \ldots, c \} \) is a prototype set, \( \| \cdot \| \) stands for 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:

\[
u_{ij}^{(t+1)} = \left( \frac{\left( \| x_j - v_i^{(t)} \|^{2} \right)^{1/(m-1)}}{\sum_{q=1}^{c} \left( \| x_j - v_q^{(t)} \|^{2} \right)^{1/(m-1)}} \right)\]

\[
v_i^{(t+1)} = \left( \frac{1}{\sum_{j=1}^{K} (u_{ij}^{(t+1)})^m} x_j \right)
\]

Here, \( t = 0, 1, 2, \ldots \) is an iterative step and \( l = 1, 2, \ldots, L \).

By presetting a threshold \( \varepsilon \), the procedure stops when \( \| U^{(t+1)} - U^{(t)} \| < \varepsilon \).

3.2 Noise Model

Consider an observed image \( X \) with \( K \) pixels. It is denoted as \( X = \{ x_j : j = 1, 2, \ldots, K \} \), where \( x_j = \{ x_{jl} : l = 1, 2, \ldots, L \} \). When \( L = 1 \), \( X \) represents a gray image. For \( L = 3 \), \( X \) is a Red-Green-Blue color image. Since there is noise in an observed image, \( X \) can be modeled as a sum of a noise-free image \( \bar{X} \) and noise \( R \):

\[
X = \bar{X} + R
\]

Mathematically speaking, \( \bar{X} = \{ \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_K \} \) is an ideal value of \( X \) and thus is unknown. \( R = \{ r_1, r_2, \ldots, r_K \} \) is viewed as the residual between \( X \) and \( \bar{X} \). Its accurate estimation can make \( \bar{X} \) instead of \( X \) 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 noise are the most common, i.e., mixed Poisson-Gaussian noise and mixed Gaussian and impulse noise. Beyond them, we focus on a mixture of a wide range of noise, 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 noise.

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

\[
\vec{x}_j = \begin{cases} 
\vec{x}_j + r'_j & j \in \Omega_1 \\
r''_j & j \in \Omega_2 := \Omega \setminus \Omega_1
\end{cases}
\]

where the subset \( \Omega_2 \) of \( \Omega \) denotes the region including the missing information of \( \vec{X} \) and is assumed to be unknown with each element being drawn from the whole region \( \Omega \) by Bernoulli trial with \( p \). In image segmentation, mixed noise model \( (3) \) is for the first time presented.

### 3.3 Residual-driven FCM

Since there exists an unknown amount of noise 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 fidelity term on residual into the objective function of FCM. Consequently, an RFCM framework is first presented:

\[
J(\vec{U}, \vec{V}, \vec{R}) = \sum_{i=1}^{c} \sum_{j=1}^{K} u''_{ij} \|\vec{x}_j - \vec{r}_j - \vec{v}_i\|^2 + \beta \cdot \Gamma(\vec{R})
\]

where \( \beta = \{\beta_l : l = 1, 2, \ldots, L\} \) is a parameter set, which controls the impact of fidelity term \( \Gamma(\vec{R}) \) on FCM. We rewrite \( \vec{R} \) as \( \{\vec{R}_l : l = 1, 2, \ldots, L\} \) with \( \vec{R}_l = (r_{j1}, r_{j2}, \ldots, r_{jK})^T \), which indicates that \( \vec{R} \) 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:

\[
\beta \cdot \Gamma(\vec{R}) = \sum_{l=1}^{L} \beta_l \Gamma(\vec{R}_l)
\]

The fidelity term \( \Gamma(\vec{R}_l) \) 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 Gaussian noise estimation, we use an \( \ell_2 \)-norm fidelity term:

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

where \( \| \cdot \|_{\ell_2} \) stands for an \( \ell_2 \) vector norm. In the presence of impulse noise, we choose an \( \ell_1 \)-norm fidelity term:

\[
\Gamma(\vec{R}_l) = \|\vec{R}_l\|_{\ell_1} = \sum_{j=1}^{K} |r_{jl}|
\]

where \( \| \cdot \|_{\ell_1} \) denotes an \( \ell_1 \) vector norm. For Poisson noise, we take the Csiszár’s I-divergence \( (5) \) of \( \vec{R} \) from \( \vec{X} \) as a fidelity term, i.e.,

\[
\Gamma(\vec{R}_l) = \sum_{j=1}^{K} ((x_{jl} - r_{jl}) - x_{jl} \log(x_{jl} - r_{jl}))
\]

For common single noise, i.e., Gaussian, Poisson, and impulse noise, the above fidelity 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 noise rather than a single noise. The fidelity 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 fidelity term for mixed or unknown noise estimation.

### 3.4 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 \( 512 \times 512 \), respectively. We show original and two observed images in Fig. 2:

![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.](image-url)

As Fig. 2(b) shows, Gaussian noise is overall organized. 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 disorganized as shown in Fig. 2(c). In Fig. 3, we portray the distributions of Gaussian and mixed noise, respectively.

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 \( \ell_1 \) norm nor \( \ell_2 \) norm can precisely characterize the residual \( \vec{R} \) in the sense of the MAP estimation.
3.5 Residual-driven FCM with Weighted $\ell_2$-norm Fidelity

Intuitively, if the fidelity term can be modified so as to make mixed noise distribution more Gaussian-like, we can still use $\ell_2$ norm to characterize residual $R$. It means that mixed noise can be more accurately estimated. Therefore, we adopt robust estimation techniques \[44, 45\] 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.

By substituting (6) into (4), combined with (5), we present RFCM with weighted $\ell_2$-norm fidelity (WRFCM) for image segmentation:

$$J(U, V, R, W) = \sum_{i=1}^{c} \sum_{j=1}^{K} w_{ij}^{m} \left( \sum_{n \in N_j} \frac{\|x_n - r_j - v_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^{L} \beta_l \sum_{j=1}^{K} \sum_{n \in N_j} \frac{\|w_{jl}r_{jl}\|^2}{1 + d_{nj}}$$

(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 [42, 54]. If there exists a small distance between a target pixel and its neighbors, they most likely belong to the same cluster. Therefore, we introduce spatial information into (8), thus resulting in our final objective function:

$$J(U, V, R, W) = \sum_{i=1}^{c} \sum_{j=1}^{K} w_{ij}^{m} \left( \sum_{n \in N_j} \frac{\|x_n - r_j - v_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^{L} \beta_l \sum_{j=1}^{K} \sum_{n \in N_j} \frac{\|w_{jl}r_{jl}\|^2}{1 + d_{nj}}$$

(9)

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 $j$, $N_j$ stands for a local window centralized in $j$, and $d_{nj}$ represents the Euclidean distance between $n$ and $j$.

3.6 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 call an Lagrangian multiplier method to minimize (9). The Lagrangian function is expressed as:

$$\mathcal{L}_\Lambda(U, V, R; W) = \sum_{i=1}^{c} \sum_{j=1}^{K} w_{ij}^{m} \left( \sum_{n \in N_j} \frac{\|x_n - r_j - v_i\|^2}{1 + d_{nj}} \right) + \sum_{l=1}^{L} \beta_l \sum_{j=1}^{K} \sum_{n \in N_j} \frac{\|w_{jl}r_{jl}\|^2}{1 + d_{nj}} + \sum_{j=1}^{K} \lambda_j \left( \sum_{l=1}^{L} w_{jl} - 1 \right)$$

(10)

where $\Lambda = \{ \lambda_j : j = 1, 2, \ldots, K \}$ 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 $\varepsilon$, input $W^{(0)}$.

1. For $t = 0, 1, \ldots$, iterate:

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

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

   (11)

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

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

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

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

\[
\begin{align*}
U^{(t+1)} &= \arg\min_L \Lambda(U, V^{(t)}, R^{(t)}; W^{(t)}) \\
V^{(t+1)} &= \arg\min_V \Lambda(U^{(t+1)}, V, R^{(t)}; W^{(t)}) \\
R^{(t+1)} &= \arg\min_R \Lambda(U^{(t+1)}, V^{(t+1)}, R; W^{(t)})
\end{align*}
\]  

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

\[
\begin{align*}
u^{(t+1)}_{ij} &= \frac{\sum_{n \in N_j} \|x_i - r^{(t)}_{n} - u^{(t)}_j\|^2}{1 + d_{nij}} \frac{-1/(m-1)}{c} \\
&= \sum_{d=1}^c \sum_{n \in N_j} \|x_i - r^{(t)}_{n} - u^{(t)}_j\|^2 \frac{-1/(m-1)}{1 + d_{nij}} \\
q^{(t+1)}_l &= \frac{\sum_{j=1}^K \left( u^{(t+1)}_{ij} \right)^m \sum_{n \in N_j} x_i - r^{(t)}_{n}}{1 + d_{nij}} \frac{-1/(m-1)}{c} \\
&= \sum_{j=1}^K \left( u^{(t+1)}_{ij} \right)^m \sum_{n \in N_j} x_i - r^{(t)}_{n} \frac{-1/(m-1)}{1 + d_{nij}} \\

\end{align*}
\]

(13, 14)

Proof. See Appendix.

In the last subproblem of (12), both \(r_j\) and \(r_n\) appear simultaneously. Since \(r_j\) is dependent to \(r_j\), 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, \(n \in N_j\) is equivalent to \(j \in N_n\). Thus we have:

\[
\sum_{j=1}^K \sum_{n \neq j} u^{m}_{ij} f(r_j) + \sum_{j \in N_j} f(r_n) = \sum_{j=1}^K \sum_{n \in N_j} u^{m}_{in} f(r_j)
\]

(15)

where \(f\) represents a function in terms of \(r_j\) or \(r_n\). By (15), we rewrite (9) as:

\[
J(U, V, R, W) = \sum_{i=1}^L \sum_{j=1}^K \sum_{n \in N_j} u^{m}_{in} \|x_i - r_n - u_j\|^2 \\
+ \sum_{l=1}^L \sum_{j=1}^K \sum_{n \in N_j} |w_{lj}r_{n}|^2 \\
\]

(16)

According to the two-step iterative algorithm, we assume that \(\hat{W}\) 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:

\[
r^{(t+1)}_{jl} = \arg\min_{r_{jl}} \left( u^{m}_{jl} \|x_j - r_{jl} - u^{(t)}_{l}\|^2 \frac{-1/(m-1)}{1 + d_{nij}} \right) \\
= \sum_{n \in N_j} \beta_l|w_{lj}r_{n}|^2 \\
+ \sum_{n \in N_j} u^{m}_{jn} \|x_j - r_{jl} - u^{(t)}_{l}\|^2 \\
\]

(17)

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

\[
r^{(t+1)}_{jl} = \frac{c \sum_{i=1}^L \sum_{n \in N_j} (u^{(t+1)}_{in})^m (x_j - u^{(t)}_{l})}{1 + d_{nij}} + \sum_{n \in N_j} |w_{lj}r_{n}|^2
\]

(18)

To show the impact of weighted \(\ell_2\)-norm fidelity 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 \(\xi\) and \(\beta\) are discussed in the later section.

![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.](image)

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.

**Algorithm 2** Residual-driven FCM with weighted \(\ell_2\)-norm fidelity (WRFCM)

**Input:** Observed image \(X\), fuzzification exponent \(m\), number of clusters \(c\), and threshold \(\varepsilon\).

**Output:** Segmented image \(\hat{X}\).

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

### 3.7 Convergence Analysis

In WRFCM, we set \(\|U^{(t+1)} - U^{(t)}\| < \varepsilon\) as the termination condition. In order to analyze the convergence of WRFCM, we take Fig. 5 as a case study. We set \(\varepsilon = 1 \times 10^{-6}\). In Fig. 6 we draw the curves of \(\theta = \|U^{(t+1)} - U^{(t)}\|\) and \(J\) versus iteration step \(t\), respectively.
As Fig. 5 indicates, since the prototypes are randomly initialized, the convergence of WRFCM oscillates slightly at the beginning. Nevertheless, it reaches stability after a few iterations. In addition, even though $\theta$ exhibits an oscillating process, $J$ keeps decreasing until the iteration stops. To sum up, WRFCM has outstanding convergence since the weight $\ell_2$-norm fidelity makes mixed noise distribution estimated accurately so that the residual is gradually separated from observed data as iterations proceed.

4 EXPERIMENTAL STUDIES

In this section, to show the performance, efficiency and robustness of WRFCM, we provide numerical experiments on synthesis, 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 [14], FRFCM [30], WFCM [22], and DSFCM-N [34]. They are the most representative ones in the field.

4.1 Evaluation Indicators

To quantitatively evaluate the performance of WRFCM, we adopt three objective evaluation indicators, i.e., segmentation accuracy (SA) [46], 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:

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

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

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

4.2 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) BrianWeb[1]. 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) [50]. This database contains 200 training, 100 validation and 200 testing images. Golden standard segmentation is annotated by different subjects for each image of size $321 \times 481$ or $481 \times 321$.

3) Microsoft Research Cambridge Object Recognition Image Database (MSRC) [51]. This database contains 591 images and 23 object classes. Golden standard segmentation is provided.

4) NASA Earth Observation Database (NEO) [52]. 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 $1440 \times 720$ contain unknown noise. Therefore, their ground truth is unknown.

4.3 Parameter Settings

Prior to numerical simulations, we report the parameter settings of WRFCM and seven comparative algorithms. Since spatial information is used in all algorithm, a local window of size $3 \times 3$ is selected for all. We set $m = 2$ and $\varepsilon = 1 \times 10^{-6}$ across all algorithms. The setting of $c$ is presented in each experiment.

Except $m$, $\varepsilon$ and $c$, FLICM and KWFLICM are free of any other parameters. However, the remaining algorithms involve different parameters. In FCM$_{S1}$ and FCM$_{S2}$, $\alpha$ 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 $3 \times 3$ structuring element. WFCM requires one parameter $\mu \in [0.55, 0.65]$ only, which constrains the neighbor term. For DSFCM-N, $\lambda$ is set based on the standard deviation of each channel of image data.

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

$$\beta_l = \frac{\phi \cdot \delta_l}{100}$$

where $\delta_l$ is the standard deviation of the $l$-th channel of $X$. In fact, $\beta$ is equivalently replaced by $\phi$. Here, we give an example to show the setting of $\phi$, refer to Fig. 7. We impose a mixture of Poisson, Gaussian, and impulse noise on the first three synthetic images in the second row of Fig. 8 respectively. The noise level is $\sigma = 30$ and $p = 20\%$.

1. http://www.bic.mni.mcgill.ca/brainweb/
2. https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html
3. http://research.microsoft.com/vision/cambridge/recognition/
4. http://neo.sci.gsfc.nasa.gov/
As Fig. 7(a) indicates, when coping with the first image, the SA value reaches its maximum gradually as the value of $\phi$ increases. Afterwards, it decreases rapidly and tends to be stable. As shown in Fig. 7(b), for the other two images, 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 $\phi$ increases from a small value.

4.4 Experimental Results and Analysis

4.4.1 Results on Synthetic Images

In the first experiment, we representatively choose five synthetic images of size $256 \times 256$, as shown in the second row of Fig. 8. A mixture of Poisson, Gaussian, and impulse noise 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 five images, we set $c$ to 4, 4, 4, 3, and 3, respectively. The segmentation results are given in Fig. 9 and Table 1. 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, KWFICLM, 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$_{N}$ gives rise to the change of topological shapes to some extent. Unlike them, FRFCM and WFCM achieve sufficient noise removal. 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 2 shows, WRFCM obtains optimal SA, SDS and MCC results for all five medical images. As a conclusion, it outperforms its peers visually and quantitatively.

4.4.2 Results on Medical Images

Next, we representatively segment five medical images from BrianWeb. They are represented as five slices in the axial plane with a sequence of 70, 80, 90, 100 and 110, which 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 comparison between WRFCM and its peers are shown in Fig. 9 and Table 1. The best values are in bold.

By a view of the marked red square in Fig. 9, we find that FCM$_{S1}$, FCM$_{S2}$, FLICM, KWFICLM 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. 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 2 shows, WRFCM obtains optimal SA, SDS and MCC results for all five medical images. As a conclusion, it outperforms its peers visually and quantitatively.

4.4.3 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 segmentation results of all algorithms are shown in Fig. 10 and Table 3.

Fig. 10 visually shows the comparison between WRFCM and seven peers while Table 3 gives the quantitative comparison. Apparently, WRFCM achieves better segmentation results than its peers. FCM$_{S1}$, FCM$_{S2}$, FLICM, KWFICLM and DSFCM$_{N}$ obtain unsatisfactory results on all tested images. Superior 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

| Algorithm   | SA  | SDS | MCC | SA  | SDS | MCC | SA  | SDS | MCC |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| FCM$_{S1}$ | 92.902 | 98.187 | 96.362 | 92.652 | 98.414 | 95.528 | 87.298 | 99.582 | 97.606 |
| FCM$_{S2}$ | 96.157 | 98.999 | 95.082 | 96.292 | 99.127 | 97.520 | 92.345 | 99.791 | 98.808 |
| FLICM      | 85.081 | 90.145 | 85.667 | 95.894 | 88.576 | 81.902 | 83.077 | 96.074 | 88.031 |
| KWFICLM    | 99.706 | 99.858 | 99.715 | 99.730 | 99.904 | 99.725 | 99.310 | 99.938 | 99.648 |
| FRFCM      | 99.652 | 99.920 | 99.939 | 99.675 | 99.995 | 99.698 | 99.629 | 99.924 | 99.568 |
| WFCM       | 97.827 | 99.325 | 98.652 | 98.079 | 99.363 | 98.197 | 96.645 | 99.735 | 98.485 |
| DSFCM$_{N}$ | 98.074 | 99.256 | 99.757 | 99.503 | 99.003 | 99.303 | 98.501 | 99.931 | 99.563 |
| WRFCM      | 99.859 | 99.937 | 99.843 | 99.902 | 99.958 | 99.792 | 99.785 | 99.931 | 99.853 |
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 2000–2019. The visual results of all algorithms are shown in Figs. 11 and 12. The corresponding SA, SDS, and MCC values are given in Table 4.

### 4.5 Performance Improvement

Besides segmentation results reported for all algorithms, we also present the performance improvement of WRFCM over seven comparative algorithms in Table 5. Clearly, for all types of images, the average SA, SDS and MCC improvements of WRFCM over other peers are within the value span 0.238%–27.836%, 0.039%–41.989%, and 0.047%–58.681%, respectively.

### 4.6 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. For a fair comparison, 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. The execution time of all algorithms for segmenting synthetic, medical, real-world images is presented in Table 6. The mean values are in bold. Moreover, we portray them in Fig. 13.

As Table 6 and Fig. 13 show, for gray and color image segmentation, the computational efficiency of KWFLICM is far lower than the 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 each iteration, FCM_S1 and FCM_S2 are more time-consuming than the others except KWFLICM.

Even though FLICM, 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 2.642 seconds longer than...
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? Answering them needs more research efforts.

**APPENDIX**

**Proof of Theorem 3.1**

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

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

where $D_{ij} = \sum_{n \in N_j} \| \mathbf{x}_n - \mathbf{r}_{ij} \|_2^2$.

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

$$\frac{\partial L_x}{\partial u_{ij}} = m D_{ij} u_{ij}^{m-1} + \lambda_j = 0.$$

Thus, $u_{ij}$ is expressed as:

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

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

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

In the sequel, one can get

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

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

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

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

$$\frac{\partial L_x}{\partial v_i} = -2 \cdot \sum_{j=1}^{K} u_{ij} \sum_{n \in N_j} \frac{1}{1 + d_{nj}}.$$

**TABLE 5**

Average performance improvements (%) of WRFCM over comparative algorithms

| Algorithm       | Synthetic images | Medical images | Real-world images in BSDS and MSRC | Real-world images in NEO |
|-----------------|------------------|---------------|------------------------------------|--------------------------|
|                 | SA    | SD | MCC | SA    | SD | MCC | SA    | SD | MCC | SA    | SD | MCC |
| FCM             | 8.327 | 1.782 | 3.677 | 4.438 | 1.309 | 2.118 | 26.708 | 26.221 | 57.938 | 13.841 | 4.148 | 5.329 |
| FCM             | 4.863 | 3.767 | 7.286 | 4.407 | 0.445 | 0.755 | 26.783 | 41.989 | 58.272 | 13.853 | 4.079 | 5.254 |
| FLICM           | 15.275 | 9.177 | 16.950 | 5.024 | 0.444 | 0.764 | 21.045 | 28.390 | 47.687 | 13.072 | 4.268 | 5.038 |
| KWFICM          | 0.238 | 0.038 | 0.047 | 5.004 | 0.494 | 0.864 | 27.833 | 36.791 | 58.681 | 8.528 | 10.736 | 12.066 |
| FRFCM           | 0.293 | 2.988 | 5.627 | 4.270 | 0.774 | 1.339 | 3.976 | 2.467 | 9.995 | 10.136 | 10.010 | 10.829 |
| WFCM            | 2.100 | 0.716 | 1.484 | 4.883 | 1.769 | 3.087 | 3.852 | 3.831 | 9.689 | 2.368 | 0.830 | 0.811 |
| DSFCM_N         | 0.702 | 3.130 | 6.071 | 4.278 | 4.928 | 8.445 | 23.087 | 30.119 | 46.672 | 10.595 | 9.195 | 10.030 |

**TABLE 6**

Comparison of execution time (in seconds) of all algorithms

| Image           | FCM_S1 | FCM_S2 | FLICM | KWFICM | FRFCM | WFCM | DSFCM_N | FCM_S1 | FCM_S2 | FLICM | KWFICM | FRFCM | WFCM | DSFCM_N |
|-----------------|--------|--------|-------|--------|-------|------|---------|--------|--------|-------|--------|-------|------|---------|
| Fig. 8 column 1 | 40.453 | 32.567 | 4.131 | 63.069 | 2.555 | 2.387 | 8.245   | 4.313  |
| Fig. 9 column 1 | 44.116 | 38.982 | 4.567 | 72.607 | 0.270 | 5.157 | 7.271   | 4.598  |
| Fig. 10 column 1| 67.155 | 49.889 | 3.172 | 102.019| 0.265 | 3.214 | 7.501   | 4.877  |
| Fig. 11 column 1| 40.030 | 31.835 | 3.560 | 71.339 | 0.236 | 2.536 | 4.561   | 3.873  |
| Fig. 12 column 1| 53.060 | 38.189 | 3.341 | 109.381| 0.265 | 2.473 | 4.242   | 3.570  |

**5 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 fidelity term coming with an analysis of noise distribution. On the basis of the framework, RFCM with weighted $\ell_2$-norm fidelity (WRFCM) is presented for coping with image segmentation with mixed or unknown noise. Spatial information is also considered in WRFCM for making residual estimation more reliable. A two-step iterative algorithm is presented to implement WRFCM. Experiments reported for four benchmark databases demonstrate that it outperforms existing FCM variants. Moreover, differing from popular supervised-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], [52], [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]?
Fig. 8. Segmentation results for five synthetic images. The parameters: \(\phi_1 = 5.58\), \(\phi_2 = 7.45\), \(\phi_3 = 8.17\), \(\phi_4 = 5.79\), and \(\phi_5 = 9.99\). From top to bottom: noisy images, noise-free images, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

The intermediate process is presented as:

\[
\sum_{j=1}^{K} u_{ij}^{m} \left( \sum_{n \in \mathcal{N}_j} \frac{x_n - r_n}{1 + d_{nj}} \right) = \sum_{j=1}^{K} u_{ij}^{m} \left( \sum_{n \in \mathcal{N}_j} \frac{v_i}{1 + d_{nj}} \right).
\]

The optimal \(v_i\) is computed:

\[
v_i = \frac{\sum_{j=1}^{K} \left( u_{ij}^{m} \sum_{n \in \mathcal{N}_j} \frac{x_n - r_n}{1 + d_{nj}} \right)}{\sum_{j=1}^{K} \left( u_{ij}^{m} \sum_{n \in \mathcal{N}_j} \frac{1}{1 + d_{nj}} \right)}.
\]

Fig. 9. Segmentation results on five medical images. The parameter: \(\phi = 5.35\). From top to bottom: noisy images, ground truth, and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

ACKNOWLEDGMENTS

This work is 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, in part by
Fig. 10. Segmentation results on five real-world images in BSDS and MSRC. The parameters: $\phi_1 = 6.05$, $\phi_2 = 10.00$, $\phi_3 = 9.89$, $\phi_4 = 9.98$, 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.

Fig. 11. Segmentation results on the first real-world image in NEO. The parameter: $\phi = 6.10$. 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. The parameter: $\phi = 9.98$. From (a) to (i): observed image and results of FCM_S1, FCM_S2, FLICM, KWFLICM, FRFCM, WFCM, DSFCM_N, and WRFCM.

Fig. 13. Average execution time (in seconds) on different types of images.

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. Sziergy, Z. Benyo, S. Sziergy, 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
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, Xi’an University, Xi’an, China.

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

Witold Pedrycz (M’88-SM’90-F’99) received the M.Sc., Ph.D., and D.Sc., degrees from the Silesian University of Technology, Gliwice, Poland.

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 (M’06-SM’07-F’16) 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 Melikshas Universities, 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 (S’88-M’90-SM’93-F’03) 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 in 1990.

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 800 publications including 12 books, 500+ journal papers (400+ in IEEE TRANSACTIONS), 23 patents and 29 book-chapters. He is the founding Editor of IEEE Press Book Series on Systems Science and Engineering and Editor-in-Chief of IEEE/CAA Journal of Automatica Sinica. He is a recipient of a 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. He is founding Co-chair of Enterprise Information Systems Technical Committee (TC) and Environmental Sensing, Networking, and Decision-making TC of IEEE SMC Society. He has been among most highly cited scholars for years and ranked top one in the field of engineering worldwide in 2012 by Web of Science/Thomson Reuters and now Clarivate Analytics. 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).