Fuzzy C-Means Clustering With Local Information and Kernel Metric for Image Segmentation

Maoguo Gong, *Member, IEEE*, Yan Liang, Jiao Shi, Wenping Ma, and Jingjing Ma

***Abstract*— In this paper, we present an improved fuzzy C-means (FCM) algorithm for image segmentation by introducing a tradeoff weighted fuzzy factor and a kernel metric. The tradeoff weighted fuzzy factor depends on the space distance of all neighboring pixels and their gray-level difference simultaneously. By using this factor, the new algorithm can accurately estimate the damping extent of neighboring pixels. In order to further enhance its robustness to noise and outliers, we introduce a kernel distance measure to its objective function. The new algorithm adaptively determines the kernel parameter by using a fast bandwidth selection rule based on the distance variance of all data points in the collection. Furthermore, the tradeoff weighted fuzzy factor and the kernel distance measure are both parameter free. Experimental results on synthetic and real images show that the new algorithm is effective and efficient, and is relatively independent of this type of noise.**

***Index Terms*— Fuzzy clustering, gray-level constraint, image segmentation, kernel metric, spatial constraint.**

I. Introduction

IMAGE segmentation is one of the key techniques in image understanding and computer vision. The task of image segmentation is to divide an image into a number of non-overlapping regions, which have same characteristics such as gray level, color, tone, texture, etc. A lot of clustering-based methods have been proposed for image segmentation [1]–[5]. Among the clustering methods, one of the most popular methods for image segmentation is fuzzy clustering, which can retain more image information than hard clustering in some cases.

Fuzzy c-means (FCM) algorithm is one of the most widely used fuzzy clustering algorithms in image segmentation. FCM algorithm was first introduced by Dunn [6] and later extended by Bezdek [7]. Although the conventional FCM algorithm works well on most noise-free images, it fails to segment images corrupted by noise, outliers and other imaging artifacts. Its non-robust results are mainly because of ignoring spatial contextual information in image and the use of non-robust Euclidean distance.

Manuscript received December 28, 2011; revised August 20, 2012; accepted September 6, 2012. Date of publication September 18, 2012; date of current version January 10, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 61273317, Grant 61202176, and Grant 61203303, the National Top Youth Talents Support Program of China, and the Fundamental Research Fund for the Central Universities under Grant K50510020001. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kenneth Lam.

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China (e-mail: gong@ieee.org; mggong@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/TIP.2012.2219547

To deal with the first problem, many improved FCM algorithms have been proposed by incorporating local spatial information into original FCM objective function [8]–[12]. Ahmed *et al.* [8] proposed FCM_S, which modified the objective function of FCM by introducing the spatial neighborhood term. One drawback of FCM_S is that the spatial neighborhood term is computed in each iteration step, which is very time-consuming. To reduce the computational complexity of FCM_S, Chen and Zhang [9] proposed two variants, FCM_S1 and FCM_S2, which replace the neighborhood term of FCM_S by introducing the extra mean-filtered image and median-filtered image, respectively. The mean-filtered image and median-filtered image can be computed in advance, so the computational costs can be reduced. To speed up the image segmentation process, Szilagyi *et al.* [10] proposed the enhanced FCM (EnFCM), which form a linearly-weighted sum image from both the local neighborhood average gray level of each pixel and original image, and then clustering is performed on the basis of the gray level histogram of summed image. Thus, the computational time of EnFCM is very small. Cai *et al.* [11] proposed the fast generalized FCM (FGFCM) algorithm. This method introduces a local similarity measure that combines both spatial and gray level information to form a non-linearly weighted sum image. Clustering is performed on the basis of the gray level histogram of the summed image. Thus, its computational time, similar to EnFCM, is also very small. However, these algorithms do not directly apply on the original image. They need some parameters a (or λ) to control the trade-off between robustness to noise and effectiveness of preserving the details. The selection of these parameters is not an easy task, and has to be made by experience or by using the trial-and-error method.

To overcome the above mentioned problems, Stelios *et al.* [1] presents a novel robust fuzzy local information c-means clustering algorithm (FLICM), which is free of any parameter selection, as well as promoting the image segmentation performance. In FLICM, a novel fuzzy factor is defined to replace the parameter a used in above algorithms and its variants. More recently, we [12] proposed a variant of FLICM algorithm (RFLICM), which adopts the local coefficient of variation to replace the spatial distance as a local similarity measure. Furthermore, it presents a more robust result. Although RFLICM algorithm can exploit more local context information to estimate the relationship of pixels in neighbors since the local coefficient of variation, it is still unreasonable to ignore the influence of spatial constraint on the relationship between central pixel and pixels in neighbors.

Hence, in this study, one of our motivations is to design a trade-off weighted fuzzy factor for adaptively controlling the local spatial relationship. This factor depends on space distance of all neighbor pixels and their gray level difference simultaneously.

In order to further improve the performance of FLICM in restraining noise, another novelty in this study is introducing the kernel distance measure to its objective function. In recent years, kernel methods have received an enormous amount of attention in machine learning community. Its main idea is to transform complex nonlinear problems in original low-dimensional feature space to the problems easily solved in the transformed space. Typical examples are support vector machines (SVM) [13], [14] kernel principle component analysis (KPCA) [15] and kernel perceptron algorithm [16]. Particularly, the clustering algorithms based on the kernel methods have been shown to be robust to the outliers or noises of the dataset [17]. So, the clustering algorithms based on kernel methods have been applied to many fields of image segmentation [18]–[21]. Because of this advantage, Zhang *et al*. [21] introduced a new kernel-induced distance measure for the original data space into the objective function of FCM (KFCM) to replace the conventional measures. Meanwhile, a penalty term takes into account the influence the neighboring pixels on the central pixel. It is shown to be more robust to noise and the outlier in image segmentation than the algorithms without the kernel substitution. However, as mentioned above, this penalty term must be computed in each iteration step, which is very time-consuming. Therefore, Chen et al. [9] also proposed two variants of KFCM, which replace this term using the mean-filtered (KFCM_S1) or median-filtered (KFCM_S2) image to reduce the computational cost. In general, the Gaussian RBF (GRBF) kernel [9] is adopted in the kernel function for its facility. But the parameter σ in GRBF has a great influence on the performance of the algorithm. Therefore, we use a fast bandwidth selection rule, which can adaptively compute the parameter σ.

The rest of this paper is organized as follows. In the next section, the main ideas of the proposed approach and our motivation will be introduced. Section III will describe the proposed algorithm in details. In Section IV, experimental results on synthetic images, medical images and natural images will be described. Conclusions will be drawn in Section V.

II. Motivation

Let us consider an image Ω be composed of N points, each point $i \in \Omega$ having a given value (gray-level) x_i. Let us suppose that this image has to be segmented into c ($c \geq 2$) classes. In the FCM approach, the segmentation process of a gray-level image can be defined as the minimization of an energy function.

FLICM [1] introduces a novel fuzzy factor G_{ki} as a fuzzy local similarity measure in its objective function, which is aimed at guaranteeing noise insensitiveness and image detail preservation. Its objective function for partitioning a dataset $\{x_i\}_{i=1}^N$ (in the gray-level space) into c clusters is defined in terms of

$$J_m = \sum_{i=1}^{N}\sum_{k=1}^{c}\left[u_{ki}^m \|x_i - v_k\|^2 + G_{ki}\right] \tag{1}$$

where $\|\bullet\|$ denotes the Euclidean norm, $\{v_k\}_{k=1}^c$ stands for the centers or prototypes of the clusters. The parameter m is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. This fuzzification factor m in the proposed algorithm and the above algorithms plays the same role as FCM which has been discussed in [7]. Here, we set $m = 2$ for the following experiments. The array $U = \{u_{ki}\}$ represents a membership matrix satisfying

$$U \in \left\{ u_{ki} \in [0,1] \left| \sum_{k=1}^{c} u_{ki} = 1, \forall i \text{ and } 0 < \sum_{i=1}^{N} u_{ki} < N, \forall k \right. \right\} \tag{2}$$

while the fuzzy factor G_{ki} is defined mathematically as follow

$$G_{ki} = \sum_{j \in N_i} \frac{1}{d_{ij}+1}\left(1 - u_{kj}\right)^m \left\|x_j - v_k\right\|^2 \tag{3}$$

where the ith pixel x_i is the center of the local window and the jth pixel x_j represents the neighboring pixels falling into the window around x_i, d_{ij} is the spatial Euclidean distance between pixels i and j. N_i stands for the set of neighbors in a window around x_i. v_k represents the prototype of the center of cluster k, and u_{kj} represents the fuzzy membership of gray value x_j with respect to the kth cluster.

Then, the objective function can get the minimum by updating the membership $\{u_{ki}\}$ and the cluster center $\{v_k\}_{k=1}^c$ as follows

$$u_{ki} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_i - v_k\|^2 + G_{ki}}{\|x_i - v_j\|^2 + G_{ji}}\right)^{1/(m-1)}} \tag{4}$$

$$v_k = \frac{\sum_{i=1}^{N} u_{ki}^m x_i}{\sum_{i=1}^{N} u_{ki}^m}. \tag{5}$$

When the algorithm has converged ($\max\left|V^{(b)} - V^{(b+1)}\right| < \varepsilon$, where $V = [v_1, v_2, ...v_c]$ are the vectors of the cluster prototypes), a defuzzification process takes place in order to convert the fuzzy partition matrix U to a crisp partition. In a general way, the maximum membership procedure method is adopted. This procedure assigns the pixel i to the class C_k with the highest membership

$$C_k = \arg\left\{\max\left\{u_{ki}\right\}\right\}, \qquad (k = 1,2, ..., c). \tag{6}$$

It is used to convert the fuzzy image achieved by the proposed algorithm to the crisp segmented image.

A. Motivation of Introducing the Trade-Off Weighted Fuzzy Factor

In FLICM [1], with the application of the fuzzy factor G_{ki}, the corresponding membership values of the non-noisy pixels, as well as the noisy pixels that falling into the local

A 90	22	13
32	20	35
28	B 120	27

(a)

87	99	116
90	20	67
110	88	75

(b)

0.414	0.5	0.414
0.5		0.5
0.414	**0.5**	0.414

(c)

Fig. 1. 3×3 window with noise and the damping extent of the neighbors. (a) Central pixel with no noise. (b) Central pixel is corrupted by noise. (c) Damping extent of the neighbors.

window will converge to a similar value, and thereby balance the membership values of the pixels that located in the window. Thus, FLICM becomes more robust to outliers.

As shown in Eq. (3), it can be seen that the factor G_{ki} is formulated without setting any artificial parameter that controls the trade-off between image noise and the image details. What's more, the influence of pixels within the local window in G_{ki} is exerted flexibly by using their spatial Euclidean distance from the central pixel. Therefore, G_{ki} can reflect the damping extent of the neighbors with the spatial distance from the central pixel. Such as the noisy pixels A and B in Fig. 1, the gray level difference between A and central pixel is less than pixel B, shown in Fig. 1 (a). But the damping extent of them is opposite. While the central pixel is noisy pixel, this situation as Fig. 1(b) shown, it fails to analyze the impact of each neighboring pixel. It is not conformity with the discrepancy between the pixels in the neighborhood. In [12], we proposed a variant of the factor G_{ki} by utilizing the local coefficient of variation to replace the spatial distance for overcoming above shortcomings and contribute to use more local context information. It is described as follows

$$G_{ki'} = \begin{cases} \sum\limits_{j \in N_i} \dfrac{1}{2+\min\left(\left(C_u^j / C_u\right)^2, \left(C_u / C_u^j\right)^2\right)} \left(1-u_{kj}\right)^m \\ \qquad \left\|x_j - v_k\right\|^2, \ \text{if } C_u^j \geq \overline{C_u} \\ \sum\limits_{j \in N_i} \dfrac{1}{2-\min\left(\left(C_u^j / C_u\right)^2, \left(C_u / C_u^j\right)^2\right)} \left(1-u_{kj}\right)^m \\ \qquad \left\|x_j - v_k\right\|^2, \ \text{if } C_u^j < \overline{C_u} \end{cases} \quad (7)$$

where C_u represents the local coefficient of variation of central pixel. C_u^j represents the jth local coefficient of variation in neighbors, $\overline{C_u}$ is the mean value of C_u^j that located in a local window.

Visually, G'_{ki} makes the relationship between neighboring pixels and the central pixel is relatively accordance with the gray level difference between them. But it is still unreasonable to ignore the effect of spatial distance constraint on the relationship between central pixel and pixels in the neighborhood when the size of window is enlarged. In addition, the damping extent of the neighbors can not be accurately calculated, as the same gray-level distribution and different spatial constraint. Hence, in this paper, we define a new trade-off weighted fuzzy factor to adaptively control the local neighbor relationship. This factor depends on space distance of all neighbor pixels and their gray level discrepancy simultaneously. In Section III.B, we will present this tradeoff weighted fuzzy factor in details.

B. Motivation of Using Non-Euclidean Distance

It can be seen that the measure used in the objective function of FLICM is still the Euclidean metric as in FCM. Although this metrical method is computationally simple, the use of Euclidean distance can lead to non-robust results on segmentation of image corrupted by noise, outliers, and other imaging artifacts. Therefore, some researchers adopted so-called robust distance measures, such as L_p norms ($0 < p < 1$) [22], [23] to replace the L_2 norm in the FCM objective function to reduce the effect of outliers on clustering results.

On the other hand, there is a trend in recent machine learning work to construct a nonlinear version of a linear algorithm using the 'kernel method', which aims at transforming the complex nonlinear problems in original low-dimensional feature space to the problems which can be easily solved in the transformed space. And this method can also be used in clustering, as shown in support vector clustering [14] and kernel (fuzzy) c-means algorithms [9].

A kernel in the feature space can be represented as the following function K

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (8)$$

where $\Phi(\bullet)$ refers to an implicit nonlinear map, $\langle \Phi(x), \Phi(y) \rangle$ denotes the inner product operation.

There are many kernel functions in literature [24]. Different kernels will induce different measure for the original space. Gaussian Radial basis function (GRBF) kernel is a commonly-used method. And its mathematical formulation presents as follow

$$K(x, y) = \exp\left(\frac{-\left(\sum\limits_{i=1}^{d} |x_i - y_i|^a\right)^b}{\sigma}\right) \quad (9)$$

where d is the dimension of vector x; σ is the kernel bandwidth, a self-custom parameter; $a \geq 0$; $1 \leq b \leq 2$. Obviously, $K(x, x) = 1$ for all x and the above RBF kernels. Note, that the parameter σ in Gaussian RBF has a very important effect on performance of the algorithms. However, the selection of an appropriate band-width value for a kernel-based clustering algorithm could be very troublesome since all the data points are unlabeled and their true classes are unknown.

According to the above mentions, in this study, we use a fast bandwidth selection rule based on the distance variance of all data points in the collection to determine the parameter σ. The self-adaptive distance metric based on kernel method will be detailed described in Section III.C.

III. METHODOLOGY

Motivated by the above descriptions, we improve FLICM by introducing a trade-off weighted fuzzy factor and kernel method. The details of the proposed algorithm, termed as KWFLICM for short, will be described in this section.

A. General Framework of KWFLICM Algorithm

In KWFLICM, the objective function is defined as follow

$$J_m = \sum_{i=1}^{N}\sum_{k=1}^{c} u_{ki}^m (1 - K(x_i, v_k)) + G'_{ki} \tag{10}$$

while the reformulated fuzzy factor is written as follow

$$G'_{ki} = \sum_{i=1}^{N}\sum_{k=1}^{c} u_{ki}^m \sum_{\substack{i \neq j \\ j \in N_i}} w_{ij}(1 - u_{ki})^m (1 - K(x_j, v_k)) \tag{11}$$

where N_i stands for the set of neighbors in a window around x_i, w_{ij} is the trade-off weighted fuzzy factor of jth in a local window around x_i, $1 - K(x_i, v_k)$ represents a non-Euclidean distance measure based on kernel method, $(1 - u_{ki})^m$ is a penalty which can accelerate the iterative convergence to some extent. $\{v_k\}_{k=1}^c$ is the centers or prototypes of the clusters and the array $\{u_{ki}\}$ represents a membership matrix which also must satisfy the Eq. (2).

The two updating formulas for minimizing J_m, with respect to u_{ki} and v_k is obtained as follows

$$u_{ki} = \frac{1}{\sum_{l=1}^{c} \left(\frac{(1-K(x_i,v_k)) + \sum_{\substack{j \in N_i \\ j \neq i}} w_{ij}(1-u_{kj})^m (1-K(x_j,v_k))}{(1-K(x_i,v_l)) + \sum_{\substack{j \in N_i \\ j \neq i}} w_{ij}(1-u_{kj})^m (1-K(x_j,v_l))} \right)^{1/(m-1)}} \tag{12}$$

$$v_k = \frac{\sum_{i=1}^{N} \left(u_{ki}^m K(x_i, v_k)\, x_i\right)}{\sum_{i=1}^{N} \left(u_{ki}^m K(x_i, v_k)\right)} \tag{13}$$

Thus, the proposed algorithm can be summarized as follows

Step 1: Set the number c of the cluster prototypes, fuzzification parameter m, window size N_i and the stopping condition ε.
Step 2: Initialize randomly the fuzzy cluster prototypes.
Step 3: Set the loop counter $b = 0$.
Step 4: Calculate the trade-off weighted fuzzy factor w_{ij} and the modified distance measurement D_{ik}^2, as described in Section III.B and C, respectively.
Step 5: Update the partition matrix using Eq. (12).
Step 6: Update the cluster prototypes using Eq. (13).
Step 7: If $\max |V_{new} - V_{old}| < \varepsilon$ then stop, otherwise, set $b = b + 1$ and go to step 4.

where $V = [v_1, v_2, \ldots, v_c]$ are the vectors of the cluster prototypes.

When the algorithm has converged, a defuzzification process takes place to convert the fuzzy image to the crisp segmented image.

B. Trade-Off Weighted Fuzzy Factor

The noise resistance property of the proposed KWFLICM mainly relies on the fuzzy factor G'_{ki}, as shown in Eq. (11). The adaptive trade-off weighted fuzzy factor depends on the local spatial constraint and local gray-level constraint.

For each pixel x_i with coordinate (p_i, q_i), the spatial constraint reflects the damping extent of the neighbors with the spatial distance from the central pixel and defined as

$$w_{sc} = 1/\left(d_{ij} + 1\right) \tag{14}$$

where the ith pixel is the center of the local window N_i and the jth pixel represents the set of the neighbors falling into the window around the ith pixel, d_{ij} is the spatial Euclidean distance between the jth pixel in neighbors and the central pixel. The definition of the spatial component makes the influence of the pixels within the local window change flexibly according to their distance from the central pixel and thus more local information can be used. The 3×3 window, in Fig. 1(a), extracted from the noise image is an example of this situation, and Fig. 1(c) depicts its damping extent of the neighbors with the spatial distance. The 3×3 window, in Fig. 1(a), extracted from the noise image is an example of this situation, and Fig. 1(c) depicts its damping extent of the neighbors with the spatial distance. Fig. 2 shows the changes of membership values corresponding to Fig. 1 (a) by introducing the fuzzy factor G'_{ki}. As shown in Fig. 2, it is clearly shown that the corresponding membership values of the noisy, as well as of the no-noisy pixels gradually tend to a similar value after iteration by iteration, ignoring the noisy pixels. And after five iterations the algorithm converges. In such case, the gray level values of the noisy pixels are different from the other pixels within the window, while the fuzzy factor G'_{ki} balances their membership values. Thus, all pixels within the window belong to one cluster. Therefore, the combination of the spatial and the gray level constraints incorporated in the factor G'_{ki} suppress the influence of the noisy pixels. Moreover, the factor G'_{ki} is automatically determined rather than artificially set, even in the absence of any prior noise knowledge. Hence, the algorithm becomes more robust to the outliers.

After that, we get the local coefficient of variation C_j for each pixel j as follow

$$C_j = \frac{var(x)}{(\bar{x})^2} \tag{15}$$

where $var(x)$ and $\bar{x}$ are the intensity variance and mean in a local window of the image, respectively.

Next we project C_j into kernel space. Then, the weights are normalized. Due to the fast decay of the exponential kernel, large distance between C_j and the mean of these local coefficients of variation will lead to nearly zero weights. Finally, according to comparing C_j with $\overline{C}$ (the mean of C_j in local window), we give a varying compensation to C_j, which can enlarge the discrepancy of damping extent in neighborhood.

Thus, the formulas show as follows

$$\overline{C} = \frac{\sum_{j \in N_i} C_j}{n_i} \tag{16}$$

$$\xi_{ij} = \exp\left[-\left(C_j - \overline{C}\right)\right], \qquad j \in N_i \tag{17}$$

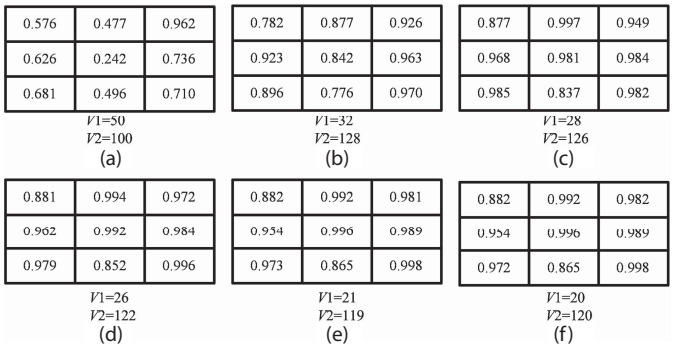

Fig. 2. Corresponding membership values of a 3×3 window with noise [as shown in Fig. 1(a)] and the cluster centers $v1$ and $v2$. (a) Initial membership values. Membership values (b) after one iteration, (c) after two iterations, (d) after three iterations, and (e) after four iterations. (f) Convergent membership values.

$$\eta_{ij} = \frac{\xi_{ij}}{\sum\limits_{k \in N_i} \xi_{ik}} \tag{18}$$

$$w_{gc} = \begin{cases} 2 + \eta_{ij} & C_j < \overline{C} \\ 2 - \eta_{ij} & C_j \geq \overline{C} \end{cases} \tag{19}$$

where the ith pixel is the center of the local window N_i, the jth pixel represents the set of the neighbors falling into the window around the ith pixel. The constant 2 guarantees the weight w_{gc} be non-negative. C_j represents the local coefficient of variation, which explain the local distribution of the jth pixel, $\overline{C}$ is the mean value of C_j that located in a local window and n_i is its local cardinality.

Therefore, the trade-off weighted fuzzy factor is written as

$$w_{ij} = w_{sc} \cdot w_{gc}. \tag{20}$$

The value of C_j reflects gray value homogeneity degree of the local window. It exhibits high values at edges or in the area corrupted by noise and produces low values in homogeneous regions. The damping extent of the neighbors with local coefficient of variation is measured by the areal type of the neighbor pixels located. If the neighbor pixels and the central pixel are located in the same region, such as homogeneous region or the area corrupted by noise, the results of local coefficient of variation obtained by them will be very close, and vice versa. In addition, it helps to exploit more local context information since the local coefficient of variation of each pixel is computed in its local window. Furthermore, the weight of the neighboring pixel will be increased to suppress the influence of outlier after transformed into the kernel space and added the spatial constraint. Here, two cases are presented for examples.

Case 1: The central pixel is not a noise and some pixels within its local window may be corrupted by noise. A 3×3 window that extracted from Fig. 3(a) depicts this situation, as shown Fig. 3(b) A, marked by bold square. According to the results from the Fig. 3, situation A, we can see that the weighting of the pixel with the noisy pixels in its neighbors is smaller than the pixels without the noisy pixels. The more the noisy pixel is contained, the smaller the weight value is. In other words, as long as the pixel is around the noisy pixel, the damping extent of the neighbors will be increase. Then, the weighting added the spatial constraint is increased to suppress the influence of the outlier, as shown in Fig. 3(f).

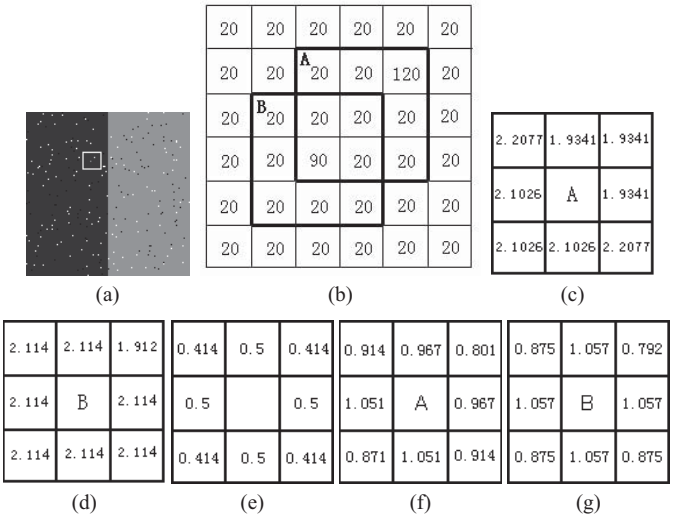

Fig. 3. Weights of a synthetic image in different conditions. (aInitial image. (b) 6×6 window with noise, marked with a rectangle in the initial image where the number in each cell represents the intensity value. (c) and (d) Obtained weight in two different cases. (e) Local spatial constraint. (f) and (g) Weight added to the spatial constraint in two cases.

Case 2: The central pixel is corrupted by noise, while the other pixels within its local window are homogenous, not corrupted by noise. An example illustrates this situation, demonstrated in Fig. 3(b), situation B marked by bold square. In this case, the gray level differences between the neighboring pixels and the central pixel are somewhat different. If the distribution of the neighboring pixels is the same, the damping extent of the neighbors principally depends on the spatial distance difference, as shown in Fig. 3(g). Hence, by using spatial restriction and local gray-level relationship, we can accurately estimate the discrepancy of neighboring pixels.

C. Non-Euclidean Distance Based on Kernel Metric

Following [9], the objective function in KWFLICM is

$$J_m = \sum_{i=1}^{N} \sum_{k=1}^{c} u_{ki}^m \left\| \Phi(x_i) - \Phi(v_k) \right\|^2 + G'_{ki} \tag{21}$$

where $\Phi(\bullet)$ is an implicit nonlinear map. The inner product between $\Phi(x_i)$ and $\Phi(v_k)$ in the feature space is $\Phi(x_i)^T \Phi(v_k) = K(x_i, v_k)$.

Then, through the kernel substitution, we have

$$\begin{aligned} \left\| \Phi(x_i) - \Phi(v_k) \right\|^2 &= (\Phi(x_i) - \Phi(v_k))^T (\Phi(x_i) - \Phi(v_k)) \\ &= \Phi(x_i)^T \Phi(x_i) - \Phi(v_k)^T \Phi(x_i) \\ &\quad - \Phi(x_i)^T \Phi(v_k) + \Phi(v_k)^T \Phi(v_k) \\ &= K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k). \end{aligned} \tag{22}$$

In this way, a new class of non-Euclidean distance measures in original data space is obtained. Because of $K(x, x) = 1$ for all $\mathbf{x}$ and the GRBF kernels, then Eq. (21) can be rewritten

as

$$J_m = \sum_{i=1}^{N}\sum_{k=1}^{c} u_{ki}^m \left(1 - K\left(x_i, v_k\right)\right) + G'_{ki} \tag{23}$$

where

$$K\left(x_i, v_k\right) = \exp\left(-\frac{\|x_i - v_k\|^2}{\sigma}\right) \tag{24}$$

here, the parameter σ is the bandwidth. The bandwidth setting rule based on the distance variance of all data points is defined as follows

Given the dataset $\Omega = \{x_1, x_2, \ldots, x_N\}$, then the data center of dataset Ω is given by

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}. \tag{25}$$

Let $d_i = \|x_i - \overline{x}\|$ be the distance from data point x_i to the data center $\overline{x}$. The mean distance of d_i is then calculated by

$$\overline{d} = \frac{\sum_{i=1}^{N} d_i}{N}. \tag{26}$$

The bandwidth is set to the variance of d_i show as follow

$$\sigma = \left(\frac{1}{N-1}\sum_{i=1}^{N}\left(d_i - \overline{d}\right)^2\right)^{\frac{1}{2}}. \tag{27}$$

The distance variance of the data points represents the degree of aggregation around the clusters. The small value of variance means the clusters are compact and well separated around cluster centers. That is, if the data set with distinguishable clusters, the membership should present separate. However, if the data set with fuzzy or undistinguished clusters, the value of variance is large. The membership should present well fuzzy distribution for all data points. This could contribute to suppressing the influence of the outliers. Fig. 4 shows different membership curves for different parameter σ in the case of two clusters. The range of the value for the tested artificial data set $\mathbf{x}$ is 0 to 3, with two cluster center. One center is 0, and the other center is 2. The distribution of these data points, as shown in Fig. 4, is undistinguished. It is just like the two compact clusters corrupted by the noisy point. Therefore, the value of distance variance for this data distribution is relatively large. So, the distance variance of them is relatively a little large. The membership should present fuzzy distribution. As we can see from these curves, when $\sigma = 0.05$, the membership curve presents that the belonging of some data points are ambiguous since the membership values of them are close to 0.5. When $\sigma = 20$, the membership curve well fits the concentrative distribution of the data set. The belonging of these data points is distinct. Therefore, the result reveals that the distance variance of the data points, as form Eq. (27), well estimates the kernel parameter σ.

Then, the distance metric based on kernel method can be transformed as

$$D_{ik}^2 = 1 - K\left(x_i, v_k\right) = 1 - \exp\left(-\frac{\|x_i - v_k\|^2}{\sigma}\right). \tag{28}$$

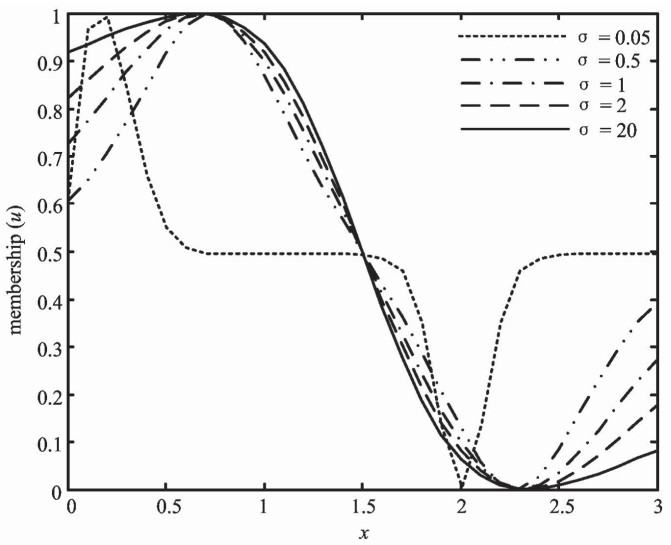

Fig. 4. Different membership curves for different parameter σ based on the dataset x with 0 as one center and 2 as the other center.

From the above descriptions, we can see that the trade-off weighted fuzzy factor and the kernel distance measure are both free of the empirically adjusted parameters which can be incorporated into other fuzzy c-means algorithms easily.

IV. Experimental Study

In this section, we describe the experimental results on three synthetic images, four medical images and three natural images with different types of noises. In addition, we test and compare the efficiency and the robustness of the proposed method (KWFLICM) with spectral graph grouping using *Nyström* method (NNcut) [25] and its predecessors, FLICM [1], RFLICM [12] and WFLICM (only introduce the trade-off weighted fuzzy factor into FLICM algorithm, here termed as WFLICM for short). The effectiveness of the trade-off weighted fuzzy factor can be validated by comparisons between FLICM and WFLICM. And the effectiveness of the kernel metric can be validated by comparisons between WFLICM and KWFLICM.

A. Results on the Synthetic Images

Our first experiment applies these algorithms to three synthetic test images. The first image with 128×128 pixels includes two classes with two intensity value taken as 20 and 120, as shown in Fig. 6(a). The number of clusters is 2. The other two images are constituted by 244×244 pixels and 256×256 pixels, as shown in Fig. 7(a) and Fig. 8(a), respectively. The number of clusters is 4. We test the algorithms on these images corrupted by different levels of Gaussian and Salt & Pepper noise. In our numerical experiments, we generally choose $\varepsilon = 0.001$ and $N_R = 8$ (a 3×3 window centered around one pixel, except the central pixel itself) [8].

In addition, the denoising performances of the above algorithms were compared with respect to the optimal segmentation accuracy (SA), where SA is defined as the sum of

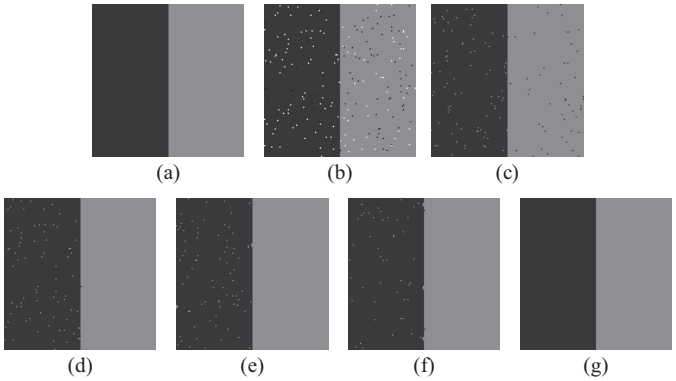

Fig. 5. Segmentation results on the first synthetic image corrupted by Salt & Pepper noise (15%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

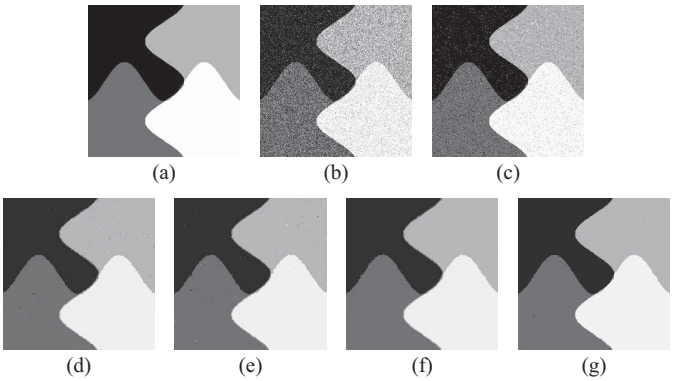

Fig. 6. Segmentation results on the second synthetic image corrupted by Gaussian noise (30%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

the correctly classified pixels divided by the sum of the total number of the pixels [5]

$$SA = \sum_{i=1}^{c} \frac{A_i \cap C_i}{\sum_{j=1}^{c} C_j} \tag{29}$$

here c is the number of clusters, A_i represents the set of pixels belonging to the ith class found by the algorithm, while C_i represents the set of pixels belonging to the ith class in the reference segmented image.

Fig. 5 illustrates the clustering results of corrupted by Salt & Pepper noise (15%) image with 128×128 pixels. As shown in Fig. 5(c)–(e), NNcut, FLICM and RFLICM are respectively affected by the noise to different extents. Especially, NNcut have had bad performance in the presence of 'Salt & Pepper' noise. Visually, Fig. 5(f) shows that WFLICM removes a large proportion of the noise, but its result is still not satisfactory enough. Its denoising result shows that this trade-off factor can adaptively control the relationship of neighbors and accurately estimate the damping extent of neighboring pixels. On the other hand, Fig. 5(g) shows that the proposed algorithm removes almost all the added noise achieving robust and satisfactory result.

Table I gives the segmentation accuracy of the above mentioned algorithms on noisy images corrupted by Gaussian noise and Salt & Pepper noise with varying degrees, respectively. The noise levels are 15%, 20% and 30%. It can be seen that the SA value of KWFLCM is larger than the other three compared algorithms in different noise levels (KWFLICM were up to 99.98%, corrupted by 30% Salt & Pepper noise). Experimental results show that the area segmented by the proposed method is clear and with high veracity.

Furthermore, the segmentation results of these two complex synthetic images are shown in Figs. 6 and 7. These two images are constituted by 244×244 pixels and 256×256 pixels, added Gaussian noise (30%) and Salt & Pepper noise (30%), respectively.

Fig. 6 shows the segmentation results of these five methods on the image corrupted by Gaussian noise (30%). Although, the segmentation result of NNcut, as shown in Fig. 6(c), keeps the clear edges, the noise was still remained in image. As shown in Fig. 6(d) and Fig. 6(e), FLICM and RFLICM are both affected by the noise to different extents. Visually, Fig. 6(f) shows that WFLICM removes a large proportion

TABLE I
SEGMENTATION ACCURACY (SA%) ON THE FIRST SYNTHETIC IMAGE WITH DIFFERENT NOISES

Noise	NNcut	FLICM	RFLICM	WFLICM	KWFLICM
Gaussian 15%	94.672	99.969	99.972	99.981	**99.988**
Gaussian 20%	91.699	99.945	99.951	99.945	**99.982**
Gaussian 30%	87.103	99.786	99.863	99.872	**99.878**
Salt & Pepper 15%	99.225	99.518	99.561	99.652	**99.996**
Salt & Pepper 20%	99.061	99.426	99.499	99.536	**99.994**
Salt & Pepper 30%	98.621	99.145	99.182	99.188	**99.988**

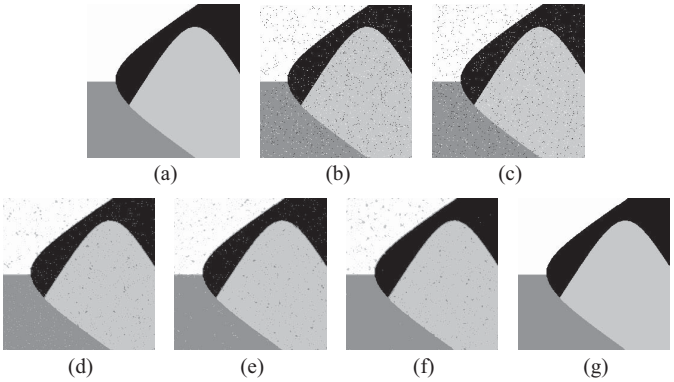

Fig. 7. Segmentation results on the third synthetic image corrupted by Salt & Pepper noise (30%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

of the noise, but the image edges are still blurry. On the other hand, Fig. 6(g) shows that KWFLICM removes almost all the added noise and maintains the image clear edge, achieving satisfactory segmentation performance. So we can find that the denoising performance of algorithm with the trade-off weighted fuzzy factor is well since the accurate estimation of relationship in neighbors, compared with FLICM and RFLICM, but it is deficient about clear edges and details persevering, particularly real images corrupted by uniform distribution noise. However, the proposed algorithm using kernel metric makes up for it.

Fig. 7 shows the segmentation results of these five methods on the image corrupted by Salt & Peppernoise (30%). As shown in Fig. 7(c)–(f), NNcut, FLICM, RFLICM and WFLICM are affected by the noise to different extents. Visually, Fig. 7(g) shows that KWFLICM removes almost all the added noise and maintains the image clear edge, achieving satisfactory segmentation performance.

The experiment further proves that the proposed algorithm can achieve more satisfying segmentation performance than the other four compared algorithms. And the result is verified by the S values, a quantitative index [8], [26]. The mathematical expression of S presents as follow

$$S = \sum_{i=1}^{c} \frac{A_i \cap A_{refi}}{A_i \cup A_{refi}} \tag{30}$$

where A_i represents the set of pixels belonging to the *ith* class found by the algorithm, while $A_{ref}i$ represents the set of the pixels belonging to the *ith* class in the reference segmented image.

This index is in fact a fuzzy similarity measure, indicating the degree of equality between A_i and A_{refi}, and the larger the S value is, the better the clustering is. Table II gives the quantitative comparison scores on these two complex synthetic images. It is observed from Table II that the quantitative index of KWFLICM is larger than the others. Figs. 6 and 7 and Table II show that the proposed algorithm is more robust to different kinds of noises and outperforms the compared algorithms.

B. Results on Medical Images

In this experiment, four magnetic resonance (MR) images are adopted to test the performance of NNcut, FLICM, RFLICM, WFLICM and KWFLICM. In general, the MR images are contaminated by Rician noise [27]. In this experiment, Rician noise is generated by a code obtained from Ged Ridgway [28].

Figs. 8 and 9 present a comparison of segmentation results on BrainWeb images [29]. BrainWeb is a database, providing simulated brain MRI data for several acquisition modalities (T1, T2, etc.) and acquisition parameters. Each 3-D image is provided with anatomical ground truth which provides tissue class label for each intracranial voxel. For the experiments, the considered BrainWeb data have been chosen with classical acquisition parameters, namely using high-resolution T1_weighted phantom with slice thickness of 1mm resolution, 9% Rician noise ($l = 9$), leading to a size of $181 \times 217 \times 181$ voxels. Two slices in the axial plane with the sequence of 90 and 180, respectively, as shown in Fig. 8(a) and Fig. 9(a).

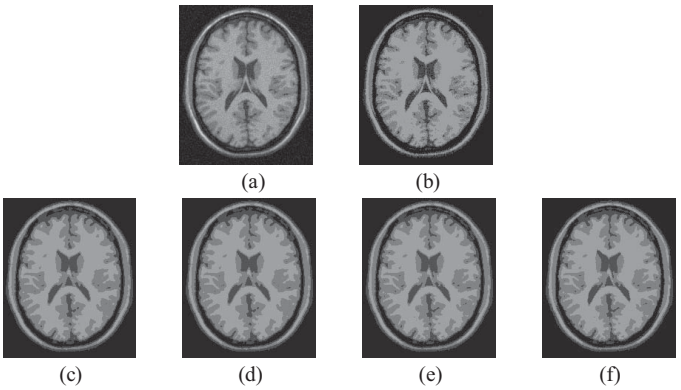

Fig. 8. Segmentation results on the first simulated MR image (*SMR1*) corrupted by Rician noise ($l = 9$). (a) Original image corrupted by Rician noise. (b) NNcut result. (c) FLICM result. (d) RFLICM result. (e) WFLICM result. (f) KWFLICM result.

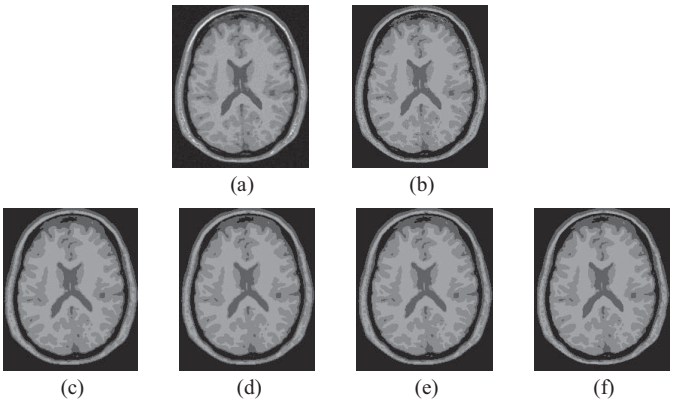

Fig. 9. Segmentation results on the second simulated MR image (*SMR2*) corrupted by Rician noise ($l = 9$). (a) Original image corrupted by Rician noise. (b) NNcut result. (c) FLICM result. (d) RFLICM result. (e) WFLICM result. (f) KWFLICM result.

Figs. 8(b)–(f) and 9(b)–(f) show the segmentation results on two slices obtained by NNcut, FLICM, RFLICM, WFLICM and KWFLICM, respectively. The images were segmented into four classes corresponding to background, cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM). From Fig. 8 and Fig. 9, it can be observed that NNcut has bad performance in the presence of Rician noise, while FLICM, RFLICM, WFLICM and KWFLICM achieve satisfactory results for removing the effect of the added noises. Furthermore, the latter two are a little superior to the other two algorithms for the effective retention capacity of the details. Table III gives the quantitative comparison scores of the experiments on these two simulated brain MRI images for accuracy of the CSF, GM and WM. From the measures of Table III, it appears that KWFLICM produces slightly better results than the other compared algorithms. The results illustrate that the proposed algorithm is robust to the noise and maintains the more details.

The other two MR images with 256×256 pixels have been shown in Figs. 10(a) and 11(a). Here, we added Rician noise of noise level ($l = 20$) to these two images, as shown in Figs. 10(b) and 11(b). The segmentation results on these two MR images with Rician noisy are presented in Figs. 10(c)–(g)

TABLE II
COMPARISON SCORES (S%) OF THE FIVE ALGORITHMS FOR THE TWO SYNTHETIC IMAGES

Image	Noise	NNcut	FLICM	RFLICM	WFLICM	KWFLICM
The second synthetic image	Gaussian 30%	92.951	98.826	98.861	98.891	**99.648**
	Salt & Pepper 30%	97.264	97.358	97.521	97.862	**99.971**
The third synthetic image	Gaussian 30%	92.872	97.217	97.581	98.239	**99.614**
	Salt & Pepper 30%	97.495	96.621	96.945	97.631	**99.956**

TABLE III
COMPARISON SCORES (S%) OF THE FIVE ALGORITHMS FOR THE TWO SIMULATED MR IMAGES

Images	regions	NNcut	FLICM	RFLICM	WFLICM	KWFLICM
SMR1	CSF	66.114	71.597	71.055	72.108	**72.298**
	GM	71.012	82.138	81.835	82.670	**83.364**
	WM	86.019	89.581	89.426	89.684	**90.149**
SMR2	CSF	65.238	65.070	65.068	65.635	**65.978**
	GM	80.469	80.564	80.372	80.756	**81.877**
	WM	85.945	87.520	87.371	87.612	**88.401**

and 11(c)–(g). According to the segmentation results, we can see that the proposed algorithm can effectively remove the noise and retain more useful image edge and details in the MR images, compared with NNcut, FLICM, RFLICM and WFLICM.

Because of without the ground truth, we can not calculate the segmentation accuracy. In order to evaluate the performances of these five methods, an objective evaluation criteria, the entropy-based information [30], is introduced here. This evaluate criteria was commonly used for assessing the segmentation performance. The entropy for region j is defined as

$$H(R_j) = -\sum_{m \in V_j} \frac{L_j(m)}{S_j} \log \frac{L_j(m)}{S_j}. \tag{31}$$

The expected region entropy of segmentation image

$$H_r(I) = \sum_{j=1}^{C} (\frac{S_j}{S_I}) H(R_j). \tag{32}$$

The layout entropy is defined as

$$H_l(I) = -\sum_{j=1}^{C} (\frac{S_j}{S_I}) \log \frac{S_j}{S_I}. \tag{33}$$

The entropy-based evaluation function, E, which combines both the layout entropy with the expected region entropy, is defined as

$$E = H_l(I) + H_r(I) \tag{34}$$

where, C is the number of regions (clusters), j (denoted as R_j) is a sub-region of original image, $L_j(m)$ denotes the number of pixels in region j (*jth* cluster) that have gray level value equal to m, V_j represents the set of all possible gray level values in region R_j, $S_j = \left|R_j\right|$ indicates the number of pixels in region R_j. The idea of this validity index is that the partition should maximize the uniformity of pixels within each segmented region, and minimize the uniformity across the regions. So the best segmentation performance is achieved when E is minimal.

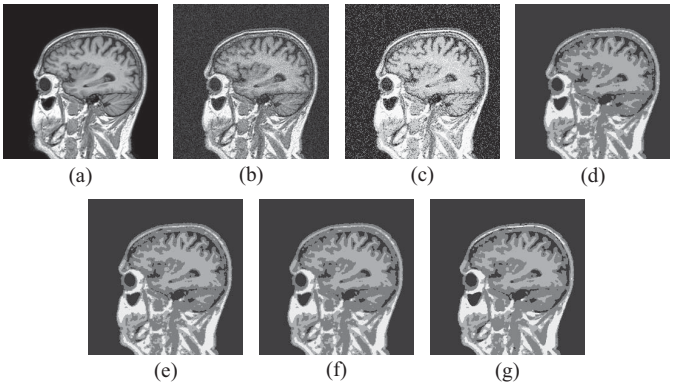

Fig. 10. Segmentation results on the first MR image (*MR1*) corrupted by Rician noise ($l = 20$). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

The comparison results on the two MR images are given in Table IV. It is observed from the results that the value E of the proposed method is less than the others, KWFLICM were up to 2.3839 (*MR1*) and 2.1069 (*MR2*), corrupted by $l = 20$ Rician noise. At the same time, the values of $H_r(L)$ and $H_l(L)$ obtained by KWFLICM are also less compared with those obtained by the other four compared algorithms. It attributes that the introduction of trade-off weighted fuzzy factor and kernel metric guarantees relative insensitivity both to noise and outlier. From Figs. 10 and 11 and Table IV, the experimental result indicates that the proposed algorithm can maintain the details of the MR images while removing the noise.

C. Results on Natural Images

In the coming experiments, we also applied the same five algorithms on three natural images. The first natural image is

TABLE IV
NUMERICAL RESULTS OF THE FIVE ALGORITHMS ON TWO MR IMAGES

Image	Metric	NNcut	FLICM	RFLICM	WFLICM	KWFLICM
MR1	$H_r(L)$	1.8687	1.9665	1.8749	1.8881	1.8747
	$H_l(L)$	0.5347	0.5093	0.5114	0.5117	0.5092
	E	2.4034	2.4759	2.3863	2.3998	**2.3839**
MR2	$H_r(L)$	1.8037	1.8060	1.8070	1.8032	1.7980
	$H_l(L)$	0.3375	0.3113	0.3114	0.3114	0.3089
	E	2.1413	2.1173	2.1148	2.1146	**2.1069**

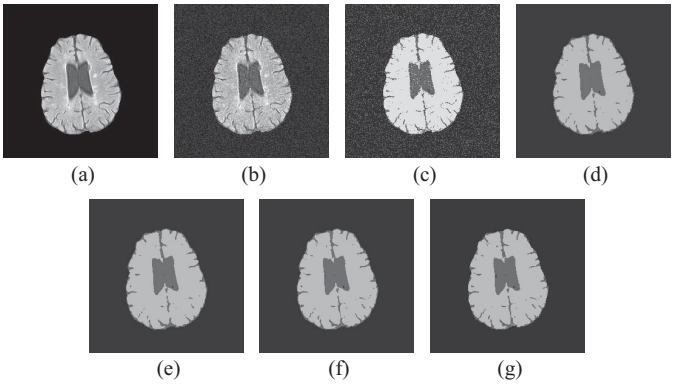

Fig. 11. Segmentation results on the second MR image (*MR2*) corrupted by Riciannoise ($l = 20$). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

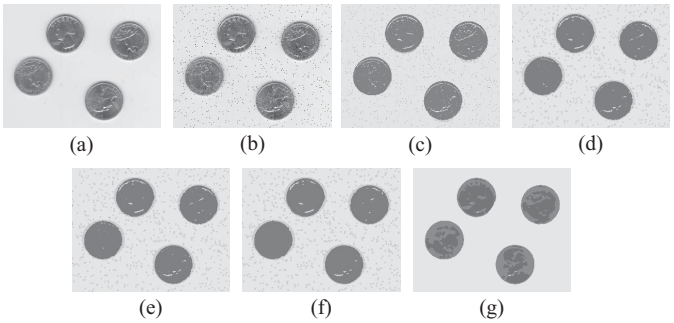

Fig. 12. Segmentation results on the *coins* image corrupted by Salt & Pepper noise (20%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

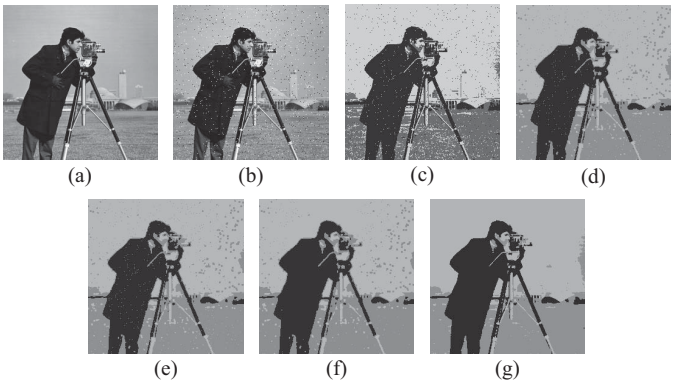

Fig. 13. Segmentation results on the *cameraman* image corrupted by Salt & Pepper noise (20%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

mainly composed of *coins* and background, with 308×242 pixels, corrupted by 20% Salt & Pepper noise and shown in Fig. 12(a)–(b). The number of clusters is 3. Fig. 12(c)–(g) shows the segmentation results obtained by NNcut, FLICM, RFLICM, WFLICM and KWFLICM, respectively. The results illustrate that the proposed algorithm is superior to the other four compared algorithms. KWFLICM is robust to the noise and maintains the clear image edges and the more details.

Figs. 13(a) and 14(a) show two natural images about a *cameraman* and *flowers*, with 256×256 pixels and 128×128 pixels, respectively. The two images are corrupted by Salt & Pepper (20%) and Gaussian noise (15%), as shown in Figs. 13(b) and 14(b), respectively. The number of clusters is 3. Figs. 13 and 14 show that the results obtained from KWFLICM have smoother regions and much clearer image edge while removing almost added noise. Table V gives the numerical results on the three natural images. Visually, the smallest value E can be obtained using the proposed method.

All the experiment results show that KWFLICM can remove the noise while preserving significant image details and obtain the good performance. Furthermore, it is relatively independent of the type of noises.

Finally, Fig. 15 illustrates the computational cost on the images with different size for NNcut, FLICM, RFLICM, WFLICM and KWFLICM. We artificially added Gaussian and Salt & Pepper noises on these images. All experiments performed on a Pentium IV (3 GHz) workstation under Windows XP Professional using MATLAB. As the result is shown, the NNcut using the *Nyström* method is much faster than other algorithms. The proposed algorithm is computational consuming, since the fuzzy factor G'_{ki} is computed in each iteration step. But this drawback is compensated for its very good performance as it was shown above. Furthermore, kernel means, performing a nonlinear data transformation into high dimensional feature space via nonlinear mapping, increases the probability of the linear separability of the data within the transformed space.

V. CONCLUSION

The kernel method has been recently applied to unsupervised clustering. In this study, we propose an unsupervised FCM algorithm based on the kernel metric for segmentation of

TABLE V
NUMERICAL RESULTS OF THE FIVE ALGORITHMS ON THE THREE NATURAL IMAGES

Image	Metric	NNcut	FLICM	RFLICM	WFLICM	KWFLICM
Coins	$H_r(L)$	1.2488	1.2239	1.2244	1.2221	1.2035
	$H_l(L)$	0.2927	0.3707	0.3711	0.3703	0.3131
	E	1.5415	1.5947	1.5955	1.5924	**1.5166**
Cameraman	$H_r(L)$	1.7402	1.7376	1.7391	1.7300	1.7041
	$H_l(L)$	0.4468	0.4612	0.4611	0.4611	0.4593
	E	2.1870	2.1988	2.2002	2.1911	**2.1635**
Flowers	$H_r(L)$	1.7006	1.6906	1.6919	1.6691	1.6090
	$H_l(L)$	0.4307	0.4414	0.4416	0.4423	0.4334
	E	2.1312	2.1320	2.1135	2.1414	**2.0424**

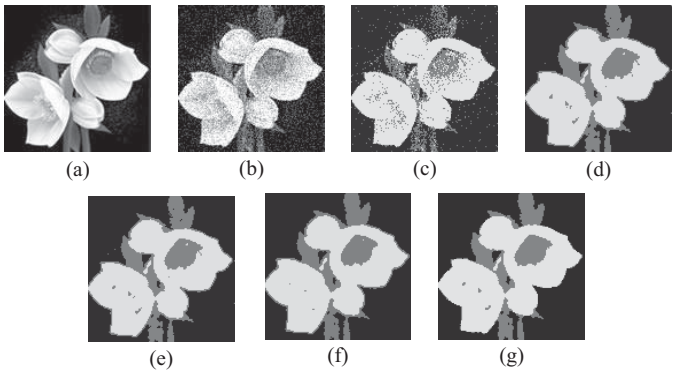

Fig. 14. Segmentation results on the *flowers* image corrupted by Gaussian noise (15%). (a) Original image. (b) Noisy image. (c) NNcut result. (d) FLICM result. (e) RFLICM result. (f) WFLICM result. (g) KWFLICM result.

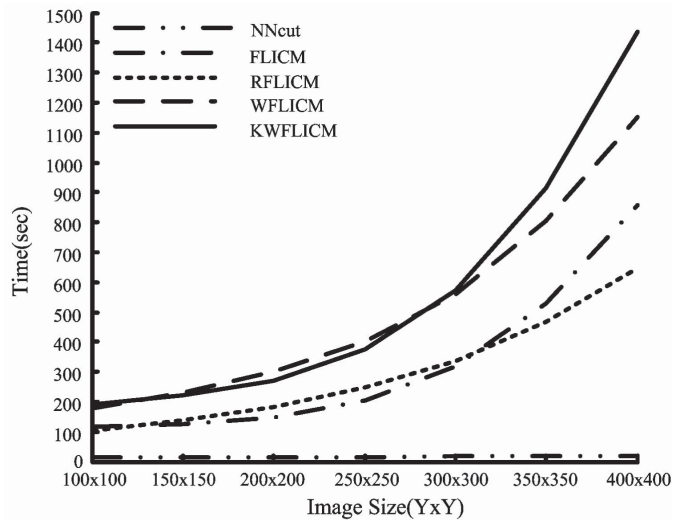

Fig. 15. Running time of the five algorithms.

images that have been corrupted by intensity inhomogeneities and noise. The results reported in this paper show that the kernel metric is an effective approach to constructing a robust image clustering algorithm. Furthermore, the improved algorithm introduced a reformulated spatial constraint, with the trade-off weighted fuzzy factor as a local similarity measure to make a trade-off between image detail and noise. And the new trade-off weight mainly depends on the distribution of the local information and local spatial constraint to affect the damping extent of the pixels in neighbors. Compared with its preexistences, it is able to incorporate the local information more exactly. In addition, the trade-off weighted fuzzy factor and the kernel distance measure are completely free of the empirically adjusted parameters determination, thereby allowing the automated applications. In our experiments, we test the proposed algorithm on synthetic images, medical images and natural images. The experiment results show that the proposed algorithm obviously improves the performance of image segmentation, as well as the robustness to the type of noises.

ACKNOWLEDGMENT

The authors would like to thank Dr. Krinidis for providing their FLICM source codes for comparison, and the editors and anonymous reviewers for their valuable comments and helpful suggestions, which greatly improved the paper's quality.

REFERENCES

[1] 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.

[2] X. Yin, S. Chen, E. Hu, and D. Zhang, "Semi-supervised clustering with metric learning: An adaptive kernel method," *Pattern Recognit.*, vol. 43, no. 4, pp. 1320–1333, Apr. 2010.

[3] L. Zhu, F. Chung, and S. Wang, "Generalized fuzzy C-means clustering algorithms with improved fuzzy partitions," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 39, no. 3, pp. 578–591, Jun. 2009.

[4] S. Tan and N. A. M. Isa, "Color image segmentation using histogram thresholding fuzzy C-means hybrid approach," *Pattern Recognit.*, vol. 44, no. 1, pp. 1–15, 2011.

[5] C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2007–2016, Jul. 2011.

[6] J. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1974.

[7] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[8] 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.

[9] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[10] 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 EMBS*, Nov. 2003, pp. 17–21.

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

[12] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141–2151, Apr. 2012.

[13] N. Cristianini and J. S. Taylor, *An Introduction to SVM's and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[14] X. Yang and G. Zhang, "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. 2011.

[15] V. Roth and V. Steinhage, "Nonlinear discriminant analysis using kernel functions," in *Advances in Neural Information Processing Systems 12*, S. A Solla, T. K. Leen, and K.-R. Muller, Eds. Cambridge, MA: MIT Press, 2000, pp. 568–574.

[16] B. Scholkopf, A. J. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[17] K. Wu and M. Yang, "Alternative C-means clustering algorithms," *Pattern Recognit.*, vol. 35, no. 10, pp. 2267–2278, 2002.

[18] L. Chen, C. Chen, and M. Lu, "A multiple-kernel fuzzy C-means algorithm for image segmentation," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 41, no. 5, pp. 1263–1274, Oct. 2011.

[19] T. C. Havens, R. Chitta, A. K. Jain, and J. Rong, "Speedup of fuzzy and possibilistic kernel C-means for large-scale clustering," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jun. 2011, pp. 463–470.

[20] D. Tsai and C. Lin, "Fuzzy C-means based clustering for linearly and nonlinearly separable data," *Pattern Recognit.*, vol. 44, no. 8, pp. 1750–1760, 2011.

[21] D. Zhang and S. Chen, "Kernel-based fuzzy clustering incorporating spatial constraints for image segmentation," in *Proc. IEEE 2nd Annu. Int. Conf. Mach. Learn. Cybern.*, Oct. 2003, pp. 2189–2192.

[22] R. Hathaway, J. C. Bezdek, and Y. Hu, "Generalized fuzzy C-means clustering strategies using norm distances," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 576–581, Oct. 2000.

[23] T. Przybyla, J. Jezewski, K. Horoba, and D. Roj, "Hybrid fuzzy clustering using LP norms," in *Proc. 3rd Int. Conf. Intell. Inf. Database Syst.*, 2011, pp. 187–196.

[24] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 12, no. 2, pp. 181–202, Mar. 2001.

[25] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nystrom method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 1–12, Feb. 2004.

[26] F. Masulli and A. Schenone, "A fuzzy clustering based segmentation system as support to diagnosis in medical imaging," *Artif. Intell. Med.*, vol. 16, no. 2, pp. 129–147, 1999.

[27] F. Zhao, L. Jiao, H. Liu, and X. Gao, "A novel fuzzy clustering algorithm with non local adaptive spatial constraint for image segmentation," *Signal Process.*, vol. 91, no. 4, pp. 988–999, 2011.

[28] MathWorks. (2011). *Image Processing Toolbox*, Natick, MA [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/14237

[29] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans. (2011). *BrainWeb: Online Interface to a 3D MRI Simulated Brain Database* [Online]. Available: http://www.bic.mni.mcgill.ca/brainweb/

[30] H. Zhang, J. Fritts, and S. Goldman, "An entropy-based objective evaluation method for image segmentation," *Proc. SPIE, Storage Retrieval Methods Appl. Multimedia*, vol. 5307, pp. 38–49, Jan. 2004.

Maoguo Gong (M'07) received the B.S. degree in electronic engineering (First Class Hons.) and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively.

He has been a Teacher with Xidian University since 2006, where he became an Associate Professor in 2008 and a Full Professor in 2010, respectively, both with exceptive admission. He is currently a Full Professor with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University. His current research interests include computational intelligence with applications, image segmentation, and image change detection.

Dr. Gong was the recipient of the New Century Excellent Talent Award from the University of the Ministry of Education of China, the Eighth Young Scientist Award of Shaanxi Province, the New Scientific and Technological Star of Shaanxi Province, the Excellent Young Contributor Award of Shaanxi Province, and the Science and Technology Award of Shaanxi Province (First Level, in 2008 and 2010). He is a member of the IEEE Computational Intelligence Society, an Executive Committee Member of the Natural Computation Society of the Chinese Association for Artificial Intelligence, and a Senior Member of the Chinese Computer Federation.

Yan Liang received the B.S. degree in automation from Three Gorges University, Yichang, China, in 2010, and is currently pursuing the M.S. degree with Xidian University, Xi'an, China.

Her current research interests include image segmentation.

Jiao Shi received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2009, where he is currently pursuing the Ph.D. degree.

Her current research interests include computational intelligence and image understanding.

Wenping Ma received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

She is currently an Associate Professor with Xidian University. Her current research interests include computational intelligence and image understanding.

Jingjing Ma received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2004 and 2012, respectively.

She is currently a Lecturer with Xidian University. Her current research interests include computational intelligence and image understanding.