Fuzzy Double C-Means Clustering Based on Sparse Self-Representation

Jing Gu, Licheng Jiao, *Senior Member, IEEE*, Shuyuan Yang, *Member, IEEE*, and Fang Liu, *Senior Member, IEEE*

***Abstract*—This paper introduces the popular sparse representation method into the classical fuzzy c-means clustering algorithm, and presents a novel fuzzy clustering algorithm, called fuzzy double c-means based on sparse self-representation (FDCM_SSR). The major characteristic of FDCM_SSR is that it can simultaneously address two datasets with different dimensions, and has two kinds of corresponding cluster centers. The first one is the basic feature set that represents the basic physical property of each sample itself. The second one is learned from the basic feature set by solving a spare self-representation model, referred to as discriminant feature set, which reflects the global structure of the sample set. The spare self-representation model employs dataset itself as dictionary of sparse representation. It has good category distinguishing ability, noise robustness, and data-adaptiveness, which enhance the clustering and generalization performance of FDCM_SSR. Experiments on different datasets and images show that FDCM_SSR is more competitive than other state-of-the-art fuzzy clustering algorithms.**

***Index Terms*—Clustering, fuzzy c-means, sparse representation (SR).**

I. Introduction

CLUSTERING is an important data processing method, which has been widely applied in the areas of pattern recognition [1], image processing [2], data mining [3], etc. The aim of clustering is to partition datasets into meaningful groups with similar samples. Fuzzy sets proposed by Zadeh [4] were introduced into the clustering [5]. So far, fuzzy clustering has been widely studied and applied in variety of substantive domains [6]–[10]. Bezdek *et al*. [11] proposed a fuzzy c-means clustering algorithm (FCM) that has become the most well-known fuzzy clustering algorithms. The FCM algorithm introduces the fuzziness into the belongingness of each sample. Each of the samples is assigned a membership grade from 0 to 1. The objective function of the classical FCM algorithm is defined by fuzzy memberships and Euclidean distances of samples to cluster centers.

After the Bezdek's masterwork, many researchers have focused on improving performance of the classical FCM algorithm by different methods and from different perspectives. First of all, to overcome the problem that a point is equidistant from two prototypes in FCM, Krishnapuram and Keller [12], [13] proposed a new clustering model named possibilistic c-means (PCM). On the basis of the PCM, two new models called fuzzy-PCM (FPCM) [14] and possibilistic fuzzy c-means (PFCM) [15] are presented consecutively. Second, spatial contextual information of image was introduced by Ahmed *et al*. [16] into the classical FCM algorithm, so a new FCM method with spatial constraints (FCM_S) is proposed. To accelerate operating speed of FCM_S, its two variants FCM_S1 and FCM_S2 [17] are presented, which, respectively, employed the mean and median filtered images obtained in advance to influence the labeling of image pixels, rather than updating the labeling of all image pixels in each iteration. Meanwhile, an enhanced FCM (EnFCM) [18] algorithm based on the gray levels rather than the image pixels is proposed to improve the running speed of FCM_S. Inspired by EnFCM, Cai *et al*. [19] used the gray and spatial information and proposed a fast generalized FCM (FGFCM) algorithm. Moreover, an idea of weighted mean was embedded into the FCM and a fuzzy weighted c-means (FWCM) algorithm [20] was proposed. After a few years, Hung *et al.* [21] proposed an improved version of FWCM, named as a new weighted fuzzy c-means (NW_FCM). Additionally, in order to guarantee noise insensitiveness and image detail preservation, the local spatial and gray level information is incorporated in a novel fuzzy way, and a new fuzzy local information c-means (FLICM) [22] method is presented. Furthermore, after a generalized fuzzy c-means clustering (GFCM) [23] is presented, a new generalized FCM algorithm with improved fuzzy partition (GIFP_FCM) [24] is proposed to quicken the convergent speed of the FCM algorithm. Afterwards, with the rise of kernel technology [25], [26], Jiao *et al*. [27] presented the kernel version of GIFP_FCM with spatial constraints (KGFCM_S1 and KGFCM_S2) to improve the clustering performance.

However, the existing fuzzy clustering methods do not use global structure information of data set. In the overwhelming

Manuscript received January 29, 2016; revised June 11, 2016 and November 2, 2016; accepted January 18, 2017. Date of publication March 23, 2017; date of current version March 29, 2018. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329402, in part by the National Natural Science Foundation of China under Grant 61573267, Grant 61473215, Grant 61571342, Grant 61572383, Grant 61501353, Grant 61502369, Grant 61271302, Grant 61272282, and Grant 61202176, in part by the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) under Grant B07048, in part by the Major Research Plan of the National Natural Science Foundation of China under Grant 91438201 and Grant 91438103, and in part by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT_15R53.

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: xuer6126@126.com; lchjiao@mail.xidian.edu.cn; syyang@xidian.edu.cn; f63liu@163.com).

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

Digital Object Identifier 10.1109/TFUZZ.2017.2686804

majority of cases, the data used to cluster in the existing fuzzy clustering methods only describe basic physical properties of a set of samples, such as position, shape, size, gray level, and so on. The basic feature only represents the single sample itself, but cannot reflect the single sample's role in whole sample set. But this is precisely crucial to the clustering. Although some existing fuzzy clustering methods employed certain spatial information, it only is spatial relation in local neighborhood. If you know what role each sample played in the overall structure of sample set, the samples with similar role may be classified into same category to achieve the desired clustering result. Though the structure of sample set is unpredictable, a sparse representation (SR) method [28] can dig out correlation among different samples which reflects the global structure of sample set. In this paper, the SR technology that contains global structure information is embedded into the classical FCM algorithm to improve its performance.

The basic theory of SR is that a sample can be represented as the linear combination of small amount of atoms in a dictionary. To avoid spending extra effort to construct the dictionary, the sample set itself is used as dictionary in the SR method. In this paper, it is called sparse self-representation, which means that a sample can be represented as the linear combination of other similar samples in sample set. In recent years, SR [29]–[31] has received significant attention and it has been successfully applied in various fields [32]–[35]. Wright *et al*. [36] first proposed an SR-based classification method which used SR coefficients to classify facial image set. Simultaneously, a new SR-based clustering algorithm, called sparse subspace clustering (SSC) [37], used dataset itself as dictionary to obtain the sparse self-representation coefficients and successfully separate different moving objects in video. In addition, Liu *et al.* proposed a low rank representation (LRR) [38] method to explore sparsity of dataset in another way, and then presented an LRR-based algorithm, named as multitask low-rank affinity pursuit (MLAP) [39], to segment a single natural image in the framework of SSC. SSC, LRR, and MLAP show that the sparse self-representation coefficients have good category distinguishing performance, and at the same time sparse self-representation method has favorable noise robustness and data-adaptiveness, which motivates us to introduce the SR method into fuzzy clustering. The author developed a method in [40], which simply combines the SR with FCM methods and reduces the time and spatial complexity by some strategies, where the SR coefficients obtained by SR model are used for fuzzy clustering as features. But it only can process a feature set just like existing fuzzy clustering algorithms.

In this paper, we propose a new fuzzy double c-means clustering algorithm based on sparse self-representation (FDCM_SSR), which can simultaneously cluster two types of features of a sample set. The two types of features can have different dimensions and distance measures, because each category in FDCM_SSR has two kinds of clustering centers with different dimensions which correspond to the two different feature sets. The first one is the basic feature describing the physical properties of samples themselves by numerical methods. The second kind of feature is obtained by solving sparse self-representation model based on the basic feature set, and called as discriminant feature. The discriminant feature comprises the similar degree between each sample with all of the other samples, which reveals the global structure of whole sample set. Since combining the basic feature with the discriminant feature revealing global structure information of sample set, the FDCM_SSR algorithm has favorable performance in data clustering and image segmentation experiments.

The rest of this paper is organized as follows. Section II reviews the classical FCM algorithm and some of its improved versions. In Section III, a sparse self-representation model which will be used later in this paper is introduced, and then the fundamental theory and formula of the FDCM_SSR method are described in detail. Experimental comparisons in different dataset and images are shown and analyzed in Section IV. Finally, Section V provides conclusion and several issues for future work.

II. Related Work

In this section, the theoretical formulations of the abovementioned fuzzy clustering algorithms are simply listed in Table I. Note that some of them can only be used in image segmentation, due to the usage of the spatial relation among image pixels. To distinguish them in Table I, the variables of equations in the fuzzy clustering algorithms applied to image and any data are denoted in general variable and vector forms, respectively. The meanings of notations used in Table I are described in Table II. In addition, to simplify the content of Table I, a constraint $\sum_{i=1}^{c} u_{ij} = 1$ is omitted for all fuzzy clustering algorithms in Table I.

III. Fuzzy Double C-Means Clustering Algorithm Based on Sparse Self-Representation (FDCM_SSR)

A. Sparse Self-Representation

The most basic SR model [28] is

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad s.t.\ \mathbf{x} = D\mathbf{z} \tag{1}$$

where $\mathbf{x}$ is a sample data; D is a $g \times h (g \ll h)$ dictionary; $\mathbf{z}$ is SR coefficients of $\mathbf{x}$; and $\|\cdot\|_0$ represents ℓ_0 norm. Because the objective function in (1) is a non-deterministic polynomial hard problem, ℓ_0 norm always is replaced with ℓ_1 norm, so that it becomes a convex optimization problem.

As mentioned above, many new SR models have been presented, where the construction of dictionary is crucial. Some of them obtain the dictionary atoms through transformation [41], seeking a large amount of similar samples [42], learning [43], and so on. However, these methods not only need take extra time and effort but also have not data-adaptiveness. So many algorithms employ the dataset itself as the dictionary of the SR model (such as SSC, LRR, and MLAP), which is called as sparse self-representation model in this paper.

In fact, any SR model can be used in the framework of this paper. To demonstrate the role of SR model in our proposed

TABLE I
THEORETICAL FORMULATIONS OF CLASSICAL FCM ALGORITHM AND SOME OF ITS IMPROVED VERSIONS

Algorithm name	Abbreviation	Objective function	Update formula of degree of membership	Update formula of clustering center
Fuzzy C-Means [11]	FCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2,\ \ s.t.\ \ \sum_{i=1}^{c}u_{ij}=1$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\left\|\mathbf{x}_j-\mathbf{v}_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}u_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{n}u_{ij}^m}$
Possibilistic C-Means [12]	PCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}t_{ij}^m\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2+\sum_{i=1}^{c}\lambda_i\sum_{j=1}^{n}\left(1-t_{ij}\right)^m$, where $\lambda_i=\frac{\sum_{j=1}^{n}u_{ij}^m\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\sum_{j=1}^{n}u_{ij}^m}$	$t_{ij}=\left(1+\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\gamma_i}\right)^{\frac{1}{m-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}t_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{n}t_{ij}^m}$
Fuzzy-Possibilistic C-Means [14]	FPCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(u_{ij}^m+t_{ij}^\theta\right)\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2,\ \ s.t.\ \ \sum_{j=1}^{n}t_{ij}=1$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\left\|\mathbf{x}_j-\mathbf{v}_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1},\ \ t_{ij}=\left(\sum_{r=1}^{n}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\left\|\mathbf{x}_r-\mathbf{v}_i\right\|^2}\right)^{\frac{1}{\theta-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}\left(u_{ij}^m+t_{ij}^\theta\right)\mathbf{x}_j}{\sum_{j=1}^{n}\left(u_{ij}^m+t_{ij}^\theta\right)}$
Possibilistic Fuzzy C-Means [15]	PFCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(\alpha u_{ij}^m+\beta t_{ij}^\theta\right)\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2+\sum_{i=1}^{c}\lambda_i\sum_{j=1}^{n}\left(1-t_{ij}\right)^\theta$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2}{\left\|\mathbf{x}_j-\mathbf{v}_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1},\ \ t_{ij}=\left(1+\left(\frac{\beta}{\gamma_i}\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2\right)^{\frac{1}{\theta-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}\left(\alpha u_{ij}^m+\beta t_{ij}^\theta\right)\mathbf{x}_j}{\sum_{j=1}^{n}\left(\alpha u_{ij}^m+\beta t_{ij}^\theta\right)}$
FCM with Spatial constraints [16]	FCM_S	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|x_j-v_i\right\|^2+\frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left(\sum_{r\in N_j}\left\|x_r-v_i\right\|^2\right)$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|x_j-v_i\right\|^2+\frac{\alpha}{N_R}\sum_{r\in N_j}\left\|x_r-v_i\right\|^2}{\left\|x_j-v_k\right\|^2+\frac{\alpha}{N_R}\sum_{r\in N_j}\left\|x_r-v_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n}u_{ij}^m\left(x_j+\frac{\alpha}{N_R}\sum_{r\in N_j}x_r\right)}{(1+\alpha)\sum_{j=1}^{n}u_{ij}^m}$
Two variants of FCM_S [17]	FCM_S1	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|x_j-v_i\right\|^2+\alpha\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\bar{x}_j-v_i\right\|^2$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|x_j-v_i\right\|^2+\alpha\left\|\bar{x}_j-v_i\right\|^2}{\left\|x_j-v_k\right\|^2+\alpha\left\|\bar{x}_j-v_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n}u_{ij}^m\left(x_j+\alpha\bar{x}_j\right)}{(1+\alpha)\sum_{j=1}^{n}u_{ij}^m}$
	FCM_S2	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|x_j-v_i\right\|^2+\alpha\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\tilde{x}_j-v_i\right\|^2$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|x_j-v_i\right\|^2+\alpha\left\|\tilde{x}_j-v_i\right\|^2}{\left\|x_j-v_k\right\|^2+\alpha\left\|\tilde{x}_j-v_k\right\|^2}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n}u_{ij}^m\left(x_j+\alpha\tilde{x}_j\right)}{(1+\alpha)\sum_{j=1}^{n}u_{ij}^m}$
Enhanced FCM [18]	EnFCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{l}\gamma_j u_{ij}^m\left(\xi_j-v_i\right)^2$	$\xi_j=\frac{1}{1+\alpha}\left(x_j+\frac{\alpha}{N_R}\sum_{r\in N_j}x_r\right)$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\xi_j-v_i}{\xi_j-v_k}\right)^{\frac{2}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{q}\gamma_j u_{ij}^m\xi_j}{\sum_{j=1}^{q}\gamma_j u_{ij}^m}$
Fast Generalized FCM [19]	FGFCM		$\xi_j=\frac{\sum_{r\in N_j}S_{jr}x_r}{\sum_{r\in N_j}S_{jr}}$, where $S_{jr}=\begin{cases}e^{-\frac{\max\left(\left\|p_j-p_r\right\|,\left\|q_j-q_r\right\|\right)}{\lambda_s}-\frac{\left\|x_j-x_r\right\|^2}{\lambda_g\sigma_j^2}}, & j\neq r\\ 0, & j=r\end{cases}$ and $\sigma_j=\sqrt{\frac{\sum_{r\in N_j}\left\|x_j-x_r\right\|^2}{N_R}}$		
Fuzzy Weighted C-Means [20]	FWCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\mathbf{x}_j-M_{ij}\right\|^2$	$M_{ij}=\frac{\sum_{r=1,r\neq j}^{n}\left\|\mathbf{x}_j-\mathbf{x}_r\right\|^{-1}u_{ir}\mathbf{x}_r}{\sum_{r=1,r\neq j}^{n}\left\|\mathbf{x}_j-\mathbf{x}_r\right\|^{-1}u_{ir}}$	$u_{ij}=\zeta_j^{1/(m-1)}\left(\left\|\mathbf{x}_j-M_{ij}\right\|^2 m\sum_{l=1}^{c}u_{ij}^{m-1}\right)^{1/(1-m)}$, where $\zeta_j=\left(\sum_{i=1}^{c}\left(\left\|\mathbf{x}_j-M_{ij}\right\|^2 m\sum_{i=1}^{c}u_{ij}^{m-1}\right)^{1/(1-m)}\right)^{1-m}$	
New Weighted Fuzzy C-Means [21]	NW_FCM		$M_{ij}=\frac{\sum_{r=1,r\neq j}^{n}\left\|\mathbf{x}_r-\mathbf{v}_i\right\|^{-1}u_{ir}\mathbf{x}_r}{\sum_{r=1,r\neq j}^{n}\left\|\mathbf{x}_r-\mathbf{v}_i\right\|^{-1}u_{ir}}$		$\mathbf{v}_i=\frac{\sum_{j=1}^{n}u_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{n}u_{ij}^m}$
Fuzzy Local Information C-Means [22]	FLICM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(u_{ij}^m\left\|x_j-v_i\right\|^2+G_{ij}\right)$, where $G_{ij}=\sum_{\substack{r\in N_j\\ j\neq r}}\frac{1}{1+d_{jr}}\left(1-u_{ir}\right)^m\left\|x_r-v_i\right\|^2$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|x_j-v_i\right\|^2+G_{ij}}{\left\|x_j-v_k\right\|^2+G_{kj}}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n}u_{ij}^m x_j}{\sum_{j=1}^{n}u_{ij}^m}$
Generalized FCM [23]	GFCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\mathbf{x}_j-\mathbf{v}_i\right\|_p^p,\ \ s.t.\ \ \sum_{i=1}^{c}u_{ij}=1$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|_p^p}{\left\|\mathbf{x}_j-\mathbf{v}_k\right\|_p^p}\right)^{\frac{1}{m-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}u_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{n}u_{ij}^m}$
Generalized FCM with Improved Fuzzy Partition [24]	GIFP_FCM	$J_m=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2+\sum_{j=1}^{n}a_j\sum_{i=1}^{c}u_{ij}\left(1-u_{ij}^{m-1}\right)$, where $a_j=\alpha\cdot\min\left(\left\|\mathbf{x}_j-\mathbf{v}_k\right\|^2\middle\|k\in\{1,\cdots,c\}\right)$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{\left\|\mathbf{x}_j-\mathbf{v}_i\right\|^2-a_j}{\left\|\mathbf{x}_j-\mathbf{v}_k\right\|^2-a_j}\right)^{\frac{1}{m-1}}\right)^{-1}$	$\mathbf{v}_i=\frac{\sum_{j=1}^{n}u_{ij}^m\mathbf{x}_j}{\sum_{j=1}^{n}u_{ij}^m}$

TABLE I
CONTINUED....

Algorithm name	Abbreviation	Objective function	Update formula of degree of membership	Update formula of clustering center
Two Kernel variants of GIFP_FCM [27]	KGFCM_S1	$J_m = 2\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m\left(1-K\left(x_j,v_i\right)\right)+\sum_{j=1}^{n} a_j\sum_{i=1}^{c} u_{ij}\left(1-u_{ij}^{m-1}\right)$ $+2\beta\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m\left(1-K\left(\bar{x}_j,v_i\right)\right)$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{2\left(1-K\left(x_j,v_i\right)\right)-a_j+2\beta\left(1-K\left(\bar{x}_j,v_i\right)\right)}{2\left(1-K\left(x_j,v_k\right)\right)-a_j+2\beta\left(1-K\left(\bar{x}_j,v_k\right)\right)}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n} u_{ij}^m\left(K\left(x_j,v_i\right)\cdot x_j+\beta K\left(\bar{x}_j,v_i\right)\cdot\bar{x}_j\right)}{\sum_{j=1}^{n} u_{ij}^m\left(K\left(x_j,v_i\right)+\beta K\left(\bar{x}_j,v_i\right)\right)}$
	KGFCM_S2	$J_m = 2\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m\left(1-K\left(x_j,v_i\right)\right)+\sum_{j=1}^{n} a_j\sum_{i=1}^{c} u_{ij}\left(1-u_{ij}^{m-1}\right)$ $+2\beta\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m\left(1-K\left(\tilde{x}_j,v_i\right)\right)$	$u_{ij}=\left(\sum_{k=1}^{c}\left(\frac{2\left(1-K\left(x_j,v_i\right)\right)-a_j+2\beta\left(1-K\left(\tilde{x}_j,v_i\right)\right)}{2\left(1-K\left(x_j,v_k\right)\right)-a_j+2\beta\left(1-K\left(\tilde{x}_j,v_k\right)\right)}\right)^{\frac{1}{m-1}}\right)^{-1}$	$v_i=\frac{\sum_{j=1}^{n} u_{ij}^m\left(K\left(x_j,v_i\right)\cdot x_j+\beta K\left(\tilde{x}_j,v_i\right)\cdot\tilde{x}_j\right)}{\sum_{j=1}^{n} u_{ij}^m\left(K\left(x_j,v_i\right)+\beta K\left(\tilde{x}_j,v_i\right)\right)}$

TABLE II
NOTATIONS USED IN TABLE I

Notation	Means	Notation	Means
$X=(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n)$	Sample set	$\bar{x}_j$	Average of neighboring pixels lying within a window around pixel x_j
n	Number of samples	$\widetilde{x}_j$	Median of neighboring pixels lying within a window around pixel x_j
c	Number of clusters	l	Number of gray levels
u_{ij}, t_{ij}	Degree of membership of the jth sample in ith cluster	γ_j	Number of pixels having gray value equal to j
$\mathbf{v}_i$	Center of ith cluster	(p_j, q_j)	Spatial coordinate of the jth pixel
m, θ	Weighting exponents	d_{jr}	Euclidean distance between pixels j and r
$\alpha, \beta, \lambda_s, \lambda_g$	Scale parameters	P	Norm
N_j	Set of neighbors falling into a window around jth pixel	$K(x,y)$	Kernel function
N_R	Cardinality of N_j		

method, we employ a simple sparse self-representation model

$$\min_{Z} \|Z\|_1 + \lambda \|X - XZ\|_F^2 \quad s.t.\ \mathrm{diag}(Z) = 0 \qquad (2)$$

where X denotes the basic feature set of all samples, a column of which represents the basic feature vector of a sample; Z is sparse self-representation coefficient of X, $\|Z\|_1 = \sum_{j,r} |z_{jr}|$; λ is parameter to balance the effect of the fidelity term; and $\|\cdot\|_F$ is Frobenius norm. The constraint $\mathrm{diag}(Z) = 0$ is to avoid that the solution of (2) is identity matrix.

The alternating direction method of multipliers (ADMM) [44] is used to solve (2). First, we introduce an auxiliary variable H, and convert it into the following equivalent form:

$$\min_{H,Z} \|H\|_1 + \lambda \|X - XZ\|_F^2 \quad s.t.\ Z = H,\ \mathrm{diag}(H) = 0. \qquad (3)$$

Then, minimize the following augmented Lagrange function:

$$\|H\|_1 + \lambda \|X - XZ\|_F^2 + \langle U, Z - H\rangle + \frac{\mu}{2}\|Z - H\|_F^2$$

where U is Lagrange multiplier and $\mu > 0$ is a penalty parameter. By ADMM, the objective function in (3) can be divided into several subproblems that have closed-form solutions. The detailed steps for solving (3) are outlined in Algorithm I.

In Algorithm I, the objective function in (4) is solved by the soft thresholding operator [45].

The larger the obtained sparse self-representation coefficient z_{jr} $(1 \le j \le n, 1 \le r \le n)$, the more similar the corresponding two samples j and r, and vice versa. So the sparse self-representation coefficients obtained by solving (2) show the similarity between every two samples in sample set, which reflect the global structure of sample set.

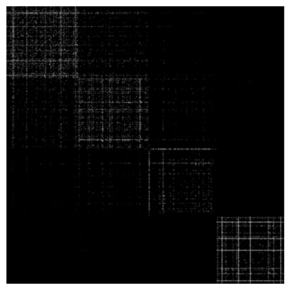

Fig. 1. Schematic of sparse self-representation coefficient matrix.

In order to intuitively understand how sparse self-representation coefficients reflect the global structure, the sparse self-representation coefficients are shown in image form. Take four-class artificial data set with ground truth as example. For viewing purpose, after the sparse self-representation coefficient matrix Z is solved, first compute $G = \frac{1}{2}(|Z| + |Z|^T)$. Then, G is sorted according to categories of samples in ground truth. Fig. 1 shows G rearranged in image form. It can be seen that the image is very dark in Fig. 1, which has two reasons. The first reason is that the range of the sparse self-representation coefficients is very large (about 10^{-10} - - 10^4) and the gray levels of image are only 256. On the other hand, the number of the large sparse self-representation coefficients is very small, which shows the coefficients solved are sparse. More importantly, as you can see from Fig. 1, G rearranged is a four-block diagonal structure. There are many bright points in every block diagonal region, which shows that the samples of same class

Algorithm I: Solving (3) by ADMM.

Input: data set $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, parameter λ.

Initialize: convergence threshold $\eta = 10^{-5}$, $\mu = 10^{-6}$ and $\varepsilon = 1.1$, $Z = H = 0$.

Iterate: while $\|Z - H\|_\infty < \eta$

1. Update each column of H

$$\arg\min_{\mathbf{h}_j} \frac{1}{\mu}\left\|\mathbf{h}_j\right\|_1 + \frac{1}{2}\left\|\mathbf{h}_j - \left(\mathbf{z}_j + \frac{\mathbf{u}_j}{\mu}\right)\right\|_2^2 \quad (4)$$

where $\mathbf{h}_j = (h_{1j}, h_{2j}, \ldots, h_{(j-1)j}, h_{(j+1)j}, \ldots, h_{nj})^T$, $j = 1, \ldots, n$, $\mathbf{h}_j = (h_{1j}, h_{2j}, \ldots, h_{nj})^T$ is the jth column vector of matrix H; similarly, $\mathbf{z}_j = (z_{1j}, z_{2j}, \ldots, z_{(j-1)j}, z_{(j+1)j}, \ldots, z_{nj})^T$, $\mathbf{u}_j = (u_{1j}, u_{2j}, \ldots, u_{(j-1)j}, u_{(j+1)j}, \ldots, u_{nj})^T$. Let $\tilde{\mathbf{h}}_j$ be the optimal solution of (4). By inserting a zero entry at the jth position of $\tilde{\mathbf{h}}_j$, we obtain jth column vector of the matrix H, $\tilde{\mathbf{h}}_j = (\tilde{h}_{1j}, \tilde{h}_{2j}, \ldots, \tilde{h}_{(j-1)j}, 0, \tilde{h}_{(j+1)j}, \ldots, \tilde{h}_{nj})^T$, which meets the constraint $\mathrm{diag}(H) = 0$.

2. Update $Z = (X^T X + \frac{\mu}{2\lambda} I)^{-1}(X^T X + \frac{1}{2\lambda}(\mu H - U))$, where I is identity matrix of size n.

3. Update multiplier $U = U + \mu(Z - H)$.

4. Update parameter $\mu = \min(\varepsilon\mu, 10^{10})$.

Output: the sparse representation coefficient matrix Z.

have high similarities. In contrast, the other regions of Fig. 1 are dark, which means that the correlations among samples from different classes are zero or near zero. Obviously, the sparse self-representation coefficient matrix containing block diagonal structure can be easily and accurately classified.

B. Fuzzy Double C-Means Based on Sparse Self-Representation (FDCM_SSR)

Many improved algorithms of FCM have been shown in previous section, but they can only address a dataset or a kind of feature of sample set. In this paper, a novel FDCM_SSR is proposed. FDCM_SSR can deal with two types of features with different dimensions. One of them is the generally mentioned feature, known as basic feature and denoted by X. The basic feature of each sample only represents the basic physical properties of itself. Another one is learned from the basic feature by utilizing the spare self-representation method. The sparse self-representation coefficients deduced by the spare self-representation model embody similarity of samples from the same category and variance among different classes' samples, which contributes to the clustering. As the sparse self-representation coefficients have good category distinguishing performance, it is taken as the second kind of feature and referred to as discriminant feature. Let Z be the discriminant feature set of all samples.

By using two datasets X and Z, the objective function of the proposed algorithm is defined by

$$J_m = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \left(\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_i\|^2\right)$$
$$s.t. \sum_{i=1}^{c} u_{ij} = 1 \quad (5)$$

where $\mathbf{x}_j$ and $\mathbf{z}_j$ are the jth column of X and Z, respectively, $\mathbf{v}_i$ and $\tilde{\mathbf{v}}_i$ are the center of X and Z in the ith cluster, respectively, and α is parameter to control the effect of the discriminant feature. In other words, $\mathbf{x}_j$ and $\mathbf{z}_j$ are the basic feature and the discriminant feature of the jth sample, respectively. Similarly, $\mathbf{v}_i$ and $\tilde{\mathbf{v}}_i$ are the ith cluster center of the basic feature set and the discriminant feature set, respectively.

To solve the minimization of (5), the cost function in (5) is first rewritten as

$$f(u_{ij}) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m \left(\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_i\|^2\right) + \beta\left(\sum_{i=1}^{c} u_{ij} - 1\right).$$

By differentiating $f(u_{ij})$ with respect to u_{ij}, $\mathbf{v}_i$, $\tilde{\mathbf{v}}_i$, and the Lagrange multiplier β, we have the following formulations:

$$u_{ij} = \left(-\frac{\beta}{m\left(\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_i\|^2\right)}\right)^{\frac{1}{m-1}}, \quad (6)$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \cdot \mathbf{x}_j}{\sum_{j=1}^{n} u_{ij}^m}, \quad (7)$$

$$\tilde{\mathbf{v}}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \cdot \mathbf{z}_j}{\sum_{j=1}^{n} u_{ij}^m}, \quad (8)$$

$$\sum_{i=1}^{c} u_{ij} = 1. \quad (9)$$

Then, through plugging (6) into the equation in (9), we can get

$$(-\beta)^{\frac{1}{m-1}} = \left(\sum_{i=1}^{c}\left(\frac{1}{m\left(\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_i\|^2\right)}\right)^{\frac{1}{m-1}}\right)^{-1}.$$

So

$$u_{ij} = \left(\sum_{k=1}^{c}\left(\frac{\|\mathbf{x}_j - \mathbf{v}_i\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_i\|^2}{\|\mathbf{x}_j - \mathbf{v}_k\|^2 + \alpha \|\mathbf{z}_j - \tilde{\mathbf{v}}_k\|^2}\right)^{\frac{1}{m-1}}\right)^{-1}. \quad (10)$$

The procedure of FDCM_SSR clustering algorithm is described in Algorithm II.

An obvious difference between the proposed FDCM_SSR algorithm with existing fuzzy clustering methods is that

Algorithm II: Fuzzy Double C-Means Based on Sparse Self-Representation (FDCM_SSR).

Step 1: Input the basic feature set $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and the number of cluster c.
Step 2: Initialize the weighting exponent m, convergence threshold η, the degree of membership matrix $U^{(0)} = \{u_{ij}^{(0)}, 1 \le i \le c, 1 \le j \le n\}$, and loop counter $f = 0$.
Step 3: Obtain discriminant feature set Z by solving a sparse self-representation model in (2).
Step 4: Update the loop counter $f = f + 1$.
Step 5: Update the cluster centers $\mathbf{v}_i$ $(1 \le i \le c)$ of the basic feature set X by using (7).
Step 6: Update the cluster centers $\tilde{\mathbf{v}}_i$ $(1 \le i \le c)$ of the discriminant feature set Z by using (8).
Step 7: Update the degree of membership matrix $U^{(f)} = \{u_{ij}^{(f)}\}$ by using (10).
Step 8: If $\max(U^{(f)} - U^{(f-1)}) < \eta$ then stop, otherwise go to Step 4.

FDCM_SSR can process simultaneously two kinds of features of same sample set. Different types of features cannot be simply concatenated, because different feature descriptors may potentially have different data distribution probabilities and belong to different feature spaces. Recently, many methods [46], [47] used multiple kinds of features to comprehensively describe objects in real world applications. Thus, we employ two types of features in FDCM_SSR algorithm to fully describe different characteristics of samples. Furthermore, one of them can reflect the global structure of sample set, which is helpful to obtain more accurate clustering result. These make sure that the FDCM_SSR algorithm has favorable robustness to noise and clustering accuracy (CA).

C. Computational Complexity Analysis

In this section, the computational complexities of the existing fuzzy clustering methods mentioned in the previous section and the proposed FDCM_SSR algorithm are analyzed one by one. Let us suppose that the number of iterations is t and the dimension of the basic feature of each sample is b. The time complexity of the traditional FCM method is $O(cnbt)$. In fact, the FCM_S, FCM_S1, FCM_S2, FLICM, GIFP_FCM, KGFCM_S1, and KGFCM_S2 algorithms basically introduce different constraints into the cost function of the classical FCM method, so their computational complexities are about $O(cnbt)$. Since the EnFCM and FGFCM methods are on the base of the gray levels, their computing complexities are $O(256cbt)$. In general, the number of samples n is greater than 256, so that $O(256cbt)$ is usually lower than $O(cnbt)$. Additionally, the time complexities of FWCM and NW_FCM are $O(cnbt(n-1))$, as there are actually $n-1$ calculations to get the unsupervised weighted mean M_{ij} in each iteration. The dimension of the discriminant feature in proposed FDCM_SSR algorithm is the same with the number of samples n; thus, the computational complexity of FDCM_SSR is $O(cnt(b+n))$, which is lower than that of FWCM and NW_FCM.

When these clustering methods are used to segment images, the number of samples n is the number of image pixels. In this case, the runtimes of FWCM, NW_FCM, and FDCM_SSR are very long. To improve their efficiency, the superpixels are employed to replace the image pixels in this paper. Generally, the range of number of superpixels is about from 600 to 1200, and the number of image pixels n is 65 536 for a small image with the size of 256×256, so the number of superpixels is much smaller than that of image pixels. Hence, the time complexity of FDCM_SSR is greatly reduced by the superpixels instead of the image pixels.

IV. Experimental Results and Analysis

To fully evaluate the performance of the proposed FDCM_SSR algorithm, the experimental results of FDCM_SSR are compared with those of other state-of-the-art clustering algorithms on UCI datasets, artificial images, natural images, and synthetic aperture radar (SAR) images. Since PCM, FPCM, FCM_S, and FWCM have the corresponding improved versions, i.e., PFCM, FCM_S1, FCM_S2, and NW_FCM, the compared methods of FDCM_SSR do not include PCM, FPCM, FCM_S, and FWCM algorithms in the experiment. Moreover, FDCM_SSR is also compared with other common clustering methods, including K-means, mean shift, and spectral clustering. Since the existing SSC, LRR, and MLAP algorithms employed spectral clustering to get final result, FDCM_SSR is compared with SR-based spectral clustering (SR + SC) (that is, discriminant feature based spectral clustering) in this paper.

First, the clustering results of FDCM_SSR and other seven clustering algorithms on UCI datasets are shown, as other fuzzy clustering algorithms mentioned in the second section only can be applied in image segmentation. In order to obtain an objective comparison, the following three metrics are used to evaluate the clustering results: CA, rand index (RI), and normalized mutual information (NMI) [48]. The CA is commonly used to evaluate the clustering performance of different methods. The RI and NMI are, respectively, defined as

$$\mathrm{NMI} = \frac{2I(P;Q)}{H(P) + H(Q)}, \tag{11}$$

$$\mathrm{RI} = \frac{h + w}{n \times (n-1)/2} \tag{12}$$

where P is known label set of samples, Q is label set of the clustering result for the feature set of samples, $I(P;Q)$ is the mutual information of P and Q, $H(P)$ and $H(Q)$ are the entropy of P and Q, respectively, h is the number of any two samples belonging to the same cluster in P and Q, w is the number of any two samples belonging to two different clusters in P and Q, and n is the number of samples. Obviously, $\mathrm{NMI} = \mathrm{RI} = 1$ if $P = Q$. The range of NMI and RI is from 0 to 1. The larger the NMI and RI, the higher the similarity between P and Q.

Then, the performance of FDCM_SSR is compared with that of 14 existing clustering methods on the segmentation of different images. Finally, the influence of the parameters λ

TABLE III
DESCRIPTIONS OF FOUR UCI BENCHMARK DATASETS

Dataset	Number of samples	Number of features	Number of clusters
Qualitative_Bankruptcy	175	6	2
Auto MPG	392	7	3
Balance Scale	625	4	3
Contraceptive Method Choice	1473	9	3

TABLE IV
PERFORMANCE OF EIGHT ALGORITHMS FOR DIFFERENT DATASETS

		K-means	Mean shift	SR + SC	FCM	PFCM	NW_FCM	GIFP_FCM	FDCM_SSR
Qualitative_Bankruptcy	CA	0.9400	0.9400	0.7943	0.9371	0.9371	0.9371	0.9371	**0.9429**
	RI	0.8828	0.8828	0.6713	0.8815	0.8815	0.8815	0.8815	**0.8916**
	NMI	0.6519	0.6660	0.2684	0.6748	0.6748	0.6748	0.6748	**0.6764**
	time	0.0819	0.0227	6.3694	0.0278	0.0316	2.4575	0.0141	2.8738
Auto MPG	CA	0.4490	0.3622	0.4847	0.4566	0.4566	0.4643	0.4617	**0.6250**
	RI	0.5481	0.5532	0.5518	0.5518	0.5518	0.5603	0.5600	**0.5912**
	NMI	0.2044	0.2091	0.2350	0.1998	0.1998	0.2086	0.2080	**0.2358**
	time	0.0682	0.0127	31.6618	0.0462	0.0376	43.1315	0.0561	22.4969
Balance Scale	CA	0.5168	0.5648	0.5536	0.4704	0.5440	0.4608	0.5376	**0.7552**
	RI	0.5780	0.5492	0.5877	0.5088	0.5727	0.4299	0.6184	**0.7043**
	NMI	0.0960	0.1334	0.1656	0.0039	0.0769	0.0051	0.2051	**0.2779**
	time	0.0645	0.0154	82.0305	0.1105	0.0638	45.9171	0.3138	46.3874
Contraceptive Method Choice	CA	0.4012	0.4087	0.4308	0.3917	0.3917	0.3978	0.3958	**0.4331**
	RI	0.5566	0.5563	0.5546	0.5582	0.5582	**0.5612**	0.5592	0.5603
	NMI	0.0321	0.0305	0.0330	0.0330	0.0330	**0.0341**	0.0328	0.0333
	time	0.0727	0.0720	401.0434	0.0718	0.0680	309.8718	0.0796	313.7476

and α to the segmentation result is analyzed by two series of experiments.

A. UCI Dataset

To validate the clustering performances of the proposed FDCM_SSR algorithm, it is compared with K-means, mean shift, SR + SC, FCM, PFCM, NW_FCM, and GIFP_FCM on four UCI benchmark clustering datasets. The brief descriptions of the four datasets are given in Table III, and the scales of these datasets are from a few hundred to a few thousand.

It is generally known that initial clustering centers and degree of membership of fuzzy clustering methods are random, so each algorithm was performed 20 times and the mean result is only retained. For qualitative_bankruptcy, auto MPG, balance scale, and contraceptive method choice datasets, the bandwidth parameters of mean shift [49] method are, respectively, set to 10, 320, 2.5, and 4.2 to make the number of classes of clustering results equivalent to the ideal value in Table III. In our experiment, the parameters of PFCM [15] in Table I are adjusted to be the best with $\alpha = 1$, $\beta = 1$, and $\theta = 2$, and the parameter α in GIFP_FCM [24] algorithm is set to 0.9. Simultaneously, the parameters of the proposed algorithm have been also adjusted to be the best with $\lambda = 0.5$ in (2) and $\alpha = 1.2$ in (5), and the parameter λ of the SR model in (2) is also set to 0.5 for the SR + SC method. In addition, we set the weighting exponent parameter $m = 2$ for all fuzzy clustering methods in this paper.

Table IV shows the clustering results and running time (s) of FDCM_SSR and other seven clustering algorithms on four UCI datasets. As can be seen from Table IV, the clustering results of FDCM_SSR are better than those of other clustering methods, except that the RI and NMI of FDCM_SSR on the contraceptive method choice dataset are slightly below that of NW_FCM. But the difference between the RI and NMI of FDCM_SSR with that of NW_FCM on the contraceptive method choice dataset is only in (e^{-4}, e^{-3}).

In this paper, the computational costs of these algorithms are assessed via runtime on a personal computer with 3.20 GHz Intel(R) Core(TM) i3 processors, 3 GB memory and Windows XP operating system, using programs written by MATLAB R2010a. As seen in Table IV, the FDCM_SSR algorithm runs slowly than K-means, mean shift, SR + SC, FCM, PFCM, and GIFP_FCM, whereas the runtime of FDCM_SSR and NW_FCM is roughly equal.

B. Artificial Images

Many existing fuzzy clustering methods, such as some algorithms mentioned in Section II, are proposed to improve performance in image segmentation. So the segmentation performance of the proposed FDCM_SSR algorithm is tested by implementing the experiments on two artificial images. The segmentation results can be directly evaluated by the runtime and segmentation accuracy (SA), which is calculated by using the corresponding ground truth. Similarly with the above-mentioned section, each algorithm is performed 20 times and the mean value of their segmentation accuracies and the corresponding segmentation result are shown.

In the image segmentation experiments, we compare the proposed FDCM_SSR algorithm with the following

TABLE V
PARAMETER SETTINGS OF DIFFERENT ALGORITHMS

Methods	SR + SC	PFCM	EnFCM	FGFCM	GIFP_FCM	KGFCM_S1	KGFCM_S2	FDCM_SSR
Parameters	$\lambda = 0.5$	$\alpha = 1,\beta = 1,\theta = 2$	$\alpha = 1$	$\lambda_g = 0.5,\lambda_s = 3$	$\alpha = 0.9$	$\alpha = 0.7,\beta = 6$	$\alpha = 0.7,\beta = 6$	$\lambda = 0.5,\alpha = 1.2$

clustering methods: K-means, mean shift, SR + SC, FCM, PFCM, FCM_S1, FCM_S2, EnFCM, FGFCM, NW_FCM, FLICM, GIFP_FCM, KGFCM_S1, and KGFCM_S2. To be fair, a mean shift based image segmentation method [50] with the highest citation is employed. For different images, the parameters (h_s, h_r, W) of mean shift are adjusted to make the visual effects and accuracy of segmentation results to be the best. Except the mean shift algorithm, the parameters of other methods are adjusted to be the best in the experiment, as shown in Table V. Other algorithms (except mean shift) that do not contain parameters are not listed in Table V. Moreover, the notations in Table V are consistent with that in Table I. In the following experiments, the parameters of different methods (except mean shift) are also set according to Table V.

Additionally, the SR + SC, NW_FCM, and FDCM_SSR algorithms belonging to pairwise grouping methods involve high cost of computation and storage, because the number of image pixels is large. So we perform the SR + SC, NW_FCM, and FDCM_SSR algorithms based on the superpixels instead of pixels in the experiment. Turbopixels [51] method is often used to obtain superpixels. In our experiment, the number of superpixels in the Turbopixels algorithm is set to 800. Then, gray histogram and mean value of each superpixel are extracted to form the feature of each superpixel, which has 257 dimensions. Finally, the segmentation results of SR + SC, NW_FCM, and FDCM_SSR are obtained by clustering the feature set of all superpixels.

Fig. 2(a) shows a two-class image with the size of 256×256, which has two gray levels 60 and 130. We artificially add salt and pepper noise 20% to the two-class image and get artificial noise image, as shown in Fig. 2(b). Superpixels obtained by implementing Turbopixels algorithm on the noise image are shown in Fig. 2(c). Fig. 2 shows the segmentation results of different clustering methods in Fig. 2(b). Fig. 2(d), (g), (h), (i), (j), (o), and (p), respectively, shows the segmentations of K-means, FCM, PFCM, FCM_S1, FCM_S2, GIFP_FCM, and KGFCM_S1, which are seriously spotty in consistent regions. The segmentation results of the EnFCM and FGFCM algorithms are dreadful, as shown in Fig. 2(k) and (l), respectively. Fig. 2(f), (m), (n), and (q), obtained, respectively, by SR + SC, NW_FCM, FLICM, and KGFCM_S2, improves the regional consistency, but many pixels are misclassified. In contrast, the segmentation results of mean shift with $(h_s, h_r, W) = (7, 6.5, 20)$ and FDCM_SSR are favorable, as shown in Fig. 2(e) and (r), respectively. The segmentation accuracies of different methods are shown in Table VI, which is consistent with the visual results.

Similarly with Fig. 2(a), we first generate a 256×256 three-class image (that has three gray levels 60, 128, and 200), as shown in Fig. 3(a). Then, Gaussian noise (20%) is added to Fig. 3(a), and the artificial noise image Fig. 3(b) is obtained. Likewise, Fig. 3(c) shows the superpixels obtained by the Turbopixels algorithm. The segmentation results of different methods for the three-class noise image are shown in Fig. 3. Although the segmentations of K-means, PFCM, FCM_S1, FCM_S2, FLICM, GIFP_FCM, KGFCM_S1, and KGFCM_S2 basically distinguish three classes, they have many spots in consistent regions, as shown in Fig. 3(d), (h), (i), (j), (n), (o), (p), and (q), respectively. Fig. 3(g), (k), (l), and (m), obtained, respectively, by FCM, EnFCM, FGFCM, and NW_FCM, wholly confuses two classes. By contrast, Fig. 3(e), (f) and (r) obtained, respectively, by mean shift, SR + SC and FDCM_SSR has good regional consistency. However, the segmentation with $(h_s, h_r, W) = (9, 12, 250)$ shown in Fig. 3(e) contains eight classes. The accuracies of segmentation corresponding to Fig. 3 are shown in Table VI, where the accuracy of FDCM_SSR is the highest.

Fig. 4 shows the histogram of segmentation accuracies of different algorithms in Figs. 2(b) and 3(b) for viewing purposes. It can be seen from Table VI and Fig. 4 that the segmentation of the proposed FDCM_SSR algorithm outperforms other methods, except that the SA of FDCM_SSR in Fig. 2(b) is slightly below that of mean shift. However, the difference between the SA of FDCM_SSR and mean shift in Fig. 2(b) is only in (e^{-4}, e^{-3}), and the SA of FDCM_SSR in Fig. 3(b) is far higher than that of mean shift.

In addition, the runtime of different algorithms is shown in Table VI. Similarly with Fig. 4, the histogram of the running time of different methods in Fig. 2(b) and 3(b) is shown in Fig. 5. Except mean shift, other methods are performed in MATLAB platform. The mean shift algorithm runs in Visual Studio, code of which is available at http://coewww.rutgers.edu/riul/research/code/ EDISON/index.html. It is generally known that the running speed of same algorithm in visual studio is quicker than MATLAB. Seen from Fig. 5, although the runtime of FDCM_SSR is longer than most of clustering methods, it obviously runs quicker than SR + SC and NW_FCM.

C. Natural Images

To further compare the performances of other state-of-the-art clustering methods and the proposed FDCM_SSR algorithm, the segmentation results of these methods on two realistic natural images are shown in Figs. 6 and 7. However, as the ground truths corresponding to the real natural images are absent, the segmentation results on natural images are evaluated by visual inspection and an index I [52]. Thus, the best visual result of each algorithm is selected from 20 runs of it. Moreover, the related parameters are set according to Table V. In addition, the color histogram and mean value of every superpixel are extracted as feature in SR + SC, NW_FCM, and FDCM_SSR.

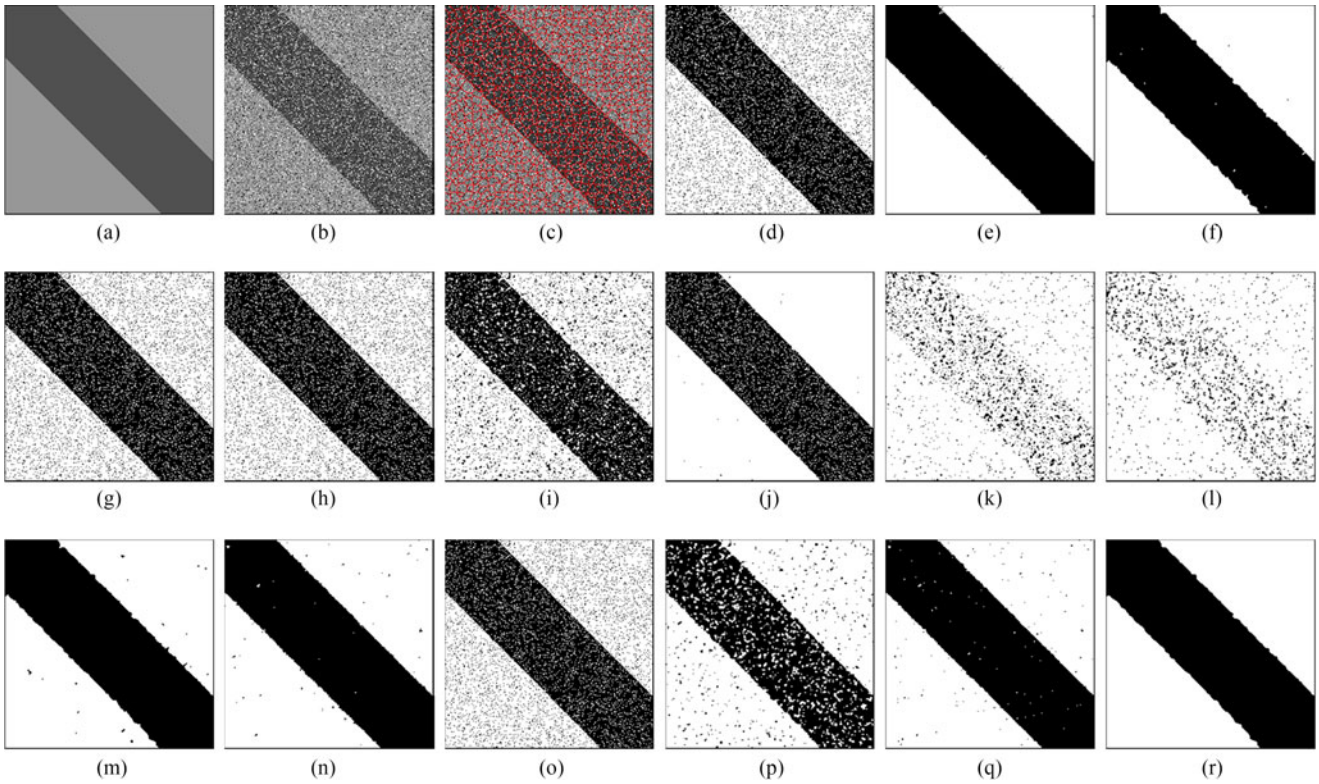

Fig. 2. Segmentation results on two-class synthetic image. (a) Original image, (b) the synthetic image with salt and pepper noise 20%, (c) superpixels obtained by Turbopixels algorithm, (d) K-means result, (e) mean shift result, (f) SR + SC result, (g) FCM result, (h) PFCM result, (i) FCM_S1 result, (j) FCM_S2 result, (k) EnFCM result, (l) FGFCM result, (m) NW_FCM result, (n) FLICM result, (o) GIFP_FCM result, (p) KGFCM_S1 result, (q) KGFCM_S2 result, and (r) FDCM_SSR result.

TABLE VI
SEGMENTATION ACCURACIES AND RUNTIME (SECOND) OF DIFFERENT ALGORITHMS ON TWO ARTIFICIAL IMAGES

		K-means	Mean shift	SR + SC	FCM	PFCM	FCM_S1	FCM_S2	En-FCM	FG-FCM	NW_FCM	FLICM	GIFP_FCM	KGFCM_S1	KGFCM_S2	FDCM_SSR
Fig. 2(b)	SA	0.9013	**0.9968**	0.9952	0.9013	0.9013	0.8779	0.9541	0.6218	0.6122	0.9944	0.9954	0.9013	0.9168	0.9933	0.996
	time	0.0700	4.8400	179.504	0.3008	0.1644	0.4296	0.4319	0.1838	0.9494	48.5706	2.6628	0.7498	6.2637	1.2963	48.4236
Fig. 3(b)	SA	0.5970	0.7304	0.9363	0.5848	0.5848	0.8624	0.8307	0.5629	0.5830	0.8303	0.9333	0.5780	0.9301	0.8571	**0.9406**
	time	0.1501	5.2900	201.1982	0.9969	0.9548	1.4015	1.6093	0.1766	0.9393	234.6238	8.0123	3.9277	14.4419	25.9049	70.5337

The first nature image is selected to test the robustness of different methods against change of light, as shown in Fig. 6(a). The parameters of mean shift are set to $(h_s, h_r, W) = (71, 8, 200)$.

Obviously, Fig. 6(a) should be of two kinds: cow and grass. As seen in Fig. 6, the segmentation of FDCM_SSR shown in Fig. 6(p) is better than that of other algorithms, in that many pixels in the head and back of cattle of other results are misclassified.

Fig. 7(a) shows a natural scenery image, which includes four types: sky, mountain, lake, and grass. The segmentations of K-means, FCM, PFCM, FCM_S1, FLICM, GIFP_FCM, and KGFCM_S2, as shown in Fig. 7(b), (e), (f), (g), (l), (m), and (o), respectively, basically distinguish four categories, but many pixels in the regions of lake and grass are misclassified. Fig. 7(c) shows the segmentation of mean shift with $(h_s, h_r, W) = (33, 16, 200)$, which contains eight categories. The segmentations of SR + SC, FCM_S2, NW_FCM, and KGFCM_S1 shown, respectively, in Fig. 7(d), (h), (k), and (n) confuse mountain with grass. Fig. 7(i) and (j) obtained, respectively, by EnFCM and FGFCM are dreadful. Fig. 7(p) shows the segmentation of the proposed FDCM_SSR algorithm, which not only accurately separates four classes but also has good regional consistency.

Table VII shows the indexes I of segmentations of different clustering methods in Figs. 6(a) and 7(a). The higher the index I, the better the performance of the clustering algorithm. As the K-means and mean shift methods have not the degree of membership that is used to calculate the index I, they are computed by the clustering center of segmentation results and dataset. Moreover, the SR + SC algorithm has not the clustering center and degree of membership, so it is not contained in Table VII. In addition, since FDCM_SSR has two types of

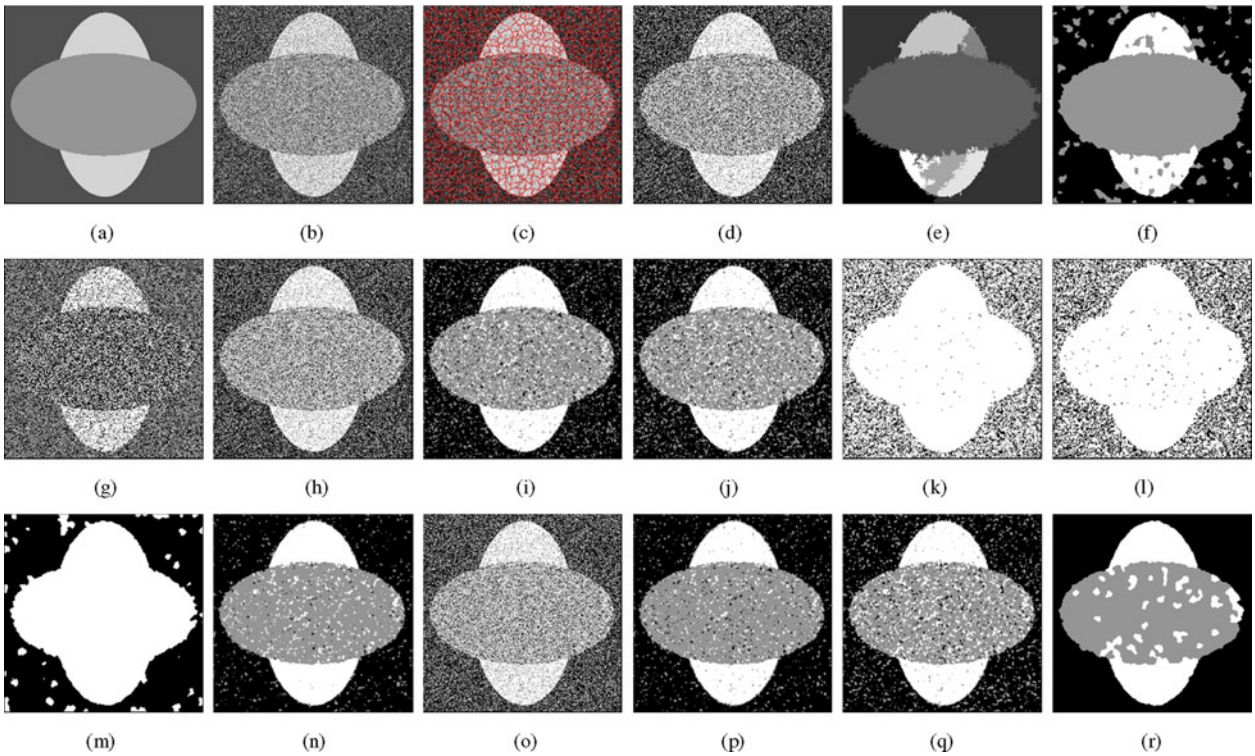

Fig. 3. Segmentation results on three-class synthetic image. (a) Original image, (b) the synthetic image with Gaussian noise 20%, (c) superpixels obtained by Turbopixels algorithm, (d) K-means result, (e) mean shift result, (f) SR + SC result, (g) FCM result, (h) PFCM result, (i) FCM_S1 result, (j) FCM_S2 result, (k) EnFCM result, (l) FGFCM result, (m) NW_FCM result, (n) FLICM result, (o) GIFP_FCM result, (p) KGFCM_S1 result, (q) KGFCM_S2 result, and (r) FDCM_SSR result.

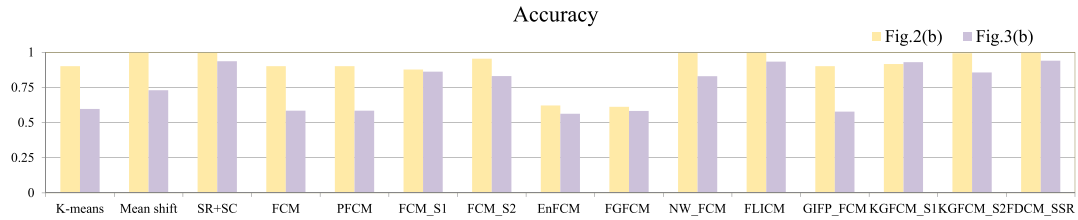

Fig. 4. Accuracy rates of the segmentations obtained by different methods for two noise images.

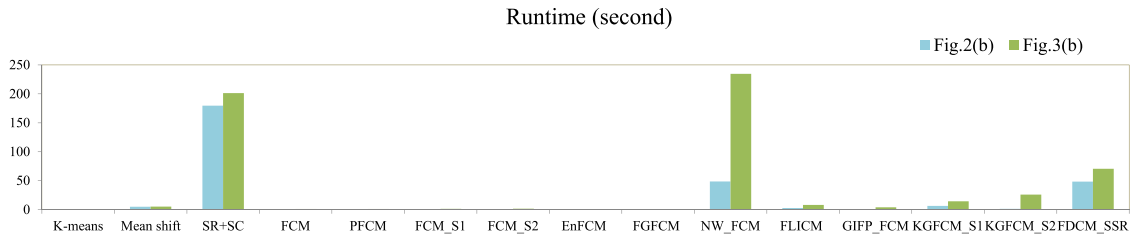

Fig. 5. Runtime (s) of different methods for two noise images.

clustering centers, the mean of two indexes I obtained by two types of clustering centers is shown in Table VII. As you can see from Table VII, the indexes I of FDCM_SSR are the highest.

D. SAR Images

To test the applicability of the proposed FDCM_SSR algorithm, the experiments on two real SAR images are implemented. The relevant environmental and parameter settings are the same with the above-mentioned section. Similarly, the visual inspection and an index I are used to evaluate the segmentations of different algorithms on SAR images, as the ground truths of SAR images are generally absent. Furthermore, the gray histogram and the local gradient ratio pattern histogram [53] of each superpixel are extracted as feature in running SR + SC, NW_FCM, and FDCM_SSR methods on the SAR images.

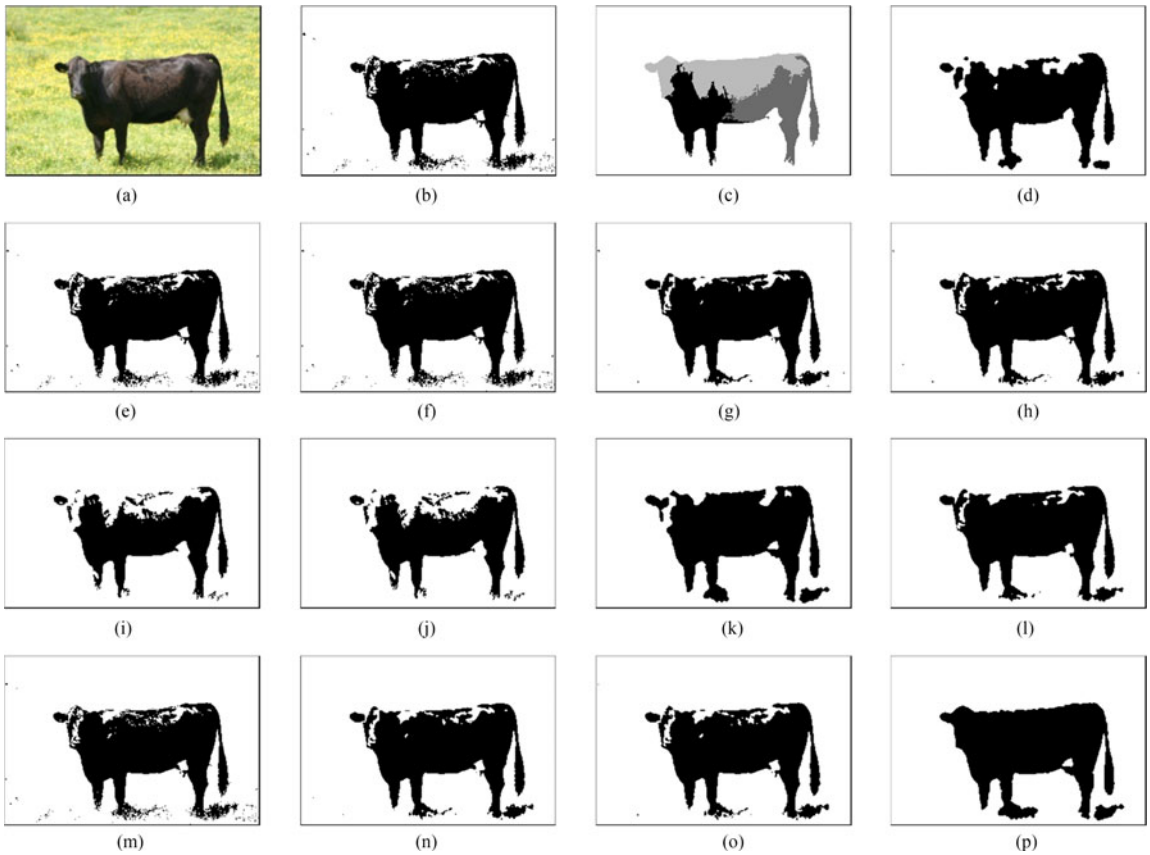

Fig. 6. Segmentation results on mountain image (213×320 pixels). (a) Original image, (b) K-means result, (c) mean shift result, (d) SR + SC result, (e) FCM result, (f) PFCM result, (g) FCM_S1 result, (h) FCM_S2 result, (i) EnFCM result, (j) FGFCM result, (k) NW_FCM result, (l) FLICM result, (m) GIFP_FCM result, (n) KGFCM_S1 result, (o) KGFCM_S2 result, and (p) FDCM_SSR result.

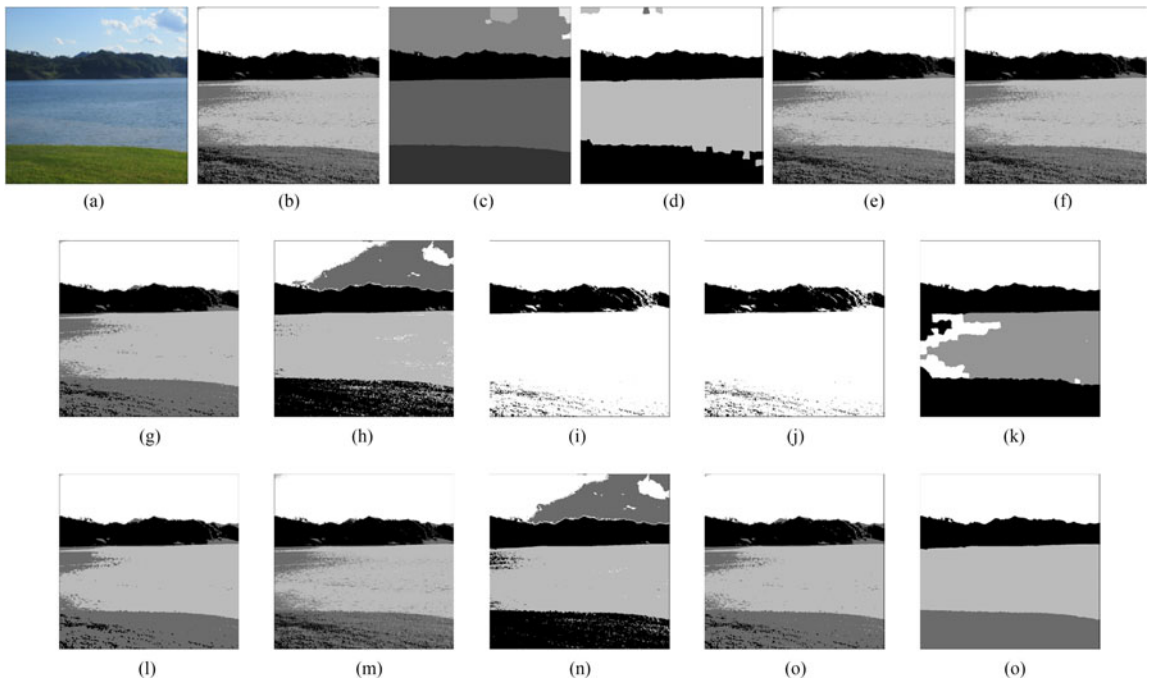

Fig. 7. Segmentation results on animal image (690×710 pixels). (a) Original image, (b) K-means result, (c) Mean shift result, (d) SR + SC result, (e) FCM result, (f) PFCM result, (g) FCM_S1 result, (h) FCM_S2 result, (i) EnFCM result, (j) FGFCM result, (k) NW_FCM result, (l) FLICM result, (m) GIFP_FCM result, (n) KGFCM_S1 result, (o) KGFCM_S2 result, and (p) FDCM_SSR result.

TABLE VII
CLUSTERING INDEX $I(e^{-5})$ OF DIFFERENT ALGORITHMS FOR TWO NATURAL IMAGES

	K-means	Mean shift	FCM	PFCM	FCM_S1	FCM_S2	En-FCM	FG-FCM	NW_FCM	FLICM	GIFP_FCM	KGFCM_S1	KGFCM_S2	FDCM_SSR
Fig. 6(a)	102.55	4.8032	5.0439	60.65	18.555	597.81	1820.4	131.72	2343.9	50.245	5.0438	411.85	470.05	**2380.3**
Fig. 7(a)	1.1706	0.8734	2.3151	5.23	2.0379	1.1067	3e–14	1096.1	1044.1	532.77	1.0051	11.639	1.3927	**1178.9**
Fig. 8(a)	1.9994	8.6702	6.5807	90.27	23.778	59.611	717.38	661.73	1712	352.77	6.5807	87.669	24.061	**2752.9**
Fig. 9(a)	1.5049	0.0134	2.0025	2.0025	5.6567	1.2738	3.1e–9	1.3914	245.38	9.0363	2.5131	19.734	2.4243	**360.09**

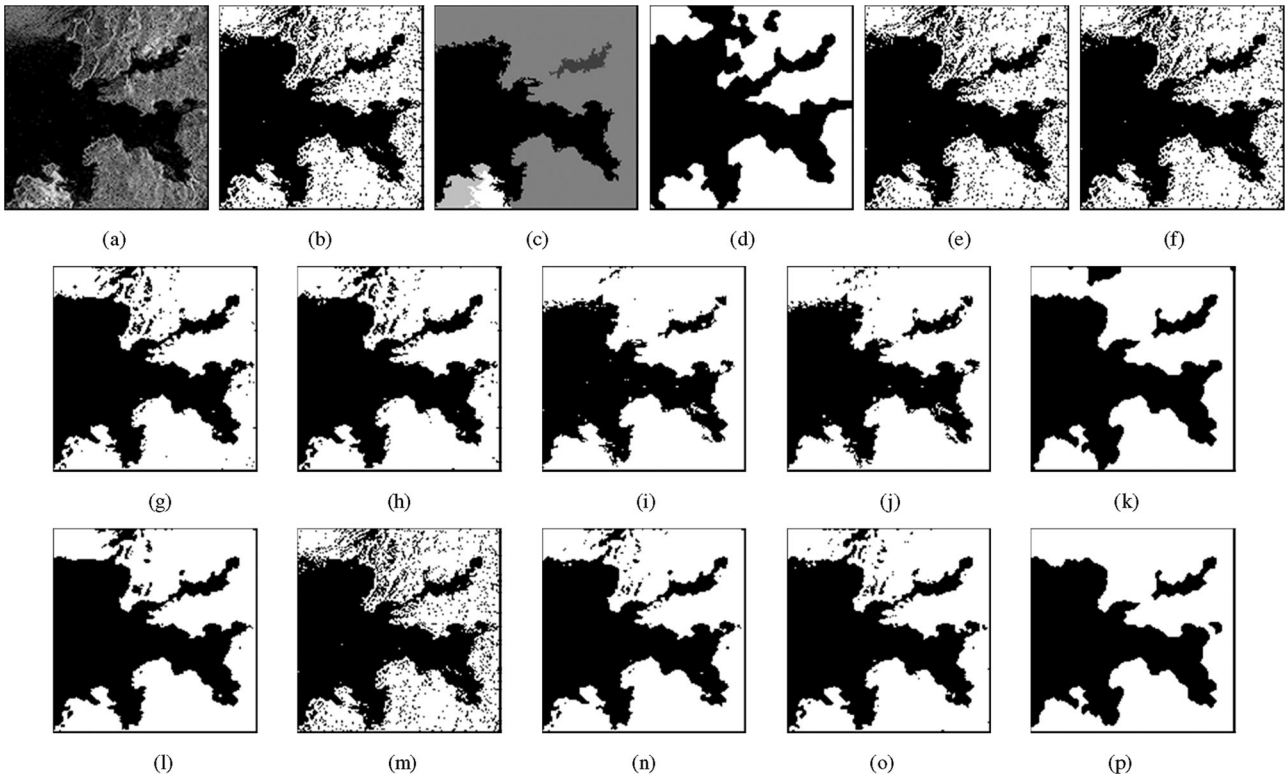

Fig. 8. Segmentation results on the SAR subimage of Aegean sea (150×150 pixels). (a) Original image, (b) K-means result, (c) mean shift result, (d) SR + SC result, (e) FCM result, (f) PFCM result, (g) FCM_S1 result, (h) FCM_S2 result, (i) EnFCM result, (j) FGFCM result, (k) NW_FCM result, (l) FLICM result, (m) GIFP_FCM result, (n) KGFCM_S1 result, (o) KGFCM_S2 result, and (p) FDCM_SSR result.

Fig. 8(a) shows a section of an SAR image over the Aegean sea of oil pollution near La Coruna in Northern Spain and provided by ERS-1 on 3 December 1992. This SAR image consists of two types of land cover: water (dark) and bare land (bright). Fig. 8(b), (e), (f), and (m) shows the segmentations of K-means, FCM, PFCM, and GIFP_FCM, respectively. In Fig. 8(b), (e), (f), and (m), the bare land region has many spots. Though the consistency in the bare land regions is good in the segmentation of mean shift (with $(h_s, h_r, W) = (33, 8, 200)$), SR + SC, and NW_FCM, respectively, shown in Fig. 8(c), (d), and (k), some pixels are misclassified of Fig. 8(c), (d), and (k). The boundaries of Fig. 8(g), (h), (i), (j), (l), (n), and (o), which are, respectively, obtained by FCM_S1, FCM_S2, EnFCM, FGFCM, FLICM, KGFCM_S1, and KGFCM_S2, are fuzzy. Compared with other methods, Fig. 8(p) obtained by FDCM_SSR has clear boundaries and favorable regional consistency. The indexes I of segmentation results of different algorithms in Fig. 8(a) are shown in Table VII. It is easy to see from Table VII that the index I of FDCM_SSR is greater than that of other methods.

Another experiment is carried out on a Ku-band SAR image with 3-m spatial resolution of the China Lake Airport, California, as shown in Fig. 9(a). The image includes three types of ground objects of runway (dark), vacancy (gray), and airport buildings (bright). Fig. 9(b), (e), (f), (g), (h), (l), (m), (n), and (o) are seriously spotty in vacancy areas, which, respectively, shows the segmentation results of K-means, FCM, PFCM, FCM_S1, FCM_S2, FLICM, GIFP_FCM, KGFCM_S1, and KGFCM_S2. Although the segmentation of mean shift with $(h_s, h_r, W) = (17, 10, 10)$ shown in Fig. 9(c) has good visual effect, it actually contains 103 categories. Many pixels are misclassified in the segmentation of SR + SC, as shown in Fig. 9(d). The vacancy and airport buildings regions of Fig. 9(i), (j), and (k) obtained, respectively, by EnFCM, FGFCM, and NW_FCM are confused. By contrast, the segmentation of FDCM_SSR

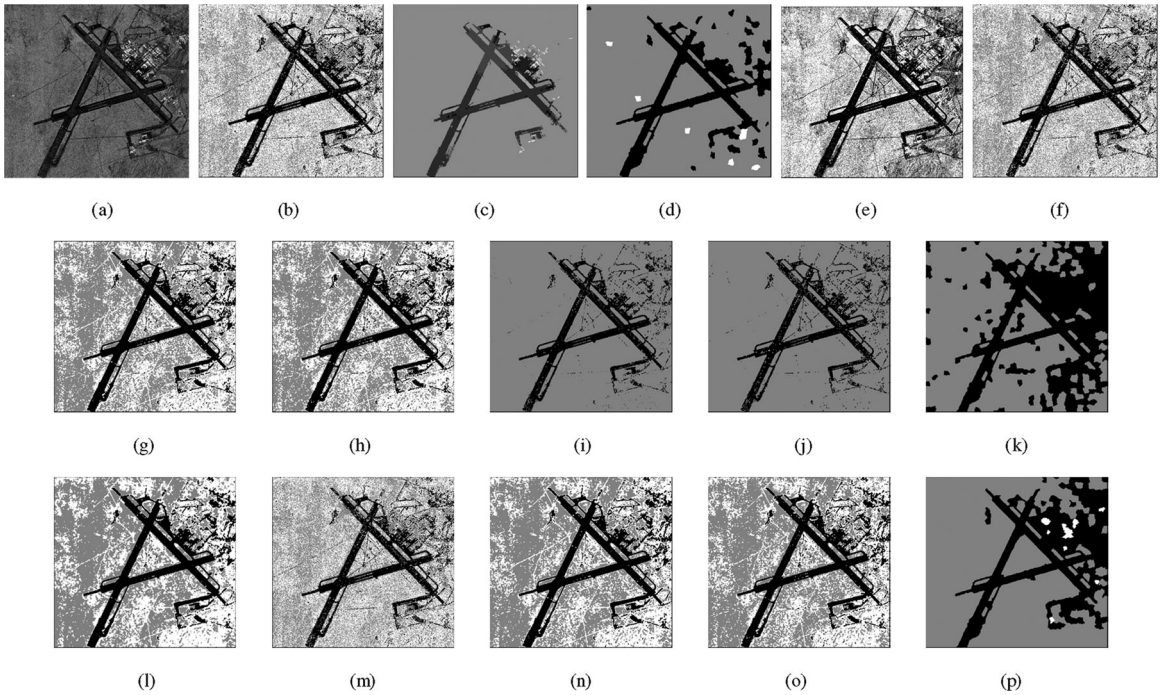

Fig. 9. Segmentation results on the SAR image of China lake airport (446×475 pixels). (a) Original image, (b) K-means result, (c) mean shift result, (d) SR + SC result, (e) FCM result, (f) PFCM result, (g) FCM_S1 result, (h) FCM_S2 result, (i) EnFCM result, (j) FGFCM result, (k) NW_FCM result, (l) FLICM result, (m) GIFP_FCM result, (n) KGFCM_S1 result, (o) KGFCM_S2 result, and (p) FDCM_SSR result.

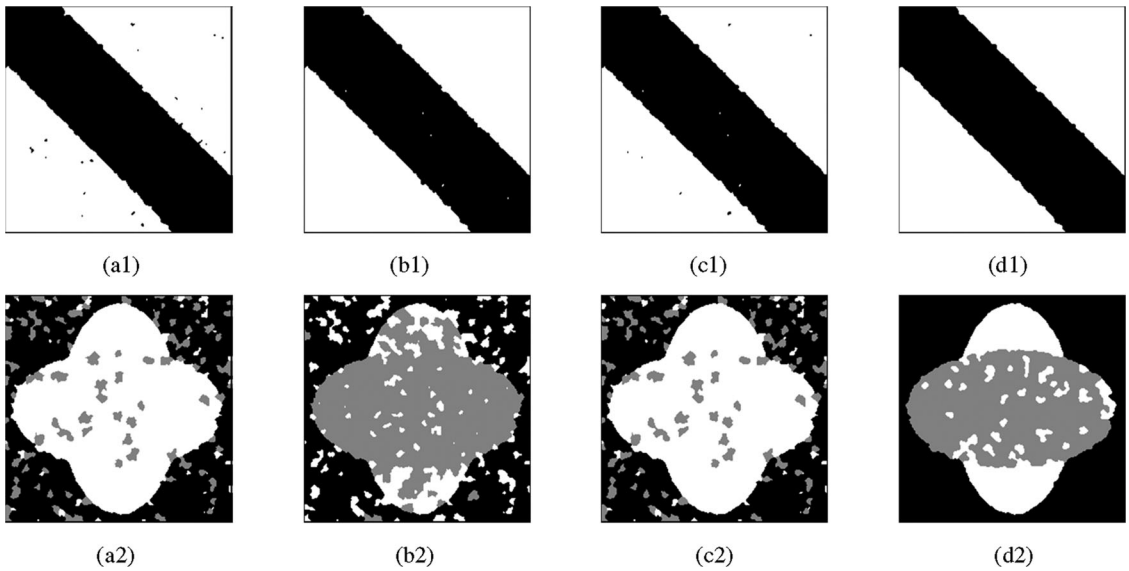

Fig. 10. (a1) Segmentation (SA: 0.9941) of FCM with basic feature for Fig. 2(b). (b1) Segmentation (SA: 0.9954) of FCM with discriminant feature in Fig. 2(b). (c1) Segmentation (SA: 0.9941) of FCM with concatenation feature for Fig. 2(b). (d1) Segmentation (SA: 0.996) of FDCM_SSR in Fig. 2(b). (a2) Segmentation (SA: 0.7508) of FCM with basic feature for Fig. 3(b). (b2) Segmentation (SA: 0.8454) of FCM with discriminant feature in Fig. 3(b). (c2) Segmentation (SA: 0.7468) of FCM with concatenation feature for Fig. 3(b). (d2) Segmentation (SA: 0.9406) of FDCM_SSR in Fig. 3(b).

shown in Fig. 9(p) has good regional consistency in runway and vacancy regions. Table VII shows the indexes I of segmentations corresponding to Fig. 9. Obviously, seen from the contrast of the indexes I for Fig. 9(a) in Table VII, the segmentation of FDCM_SSR is better than that of other methods.

E. Discussion

First, a problem whether different types of features should be concatenated is discussed. Take the two artificial images Figs. 2(b) and 3(b) as examples. Fig. 10 shows the segmentation results of FCM with basic feature, discriminant feature, concatenation of basic feature, and discriminant feature of superpixels and FDCM_SSR algorithm in Figs. 2(b) and 3(b). As can be seen from Fig. 10, the segmentations of FCM with concatenation feature shown in Fig. 10(c1) and (c2) are worse than those of FCM with a kind of feature, which means that the concatenation of different kinds of features does not enable the segmentation result to be best. Additionally, the

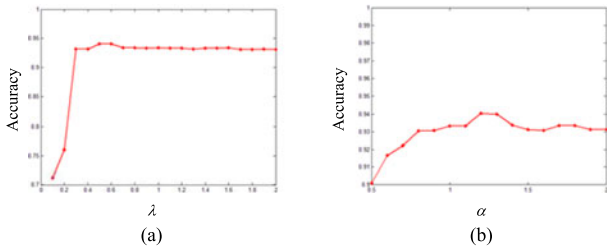

Fig. 11. Influence of different parameters on the FDCM_SSR segmentation results. (a) The influence of the parameter λ ($\alpha = 1.2$). (b) The influence of the parameter α ($\lambda = 0.5$).

segmentations of FCM based on discriminant feature are better than those based on basic feature, as shown in Fig. 10(a1), (b1), (a2), and (b2). It indicates that the SR technology is effective. Furthermore, the segmentations of FDCM_SSR shown in Fig. 10(d1) and (d2) are best, which demonstrates that we effectively make use of different types of features in the proposed algorithm.

Second, the segmentation result of the proposed FDCM_SSR algorithm may be affected by two adjusting parameters: λ in (2) and α in (5). Taking the three-class synthetic image Fig. 3(b) as an example, the influence of different parameters on the segmentation results is shown in Fig. 11. It is can be seen from Fig. 11 that two curves are basically stable after a small portion of fluctuations. When $\lambda \geq 0.5$, the segmentation accuracies are high, as shown in Fig. 11(a), which reflects that the fidelity term in (2) plays an important role in the sparse self-representation model. Moreover, the accuracy rates obviously increase when $\alpha > 1$ in Fig. 11(b). It shows that the segmentation result is improved when the effect of the discriminant feature is greater than that of the basic feature. So the discriminant feature is helpful to the clustering. Additionally, the accuracies slowly decrease when λ and α are too large, as shown in Fig. 11. That is because the impact of the second terms in (2) and (5) cannot be hugely magnified and the first items of (2) and (5) also play a certain role.

V. Conclusion

In this paper, a novel FDCM_SSR was introduced. The biggest difference between the proposed algorithm and other fuzzy clustering methods is that FDCM_SSR can simultaneously cluster two datasets and other fuzzy clustering methods only can deal with one set of data. The two datasets can have different dimensions and belong to different spaces. In fact, each set of data is a kind of feature descriptor of sample set. Two types of feature sets can describe the sample set more fully than a kind of feature set. They are named as basic feature set and discriminant feature set in this paper. The basic feature set is the basic physical properties of sample set, which is usually used to cluster in existing fuzzy clustering algorithms. By utilizing the sparse self-representation model, the discriminant feature set is learned from the basic feature set. The sparse self-representation method not only has noise-robustness and data-adaptiveness, but also reflects the global similarity among samples. Therefore, the discriminant feature set contributes to obtain good clustering result. Furthermore, most of the improved FCM algorithms are only used to segment image, while FDCM_SSR is applied in data clustering and image segmentation. Owing to the usage of the discriminant feature including the global structure, experimental results of FDCM_SSR on different datasets and images are more accurate and efficient than those of other fuzzy clustering methods.

References

[1] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition I," *IEEE Trans. Syst., Man, Cybern. Soc.*, vol. 29, no. 6, pp. 778–785, Dec. 1999.
[2] 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.
[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] L. A. Zadeh, "Fuzzy sets," *Inf. Control.*, vol. 8, pp. 338–353, 1965.
[5] I. Gitman and M. D. Levine, "An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique," *IEEE Trans. Comput.*, vol. C-19, no. 7, pp. 583–593, Jul. 1970.
[6] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98–109, Jan. 2014.
[7] T. Banerjee, J. M. Keller, M. Skubic, and E. Stone, "Day or night activity recognition from video using fuzzy clustering techniques," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 3, pp. 483–493, Jun. 2014.
[8] I-J. Chiang, C.-H. Liu, Y.-H. Tsai, and A. Kumar, "Discovering latent semantics in web documents using fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2122–2134, Dec. 2015.
[9] T. Nguyen and J. Wu, "Online feature selection based on fuzzy clustering and its applications," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 6, pp. 1294–1306, Dec. 2016.
[10] L. Hu and K. C. C. Chan, "Fuzzy clustering in a complex network based on content relevance and link structures," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 2, pp. 456–470, Apr. 2016.
[11] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984.
[12] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.
[13] R. Krishnapuram and J. M. Keller, "The possibilistic c-means algorithm: Insights and recommendations," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 3, pp. 385–393, Aug. 1996.
[14] N. R. Pal, K. Pal, and J. C. Bezdek, "A mixed c-means clustering model," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, vol. 1, Jul. 1997, pp. 11–21.
[15] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, Aug. 2005.
[16] 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, Mar. 2002.
[17] 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.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.
[18] L. Szilagyi, Z. Benyo, S. Szilagyii, and H. Adam, "MR brain image segmentation using an enhanced fuzzy c-means algorithm," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2003, pp. 17–21.
[19] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recog.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.
[20] C. H. Li, W. C. Huang, B. Kuo, and C.-C. Hung, "A novel fuzzy weighted c-means method for image classification," *Int. J. Fuzzy Syst.*, vol. 10, no. 3, pp. 168–173, Sep. 2008.
[21] C.-C. Hung, S. Kulkarni, and B. Kuo, "A new weighted fuzzy c-means clustering algorithm for remotely sensed image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 543–553, Jun. 2011.
[22] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328–1337, May 2010.

[23] R. J. Hathaway, J. C. Bezdek, and Y. Hu, "Generalized fuzzy c-means clustering strategies using Lp norm distances," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 576–582, Oct. 2000.
[24] L. Zhu, F.-L. Chung, and S. Wang, "Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions," *IEEE Trans. Syst., Man, Cybern.*, vol. 39, no. 3, pp. 578–591, Jun. 2009.
[25] H.-C. Huang, Y. Y. Chuang, and C. Chen, "Multiple kernel fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 1, pp. 120–134, Feb. 2012.
[26] X. Yang, G. Zhang, J. Lu, and M. Jun, "A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 105–115, Feb. 2014.
[27] F. Zhao, L. Jiao, and H. Liu, "Kernel generalized fuzzy c-means clustering with spatial information for image segmentation," *Digital Signal Process.*, vol. 23, pp. 184–199, 2013.
[28] M. Aharon, M. Elad, and B. Alfred, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[29] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
[30] M. Luo, F. Sun, and H. Liu, "Hierarchical structured sparse representation for T–S fuzzy systems identification," *IEEE Trans. Fuzzy Syst.*, vol. 21, no. 6, pp. 1032–1043, Dec. 2013.
[31] M. Luo, F. Sun, and H. Liu, "Joint block structure sparse representation for multi-input–multi-output T–S fuzzy system identification," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1387–1400, Dec. 2014.
[32] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–35, Jan. 2009.
[33] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 2368–2378, Apr. 2014.
[34] N. Qi, Y. Shi, X. Sun, W. Ding, and B. Yin, "Single image super-resolution via 2D sparse representation," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2015, pp. 1–6.
[35] J. Gu, L. Jiao, S. Yang, F. Liu, B. Hou, and Z. Zhao, "A multi-kernel joint sparse graph for SAR image segmentation," *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, vol. 9, no. 3, pp. 1265–1285, Mar. 2016.
[36] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[37] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2009, pp. 2790–2797.
[38] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
[39] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," in *Proc. Int. Conf. Comput. Vision*, 2011, pp. 2439–2446.
[40] J. Gu, L. Jiao, S. Yang, and J. Zhao, "Sparse learning based fuzzy c-means clustering," *Knowl.-Based Syst.*, vol. 119, pp. 113–125, 2016.
[41] J. Cheng, H. Liu, T. Liu, F. Wang, and H. Li, "Remote sensing image fusion via wavelet transform and sparse representation," *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 158–173, 2015.
[42] X. Zhang, D. S. Pham, S. Venkatesh, W. Liu, and D. Phung, "Mixed-norm sparse representation for multi view face recognition," *Pattern Recog.*, vol. 48, no. 9, pp. 2935–2946, 2015.
[43] Q. Qiu, V. M. Patel, and C. Rama, "Information-theoretic dictionary learning for image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2173–2184, Nov. 2014.
[44] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, pp. 17–40, 1976.
[45] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *Soc. Ind. Appl. Math. J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Dec. 2009.
[46] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.
[47] A. Salah, M. H. Mahoor, X. Zhang, and R. M. Voyles, "Human activity recognition using multi-features and multiple kernel learning," *Pattern Recog.*, vol. 47, no. 5, pp. 1800–1812, 2014.
[48] T. F. Covoes, E. R. Hruschka, and J. Ghosh, "Adjusted rand index, normalized mutual information," *Intell. Data Anal.*, vol. 17, no. 3, pp. 485–505, 2013.
[49] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, 1975.
[50] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[51] L. Alex, A. Stere, K. N. K., D. J. Fleet, J. D. Sven, and K. S., "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.
[52] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.
[53] T. Tang, D. Xiang, H. Liu, and Y. Su, "A new local feature extraction in SAR image," in *Proc. Asian-Pac. Conf. Synth. Aperture Radar*, 2013, pp. 377–379.

Jing Gu received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 2007 and 2010, respectively. She is currently working toward the Ph.D. degree in the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an, China.

Her research interests include image processing, machine learning, and pattern recognition.

Licheng Jiao (SM'89) received the B.S. degree from the Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from the Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

Since 1992, he has been a Professor in the School of Electronic Engineering, Xidian University, Xi'an, China, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China. He is in charge of about 40 important scientific research projects and has published more than 20 monographs and a hundred papers in international journals and conferences. His research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is a member of the IEEE Xi'an Section Execution Committee, the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, a Councilor of the Chinese Institute of Electronics, a committee member of the Chinese Committee of Neural Networks, and an Expert of the Academic Degrees Committee of the State Council.

Shuyuan Yang (M'07) received the M.S. and Ph.D. degrees in circuits and systems from Xidian University, Xi'an, China, in 2003 and 2005, respectively.

She is currently a Full Professor in the Department of Electrical Engineering, Xidian University. Her current research interests include machine learning, computer vision, and compressive sampling.

Fang Liu (M'07–SM'07) received the B.S. degree in computer science and technology from the Xi'an Jiaotong University, Xi'an, China, in 1984, and the M.S. degree in computer science and technology from Xidian University, Xi'an, China, in 1995.

Currently, she is a Professor in the School of Computer Science, Xidian University. Her research interests include synthetic aperture radar image processing, multiscale geometry analysis, optimization problems, and data mining.