Image segmentation by generalized hierarchical fuzzy C-means algorithm

Article *in* Journal of Intelligent and Fuzzy Systems · February 2015

DOI: 10.3233/IFS-141378

Zheng Yuhui
Nanjing University of Information Science & Technology

Byeungwoo Jeon
Sungkyunkwan University

Q. M. Jonathan Wu
University of Windsor

Journal of Intelligent & Fuzzy Systems 28 (2015) 961–973
DOI:10.3233/IFS-141378
IOS Press

Image segmentation by generalized hierarchical fuzzy C-means algorithm

Yuhui Zheng[a,c], Byeungwoo Jeon[d], Danhua Xu[a,c], Q.M. Jonathan Wu[a,b] and Hui Zhang[a,c,∗]
[a] *School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, China*
[b] *Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada*
[c] *Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, China*
[d] *College of Information and Communication Engineering, Sungkyunkwan University, Korea*

Abstract. Fuzzy c-means (FCM) has been considered as an effective algorithm for image segmentation. However, it still suffers from two problems: one is insufficient robustness to image noise, and the other is the Euclidean distance in FCM, which is sensitive to outliers. In this paper, we propose two new algorithms, generalized FCM (GFCM) and hierarchical FCM (HFCM), to solve these two problems. Traditional FCM can be considered as a linear combination of membership and distance from the expression of its mathematical formula. GFCM is generated by applying generalized mean on these two items. We impose generalized mean on membership to incorporate local spatial information and cluster information, and on distance function to incorporate local spatial information and image intensity value. Thus, our GFCM is more robust to image noise with the spatial constraints: the generalized mean. To solve the second problem caused by Euclidean distance (l2 norm), we introduce a more flexibility function which considers the distance function itself as a sub-FCM. Furthermore, the sub-FCM distance function in HFCM is general and flexible enough to deal with non-Euclidean data. Finally, we combine these two algorithms to introduce a new generalized hierarchical FCM (GHFCM). Experimental results demonstrate the improved robustness and effectiveness of the proposed algorithm.

Keywords: Fuzzy C-means, generalized mean, hierarchical distance function, image segmentation, spatial constraint

1. Introduction

Image segmentation is one of the most important and difficult problems in many applications, such as robot vision, object recognition and medical image processing. Although different methodologies [1–4] have been proposed for image segmentation, it remains a challenge due to overlapping intensities, low contrast of images, and noise perturbation. In the last decades, fuzzy segmentation methodologies, and especially the fuzzy c-means algorithms (FCM) [5], have been widely studied and successfully applied in image clustering and segmentation. Their fuzzy nature makes the clustering procedure able to retain more original image information than the crisp or hard clustering methodologies [6, 7].

Although the FCM algorithm usually performs well with non-noise images, it is still weak in imaging noise, outliers and other imaging artifacts. This may be caused by two aspects: one is the usage of the non-robust, Euclidean distance function which is not robust under noise perturbations, and the other does not pertain to any information about spatial context. Thus, two following questions are presented:

(1) How to choose the proper distance function to make the standard FCM more robust to image noise and outliers?

∗Corresponding author. Hui Zhang, E-mail: nrzhanghui@gmail.com.

(2) How to adopt the local spatial information to make the standard FCM more robust to image noise?

Several attempts have been made to compensate for these drawbacks of FCM. For example, in [8–11, 27, 28], various more robust alternatives for the distance function of the FCM algorithm have been proposed. Some researchers suggested replacing the Euclidean distance (l_2) with the city block distance (l_1) in FCM [8]. In [9], Bobrowski et al. used l_1 and l_∞ norm instead of Euclidean distance l_2 in FCM. A more general study using l_{p} norm distances can be found in [10]. The modified distance and kernel distance function were proposed in [11, 6], respectively.

To overcome the second shortcoming, a wide variety of approaches have been proposed to incorporate spatial information in the image [12]. A common approach is the use of a Markov Random Field (MRF) [13]. Such method aims to impose spatial smoothness constraints on the image pixel labels. Recently, a special case of the MRF model—the Hidden MRF (HMRF) Model—has been proposed [14, 15].

In this paper, we propose two simple and effective algorithms, generalized FCM (GFCM) and hierarchical FCM (HFCM), to overcome these two limitations.

The idea of HFCM is inspired by the hierarchical mixture of experts (HME) [21–24]. An additional feature of the hierarchical mixture classifier is that it provides class conditional density estimates as flat mixtures. In our HFCM, we assume the distance function is estimated by a sub-FCM in order to allow a more general setting. In standard FCM, this distance function is represented by Euclidean distance, l_2 norm. However, in our algorithm, each distance function is represented by a sub-FCM of two or three sub-components, which allows us to approximate non-Euclidean distance functions (such as gamma or logarithmic gamma distance for Synthetic Aperture Radar (SAR) image processing). Moreover, the idea of HFCM can be extended to various distance functions, such as l_{p} norm and kernel distance functions, etc. Thus, our algorithm is more flexible and general than standard FCM.

We incorporate generalized mean into FCM, called GFCM, to overcome the second shortcoming mentioned above. In fact, FCM can be considered as a linear combination of membership and distance function from the expression of its mathematical formula. We impose generalized mean on this two items to make the labeling of a pixel influenced by the labels in its immediate neighborhood. We impose generalized mean on membership to combine local spatial information and component information, and on distance function to combine local spatial information and observation information. We then combine our two proposed algorithms (GFCM and HFCM) and introduce Generalized Hierarchical Fuzzy C-Means (GHFCM). The performance of proposed approach, compared with state-of-the-art technologies, demonstrates its improved robustness and effectiveness.

The remainder of this paper is organized as follows: In Section 2, we provide a brief review of the generalized mean and the FCM algorithm. In Section 3, we introduce our GFCM algorithm which adopts generalized means as the spatial constraints. We introduce our HFCM which represents each distance function in standard FCM by a single sub-FCM in Section 4 and GHFCM in Section 5. The experimental results are given in Section 6. Finally, some concluding remarks are provided.

2. Mathematical backgrounds

2.1. Generalized mean (GDM)

In mathematics, a generalized mean is an abstraction of the Pythagorean means including arithmetic, geometric, and harmonic means. The generalized mean of $a_1,\ a_2,\ \ldots,\ a_n$ is defined as

$$M_p\left(a_1, a_2, \ldots, a_n\right) = \left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)^{1/p}, \qquad (1)$$

where $a_i \geq 0$, $p \in [-\infty, +\infty]$ and $\sum_{i=1}^{n} a_i = 1$.

For $p \longrightarrow 0$, (1) approaches the geometric mean

$$M_G\left(a_1, a_2, \ldots, a_n\right) = \left(\prod_{i=1}^{n} a_i\right)^{1/n}. \qquad (2)$$

For $p = 1$, (1) results in the arithmetic mean

$$M_A\left(a_1, a_2, \ldots, a_n\right) = \frac{1}{n}\sum_{i=1}^{n} a_i. \qquad (3)$$

There are some other special cases of GDM based on different p values. For example, if $p = -1$, M is harmonic mean; for $p = 2$, M is quadratic mean; under the condition $p \longrightarrow -\infty$, $M_{-\infty} = \min\left(a_1, a_2, \ldots, a_n\right)$ and the condition $p \longrightarrow \infty$, $M_\infty = \max\left(a_1, a_2, \ldots, a_n\right)$.

2.2. Fuzzy c-means algorithm

To deal with the problem of clustering N multivariate data points into J clusters, Dunn [16] introduced and later Bezdek [6] extended the fuzzy c-means clustering algorithm. In the standard FCM algorithm, the fuzzy objective function that needs to be minimized is given by

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}^m d_{ij}. \tag{4}$$

where y_i, $i = (1,2,\ldots,N)$, denotes the data set in the D-dimensional vector space, N is the total number of data points, J is the number of clusters, u_{ij} is the degree of membership of y_i in the j-th cluster, m is the weighting exponent on each fuzzy membership function u_{ij}, d_{ij} is a distance (similarity) measure between point y_i and cluster center μ_j, called distance function. The squared Euclidean distance is usually used in standard FCM, given as

$$d_{ij} = \left\| y_i - \mu_j \right\|^2. \tag{5}$$

where μ_j is the prototype (mean) center of cluster j. With the distance function in (5), the FCM algorithm is iterated through the necessary conditions for minimizing J_m with the following update equations:

$$\mu_j = \sum_{i=1}^{N} u_{ij}^m y_i \Big/ \sum_{i=1}^{N} u_{ij}^m. \tag{6}$$

$$u_{ij} = \left(d_{ij}\right)^{1/(1-m)} \Big/ \sum_{h=1}^{J} (d_{ih})^{1/(1-m)}. \tag{7}$$

with the constraint $\sum_{j=1}^{J} u_{ij} = 1$.

3. Generalized FCM

3.1. Model establishment

Let us first consider (4). It can be easily seen that the objective function is composed of two items: membership u_{ij} and distance function d_{ij}. Our algorithm is simple, easy and straightforward to modify these two items with local generalized mean. This modification incorporates more local spatial information to make the model more robust to image noise. After modification of (1), the local generalized mean is given as

$$M_p = \left(\frac{1}{\mathcal{N}_i} \sum_{c \in \mathcal{N}_i} a_c \right)^{1/p}. \tag{8}$$

where N_i is the neighborhood of the i-th pixel, including the i-th pixel. Considering weighted factor and local generalized arithmetic mean with distance measure, a new objective function can be given as

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}^m \sum_{c \in \mathcal{N}_i} w_c d_{cj}. \tag{9}$$

where w_c is the weighted factor to control the influence of the neighborhood pixels depending on their distance from the central pixel i. Generally, the strength of w_c should decrease as the distance between pixel c and i increases. One possible selection of w_c is the Gaussian function: $w_c = 1/\left(2\pi\delta^2\right)^{1/2}\exp\left(-d_{ci}^2/2\delta^2\right)$, where $d_{ci,}$ is the spatial Euclidean distance between pixels c and i, and $\delta = (\text{window size} - 1)/4$. By applying the optimization way similar to the standard FCM, the parameters in GFCM can be calculated iteratively as

$$\mu_j = \sum_{i=1}^{N}\sum_{c \in \mathcal{N}_i} w_c u_{ij}^m y_c \Big/ \sum_{i=1}^{N}\sum_{c \in \mathcal{N}_i} w_c u_{ij}^m. \tag{10}$$

$$u_{ij} = \left(\sum_{c \in \mathcal{N}_i} w_c d_{cj} \right)^{1/(1-m)} \Big/ \sum_{h=1}^{J} \left(\sum_{c \in \mathcal{N}_i} w_c d_{ch} \right)^{1/(1-m)}. \tag{11}$$

The membership function u_{ij} in (11) represents the probability that an image pixel i belongs to the j-th cluster. However, it can be seen that the spatial information of the membership function in the neighborhood of each pixel is not under consideration in (11). By applying local weighted generalized mean on membership, the modified membership u_{ij} can be re-calculated as

$$u_{ij} = \sum_{c \in \mathcal{N}_i} w_c u_{cj} \Big/ \sum_{h=1}^{J}\sum_{c \in \mathcal{N}_i} w_c u_{ch}. \tag{12}$$

where w_c is the weighted factor defined in (9). The "old" membership u_{ij} (on the right side of (12)) is calculated by (11). We then use "new" membership (on the left side of (12)) instead of "old" membership in our proposed fuzzy system. It is noted that "new" membership incorporates more image spatial information according to the help of local weighted generalized mean. Thus, our algorithm is more robust to image noise.

For a deep understanding of our algorithm, let us give some analysis for usage of generalized mean in detail. Distance function d_{ij} is a measure between point y_i and cluster center μ_j in standard FCM. In our GFCM, this distance is influenced by the distance in its immediate neighborhood to incorporate local spatial information and observation information (image intensity value). The new distance function in our model can be modified as (ignoring constant item)

$$d_{ij} = \sum_{c \in \mathcal{N}_i} w_c \left\| y_c - \mu_j \right\|^2 = \left\| \bar{y}_i - \mu_j \right\|^2 . \quad (13)$$

Equation (13) may explain why our method is more robust to image noise. Assuming image intensity y_i is corrupted by image noise, in this case, calculation of d_{ij} in standard FCM may be far away from the "true" distance function. However, in our model, this distance is calculated by modified $\bar{y}_i$ which is obtained by its immediate neighborhood.

The membership function u_{ij} represents the probability that image pixel i belongs to the cluster j. Similar to distance function, we also use generalized mean to make the membership of image pixel i influenced by the membership in its immediate neighborhood for incorporating local spatial information and cluster information, as shown in (12). That means the probability of i-th pixel belongs to the cluster j is not decided by the pixel i (u_{ij}), but by the neighborhood of pixel i ($\sum_{c \in \mathcal{N}_i} w_c u_{cj}$).

In short, our algorithm is based on an obvious fact. A single image pixel i is easy to be corrupted by noise. However, image pixels in the local neighborhood of i-th pixel are hard to be all corrupted by noise. As long as the "signal" pixels are more than "noise" pixels in this neighborhood, the correct function (membership or distance function) can always be calculated. This is the reason why we use the generalized mean to calculate "average" functions for eliminating the noise effect.

3.2. Connection to existing methods

In this subsection, we discuss the relationship and difference between our algorithm and existing methods as follows:

(1) *Compared to Ahmed's method*

Ahmed et al. [4] proposed modified FCM (MFCM) where the objective function of the classical FCM is modified in order to compensate the intensity inhomogeneity and allow the labeling of a pixel to be influenced by the labels in its immediate neighborhood. In [4], the objective function is given as

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{J} u_{ij}^m \left(d\left(y_i, \mu_j\right) + \frac{a}{N_R} \sum_{c \in \mathcal{N}_i} d\left(y_c, \mu_j\right) \right). \quad (14)$$

where N_i is the neighborhood of the i-th pixel including the i-th pixel itself, and N_R represents its cardinality. The parameter a is used to control the effect of the neighbor's term.

Compared (9) with (14), it can be easily seen that Ahmed's method is a special case of our algorithm when we set

$$w_c = \begin{cases} 1, & c = i \\ \frac{a}{N_R}, & c \in \mathcal{N}_i \ \ and \ \ c \neq i \end{cases} \quad (15)$$

In other words, our algorithm degrades to the Ahmed's method under condition (15). Furthermore, Ahmed's method only considers the neighborhood's effect on distance function but not on the membership function. In fact, one important image characteristic is that neighboring pixels are highly correlated. These neighboring pixels possess similar membership, and the probability that they belong to the same cluster is great [17].

(2) *Compared to Global generalized FCM*

Yu et al. [1] proposed a general FCM model which combines (weighted) generalized mean and FCM. The (weighted) generalized mean defined in [1] as

$$M_f(Y) = f\left(\sum_{j=1}^{J} \alpha_j g\left(y_j\right) \right). \quad (16)$$

Yu et al. separate the membership u_{ij} into two independent items a_i and α_j, and then apply generalized mean (function f) on whole component j

$$J_y = \frac{1}{N} \sum_{i=1}^{N} a_i f(S_i), \ \ \text{where} \ \ S_i = \sum_{j=1}^{J} \alpha_j g\left(d_{ij}\right). \quad (17)$$

Karayiannis et al. [2] proposed a generalized FCM model by modifying the membership constraint in standard FCM with weighted generalized mean. The modified constraint in [2] has the form

$$\left(\sum_{j=1}^{J} \beta_j \left(u_{ij}\right)^{\alpha} \right)^{1/\alpha} = 1. \quad (18)$$

and then the membership u_{ij} can be estimated as

$$u_{ij} = \frac{1}{J}\left(\sum_{h=1}^{J}\beta_h\left(\frac{d_{ij}}{d_{ih}}\right)^{\alpha/(1-m)}\right)^{-1/\alpha}. \quad (19)$$

It is noted that models in [1, 2] degrade to the standard FCM when $p = 1$ in (17) and $\alpha = 1$ in (18–19). Compared to our local GFCM algorithm, none local spatial information is taken into account in global generalized FCM methods in [1, 2].

(3) *Compared to MRF/HMRF*

Both generalized mean and MRF/HMRF [12–15] can be considered as spatial constraints to incorporate some local spatial information, and to make the model more robust to image noise. In this paper, we adopt GDM instead of MRF/HMRF for three reasons: (1) MRF/HMRF is complex and time consuming. In contrast, GDM is simple, easy and fast for implementation. (2) MRF/HMRF resorts to the additional parameter β to keep a balance between robustness to noise and image sharpness and details. However, our model is fully free of the empirically adjusted parameter β. (3) In MRF/HMRF, for a 2-D image, the definition of neighbors (i–1 and i+1) extends to horizontal, vertical and diagonal pixels, which become a 3×3 square window. In our model, the neighborhood window size of GDM can be selected as 3×3, 5×5, 7×7, etc. Moreover, although a square window is used in this paper, windows with other shapes (e.g. diamonds or circles) can also be suitable.

(4) *Extension of our algorithm*

Let us consider another type of fuzzy algorithm called conditional fuzzy c-means (CFCM) which is introduced by [18]. CFCM is very similar as standard FCM except the membership constraint. The constraint in CFCM is

$$u_{ij} \in [0, 1], \text{ and } \sum\nolimits_{j=1}^{J} u_{ij} = f_i. \quad (20)$$

where f_i describes a level of involvement of y_i in the constructed cluster.

Recently, Wei et al. [19] modified the FCM objective function and present a novel FCM (NFCM)

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}^m d_{ij} + \lambda \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}\log u_{ij}. \quad (21)$$

where $\lambda \geq 0$. It can be seen that our model is general enough and the idea of usage of generalized mean can also be applied to distance function and membership in CFCM and NFCM.

In [20], Yang proposed a new FCM generalization, called a penalized FCM (PFCM), based on the modified fuzzy objective function as follows

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}^m d_{ij} + \omega \sum_{i=1}^{N}\sum_{j=1}^{J} u_{ij}^m \log\alpha_j. \quad (22)$$

where $\omega, \alpha_j \geq 0$ and $\sum_{j=1}^{J}\alpha_j = 1$. It can be seen that our generalized mean algorithm can be extended and then be applied to the penalized item α_{ij} in PFCM.

From the discussion above, we can conclude that although our algorithm focus on standard FCM, it can also be easily extended to improve the performance of other FCM-like algorithms based on adding some type of penalty terms to the original FCM objective function. However, to clearly demonstrate our idea, here we impose generalized mean on standard FCM.

4. Hierarchical fuzzy c-means

In this subsection, we introduce a more flexible fuzzy algorithm called hierarchical fuzzy c-means (HFCM). Our idea is straightforward, simple and easy to implementation. We assume the distance function in traditional fuzzy model itself is a sub-fuzzy model. Thus, our algorithm is general enough and can be applied to various distance functions, whatever the function type (Euclidean distance, l_1, l_p, l_∞ norm, or kernel distance function, etc.). However, to clearly demonstrate our idea, we introduce our algorithm based on the standard FCM (Euclidean distance) in this paper. In this case, the distance function in standard FCM can be defined as

$$d_{ij} = \sum_{k=1}^{K} v_{ijk}^n \bar{d}_{ijk}. \quad (23)$$

where $\bar{d}_{ijk}$ is the sub-distance function and v_{ijk} is the sub-membership which satisfies $\sum_{k=1}^{K} v_{ijk} = 1$. It is noted that the exact expression of sub-membership is $v_{ik|j}$ which represents the sub-cluster k belongs to cluster j. Here, we use v_{ijk} instead of $v_{ik|j}$ for simple expression.

In fact, our HFCM can also be considered as a two-level FCM: (1) the data are generated by J clusters in the first level and (2) within each cluster j; the data are generated by class-labeled sources that form sub-clusters of the large cluster in the second level. Substitution of (23) into (4), the objective function of HFCM is given as

$$J_{mn} = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} u_{ij}^{m} v_{ijk}^{n} \bar{d}_{ijk}. \quad (24)$$

It can be seen that at the second level of the hierarchy, information is provided about the data along with their class labels. It is noted that we have J classes; K sub-classes correspond to each class j of the first level.

The objective function J_{mn} can be minimized in a fashion similar to the standard FCM algorithm. Taking the first derivatives of J_{mn} with respect to u_{ij}, v_{ijk} and μ_{jk} and then setting them to zero results in three necessary but not sufficient conditions for J_{mn} to be at a local extreme. Let us first consider the derivation of the fuzzy membership (and sub-membership) function values. This can be obtained by minimizing the objective function J_{mn} over u_{ij} and v_{ijk} under the constraints $\sum_{j=1}^{J} u_{ij} = 1$ and $\sum_{k=1}^{K} v_{ijk} = 1$. Then, we have

$$J'_{mn} = J_{mn} + \alpha\left(1 - \sum_{j=1}^{J} u_{ij}\right) + \beta\left(1 - \sum_{k=1}^{K} v_{ijk}\right).$$

Taking the derivative of J'_{mn} with respect to u_{ij} and setting the result to zero, we have

$$u_{ij} = \left(\sum_{k=1}^{K} v_{ijk}^{n}\bar{d}_{ijk}\right)^{1/(1-m)} \Bigg/ \sum_{h=1}^{J}\left(\sum_{k=1}^{K} v_{ihk}^{n}\bar{d}_{ihk}\right)^{1/(1-m)}. \quad (25)$$

Similar processing on sub-membership v_{ijk}, we have

$$v_{ijk} = \left(u_{ij}^{m}\bar{d}_{ijk}\right)^{1/(1-n)} \Bigg/ \sum_{h=1}^{K}\left(u_{ij}^{m}\bar{d}_{ijh}\right)^{1/(1-n)}. \quad (26)$$

According to previous analysis, our algorithm is general enough and can be applied to various distance functions. To clearly demonstrate our idea, we still adopt Euclidean distance in this paper. Thus, we define $\bar{d}_{ijk} = \left\|y_i - \mu_{jk}\right\|^2$. To minimize the objective function by solving $\partial J_{mn}/\partial\mu_{jk} = 0$, we have

$$\mu_{jk} = \sum_{i=1}^{N} u_{ij}^{m} v_{ijk}^{n} y_i \Bigg/ \sum_{i=1}^{N} u_{ij}^{m} v_{ijk}^{n}. \quad (27)$$

It is noticed that according to (23), each distance function exhibits a FCM form. This suggests that we can contrast the proposed HFCM against the traditional FCM. In addition, the HFCM with J components in its first level can be considered as an extension of the standard FCM that employs J components in total. The proposed HFCM degrades to standard FCM when we set the sub-component k equal to 1 and sub-membership $v_{ijk} = 1$. Particularly, in standard FCM, all data points belong to component j depend on there distance function d_{ij}. In contrast, the HFCM assumes that the data generated by the component j are explained in a way that depends on their sub-component labels k (for each component j, a different distance function $\bar{d}_{ijk}$ is provided). The HFCM can also be derived to a standard FCM with KJ components in total which each distance function is modeled by a FCM with J components (parameters K and J are exchangeable).

Let us take two clustering examples to demonstrate the effectiveness of HFCM. In the first example, the synthetic data consists of 1800 data point from two Gaussian components where each component contains 900 samples. These two Gaussian distributions have the following parameters (mean μ_j and covariance matrices Σ_j): $\mu_1 = (0, 5)^{\mathrm{T}}$, $\mu_2 = (0, -5)^{\mathrm{T}}$, $\Sigma_1 = \Sigma_2 = \mathrm{diag}(2, 2)$. The test data is further augmented by 200 outliers, which are drawn from a Gaussian distribution, located in the middle-left of the test data. The test data and outliers are shown in Fig. 1(a). The solution by FCM and HFCM are shown in Fig. 1(b), drawn by blue ellipse

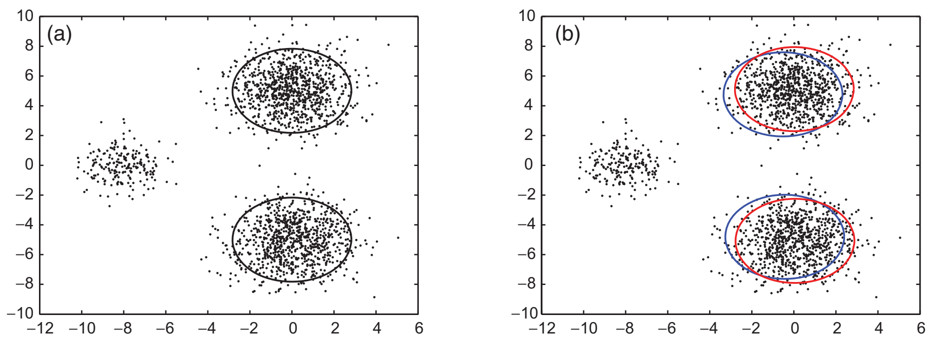

Fig. 1. Classification for the synthetic data by two different methods. (a) Original data distribution (black ellipse) (b) The solution by FCM (blue ellipse); the solution by HFCM (red ellipse).

Table 1
Estimated center for two different methods

	FCM	HFCM
Component 1	[−0.52 4.76]	[0.02 5.11]
Component 2	[−0.43 −4.81]	[0.05 −5.08]

and red ellipse, respectively. It can be seen that outliers pull estimated cluster center away from the true cluster center in the standard FCM. However, HFCM seems to fit the test data well and be not affected by outliers. The estimated center by FCM and HFCM are given in Table 1. It can be seen that HFCM estimates more accurate center than FCM.

In the second example, the synthetic data consists of 1800 data point from three Gaussian components where each component contains 600 samples. These three Gaussian distributions have the following parameters: $\mu_1=(0,\ 0)^{\mathrm{T}}$, $\mu_2=(3,\ 2)^{\mathrm{T}}$, $\mu_3=(0,\ 5)^{\mathrm{T}}$, $\Sigma_1=\Sigma_3=\mathrm{diag}(1/2,\ 1/2)$, $\Sigma_2=\mathrm{diag}(1/8,\ 1/8)$. The test data is further augmented by 1800 outliers, which are drawn from a uniform distribution on the interval $[-6,\ 6]$. The test data and outliers are shown in Fig. 2(a). The three clusters are shown by green stars (first component), red diamonds (second component) and blue circles (third component). The outliers are shown by black points. The classification results by FCM and HFCM are shown in Fig. 2(b) and (c) respectively. From Fig. 2(b), it can be seen that FCM can not classify them well. Some green, red and blue points are mixed together in the center of the figure due to the effect of the outliers. FCM also misclassify some red points which should belong to the green one (first component). However, HFCM obtains nearly perfect results shown in Fig. 2(c). The misclassification ratio (MCR) is 5.33% for FCM compared to 0.72% for HFCM. Thus, we can conclude that the additional flexibility in HFCM resolves a serious data representation drawback of the standard FCM and improves classification performance significantly.

5. Generalized hierarchical FCM

It is easy to combine HFCM and GFCM to generate a new generalized hierarchical FCM (GHFCM). GHFCM combines the merits of GFCM and HFCM to take into account more local spatial information, more flexible and general distance function. Substituting (23) into (9), the objective function of GHFCM can be given as

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} u_{ij}^m v_{ijk}^n \sum_{c\in\mathcal{N}_i} w_c d_{cjk} = \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{c\in\mathcal{N}_i} u_{ij}^m v_{ijk}^n w_c \left\|y_c-\mu_{jk}\right\|^2. \quad (28)$$

where w_c is the weighted factor defined in (9). The membership u_{ij} and sub-membership v_{ijk} satisfies the constraint $\sum_{j=1}^{J} u_{ij}=1$ and $\sum_{k=1}^{K} v_{ijk}=1$, respectively. In this paper, the sub-distance function d_{cjk} is simply selected as Euclidean distance. In fact, it can be easily extended to many other complex distance functions. By applying the optimization way similar to the standard FCM, the parameters in GHFCM can be calculated iteratively as

$$u_{ij} = \left(\sum_{k=1}^{K}\sum_{c\in\mathcal{N}_i} w_c v_{ijk}^n d_{cjk}\right)^{1/(1-m)} \Bigg/ \sum_{h=1}^{J}\left(\sum_{k=1}^{K}\sum_{c\in\mathcal{N}_i} w_c v_{ihk}^n d_{cjk}\right)^{1/(1-m)}. \quad (29)$$

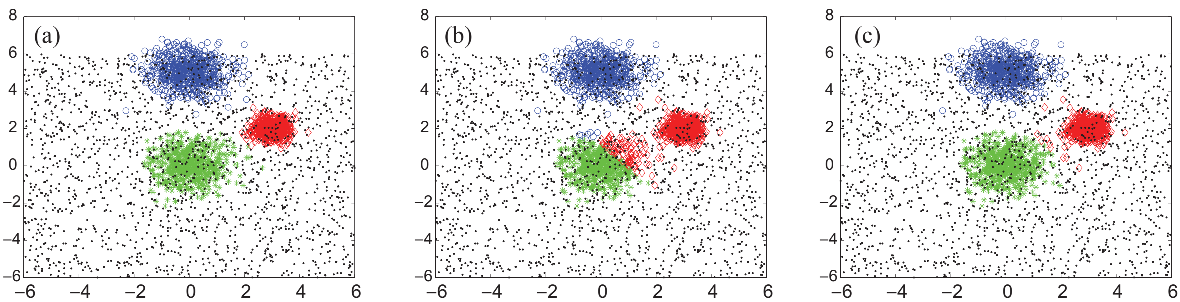

Fig. 2. Classification for the three-class data by two different methods. (a) Original data with outliers. (b) The solution by FCM, misclassification ratio is 5.33%. (c) The solution by HFCM, misclassification ratio is 0.72%.

$$v_{ijk} = \left(\sum_{c \in \mathcal{N}_i} w_c u_{ij}^m d_{cjk} \right)^{1/(1-n)} \Bigg/ \sum_{h=1}^{K} \left(\sum_{c \in \mathcal{N}_i} w_c u_{ij}^m d_{cjk} \right)^{1/(1-n)}. \tag{30}$$

The cluster center μ_{jk} is evaluated as

$$\mu_{jk} = \sum_{i=1}^{N} \sum_{c \in \mathcal{N}_i} u_{ij}^m v_{ijk}^n y_c / \sum_{i=1}^{N} u_{ij}^m v_{ijk}^n. \tag{31}$$

By applying local weighted generalized mean to incorporate spatial information and cluster information, the modified membership u_{ij} and sub-membership v_{ijk} can be re-calculated as

$$u_{ij} = \sum_{c \in \mathcal{N}_i} w_c u_{cj} / \sum_{h=1}^{J} \sum_{c \in \mathcal{N}_i} w_c u_{ch}. \tag{32}$$

$$v_{ijk} = \sum_{c \in \mathcal{N}_i} w_c v_{cjk} / \sum_{h=1}^{K} \sum_{c \in \mathcal{N}_i} w_c v_{cjh}. \tag{33}$$

GHFCM Algorithm:
Step 1. Fix the cluster number J, the sub-cluster number K, initialize fuzzy membership, sub-membership and then select initial cluster center.
Step 2. Set the loop counter $l = 0$.
Step 3. Update the new cluster center using (31).
Step 4. Update the fuzzy membership function using (29) and (32).
Step 5. Update the fuzzy membership function using (30) and (33).
Step 6. Terminate the iterations if the object function converges; otherwise, increase the iteration ($l = l+1$) and repeat steps 3 through 6.

6. Experimental results and discussion

In this section, we experimentally evaluate our proposed GHFCM in a set of synthetic images and real images. The neighborhood window size of GHFCM is set as 5×5. The fuzzy membership weighting is set $m = 2$. The sub-class is set $k = 2$ in GHFCM. We also evaluate GGFCM [2], MFCM [4], FCM_S [6], FLICM [3], FRFCM [29] and HMRF-FCM [12] for experimental comparison. The compared methods are based on different types of models, and were published in different journals recently; thus it is more significant to compare our algorithm with these state-of-the-art technologies. Our experiments have been developed in Matlab R2009b, and are executed on an Intel Pentium Dual-Core 2.2 GHZ CPU, 2 G RAM.

6.1. Synthetic images

In the first experiment, a three-class synthetic image (246×246, shown in Fig. 3(a)) is used to compare the performance of the proposed method with others. Figure 3(b) and (c) show the same image corrupted by Gaussian noise with zero mean and 0.12 variance, and by Speckle noise with noise density = 0.18, respectively. In order to evaluate the segmentation results, we employ the misclassification ratio (MCR) [15] in our experiments. The value of MCR is in the [0%–100%] range, where lower values indicate better segmentation performance. The segmentation results of the Gaussian noised image (Fig. 3(b)) by GGFCM, MFCM, FCM_S, FLICM, HMRF-FCM, FRFCM and the proposed methods are shown in Fig. 4(a) through (g). The class number J is set to 3, based on previous experience. As we observe, GGFCM, MFCM and FCM_S do not segment images well. Although FLICM, HMRF-FCM and FRFCM can reduce the effect of noise to some extent, they are still sensitive to heavy noise and misclassify some portions of pixels, as shown in Fig. 4(d), (e) and (g). However, we observe that the proposed GHFCM yields outstanding segmentation results compared to the poor performance of their competitors, as seen in Fig. 4(f). We also segment Speckle noised image (Fig. 3(c)) to evaluate the performance of various methods. Similar results are obtained and shown in Fig. 5(a) through (g). The results obtained by different noise densities are given in Table 2. As we observe, the proposed methods obtains the best results compared to the other methods, and especially for heavy noised image segmentation.

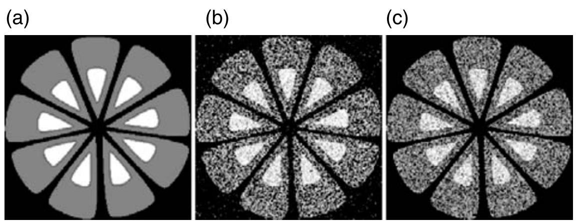

Fig. 3. (a) Original three-class image (b) Corrupted by Gaussian noise (zero mean, 0.12 variance) (c) Corrupted by Speckle noise (noise density = 0.18).

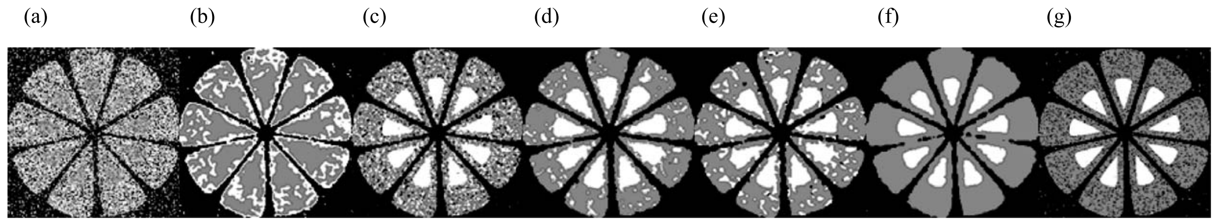

Fig. 4. Segment synthetic image with Gaussian noise. (a) GGFCM, MCR = 54.69% (b) MFCM, MCR = 29.7% (c) FCM_S, MCR = 13.11% (d) FLICM, MCR = 9.39% (e) HMRF-FCM, MCR = 11.3% (f) GHFCM, MCR = 3.66% (g) FRFCM, MCR = 5.58%.

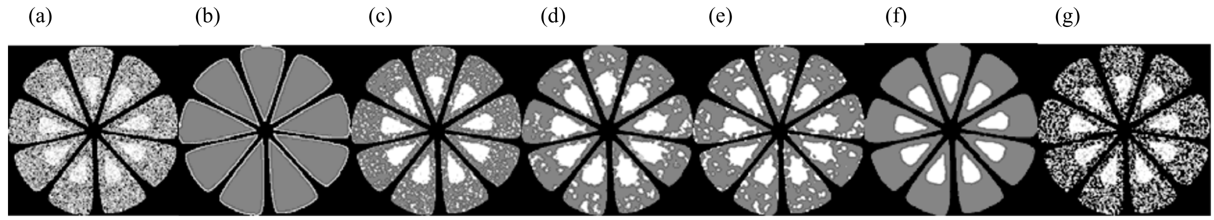

Fig. 5. Segment synthetic image with Speckle noise. (a) GGFCM, MCR = 28.57% (b) MFCM, MCR = 15.05% (c) FCM_S, MCR = 5.48% (d) FLICM, MCR = 9.05% (e) HMRF-FCM, MCR = 7.94% (f) GHFCM, MCR = 2.36% (g) FRFCM, MCR = 30.09%.

Table 2
The misclassification ratio (MCR %) of synthetic image with additive Gaussian noise for different methods

	GGFCM	MFCM	FCM_S	FLICM	HMRF-FCM	GHFCM	FRFCM
(G) noise = 0.06	28.88	15.57	3.02	3.03	4.25	2.51	3.62
(G) noise = 0.08	33.84	16.26	5.73	5.62	4.49	2.77	3.93
(G) noise = 0.10	36.99	22.77	9.12	8.82	6.17	3.14	5.42
(G) noise = 0.12	54.69	29.7	13.11	9.39	11.30	3.66	5.58
(S) speckle = 0.12	22.31	14.83	1.14	3.06	3.61	1.81	24.49
(S) speckle = 0.14	24.96	14.87	1.82	4.67	6.02	1.91	24.67
(S) speckle = 0.16	26.70	14.93	2.98	8.90	7.11	2.12	25.29
(S) speckle = 0.18	28.57	15.05	5.48	9.05	7.94	2.36	30.09
Average computation time	2.43s	1.03s	1.53s	22.79s	75.18s	1.84s	2.64

We also evaluate the computation time for all methods in the previous experiment. The average computation time of the different methods is presented on the last line of Table 2. It is noted that the computation of our method, GHFCM, is slower than MFCM and FCM_S, but is still much faster than other methods. Compared to other methods, our algorithm can be calculated more quickly and achieves the best segmentation result.

6.2. *Real images*

In this experiment, we evaluate the performance of the proposed method based on a subset of the Berkeley image dataset [25], which is comprised of a set of real-world color images along with segmentation maps provided by different individuals. We employ the Probabilistic Rand (PR) index to evaluate the performance of the proposed method, with the multiple ground truths available for each image within the dataset. It has been shown that the PR index possesses the desirable property of being robust to segmentation maps that result from splitting or merging segments of the ground truth [26]. The PR index takes values between 0 and 1, with values closer to 0 (indicating an inferior segmentation result) and values closer to 1 (indicating a better result).

Figure 6 shows the original Berkeley images used for the image segmentation experiment. These images with and without Gaussian noise are segmented by the proposed method, illustrated in Fig. 7. For fair comparison, we also evaluate the performance of GGFCM, FCM_S, FLICM, FRFCM and HMRF-FCM in addition to our methods. Table 3 presents the average PR values for all methods, corresponding to each of the test images in Fig. 6. Compared to other methods, the proposed algorithm yields the best segmentation results with the highest PR values.

We also evaluate the computation time for different methods in this experiment. The average computation

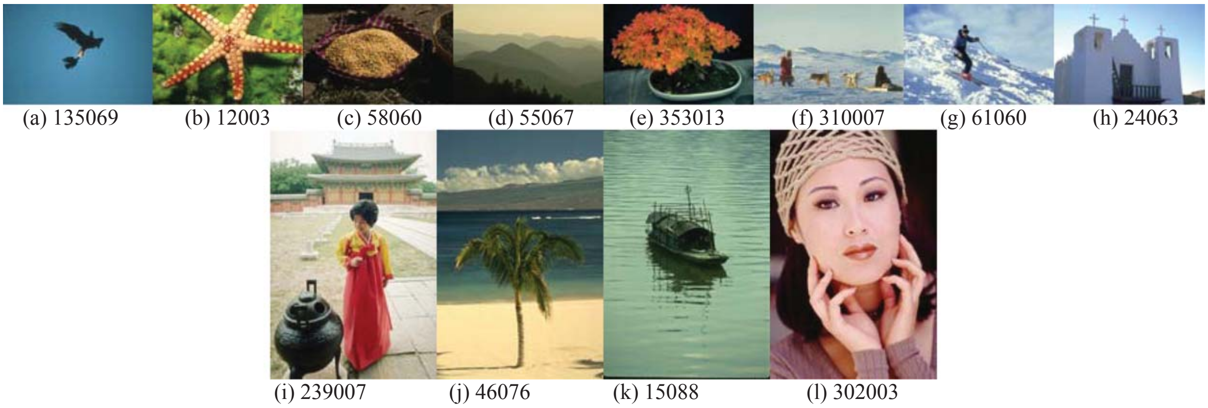

Fig. 6. Original image from the Berkeley image segmentation dataset.

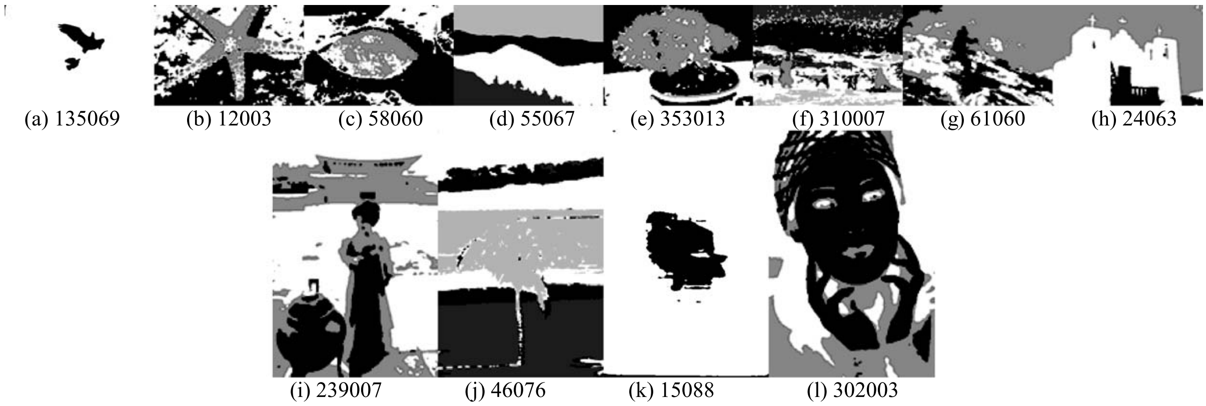

Fig. 7. Image segmentation results by GHFCM.

Table 3
Comparison of different methods for Berkeley image dataset, Probabilistic Rand (PR) Index

Image #	Class	GGFCM	FCM_S	FLICM	HMRF-FCM	GHFCM	FRFCM
135069	2	0.985	0.981	0.983	0.984	0.985	0.983
124084	3	0.708	0.510	0.510	0.526	0.756	0.499
69020	3	0.567	0.535	0.552	0.559	0.572	0.489
12003	3	0.622	0.608	0.614	0.618	0.673	0.576
58060	3	0.570	0.573	0.584	0.615	0.607	0.584
239007	3	0.665	0.633	0.645	0.668	0.659	0.658
46076	4	0.808	0.715	0.725	0.826	0.828	0.815
55067	4	0.887	0.879	0.879	0.888	0.890	0.870
353013 + 0.01 noise	3	0.702	0.633	0.663	0.741	0.751	0.661
310007 + 0.01 noise	7	0.620	0.664	0.708	0.677	0.709	0.640
61060 + 0.01 noise	3	0.546	0.617	0.625	0.575	0.650	0.627
15088 + 0.02 noise	2	0.698	0.656	0.717	0.855	0.864	0.835
24063 + 0.02 noise	3	0.777	0.819	0.826	0.834	0.840	0.766
374067 + 0.02 noise	4	0.695	0.711	0.729	0.744	0.761	0.658
302003 + 0.02 noise	3	0.704	0.705	0.713	0.715	0.718	0.708
Mean		0.704	0.683	0.698	0.722	0.751	0.691
Average computation time		9.97s	4.77s	56.71s	78.64s	9.13s	8.53s

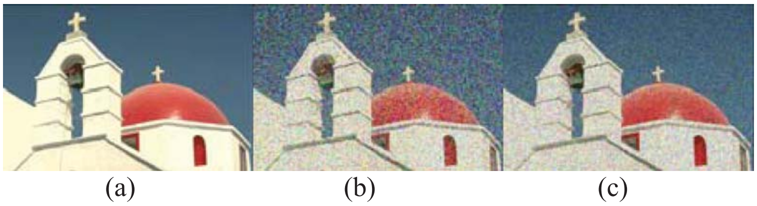

Fig. 8. (a) Original RGB image (b) Corrupted by Gaussian noise (zero mean, 0.15 variance) (c) Corrupted by Speckle noise (noise density = 0.15).

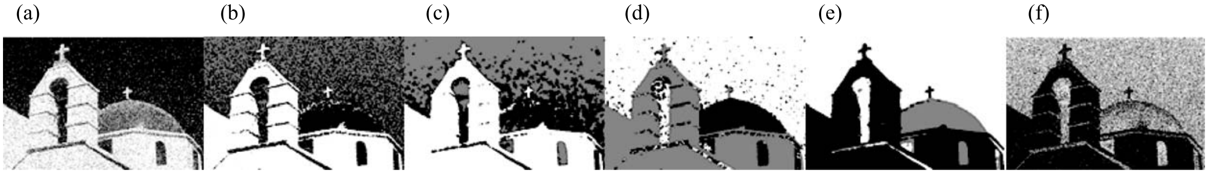

Fig. 9. Segment RGB Image with Gaussian noise. (a) GGFCM, PR = 0.667, t = 9.28 s (b) FCM_S, PR = 0.759, t = 4.23 s (c) FLICM, PR = 0.779, t = 28.18 s (d) HMRF-FCM, PR = 0.843, t = 167.53 s (e) GHFCM, PR = 0.851, t = 8.83 s (f) FRFCM, PR = 0.6167, t = 9.22 s.

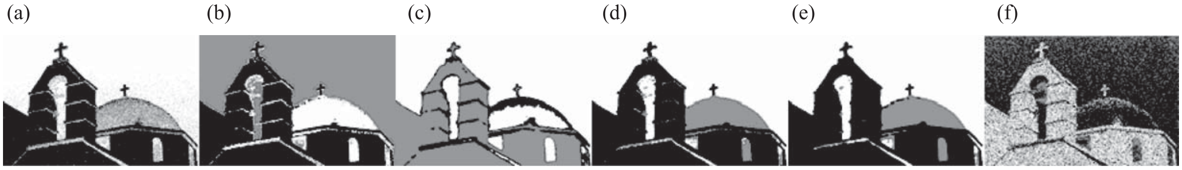

Fig. 10. Segment RGB Image with Speckle noise. (a) GGFCM, PR = 0.789, t = 8.63 s (b) FCM_S, PR = 0.849, t = 4.19 s (c) FLICM, PR = 0.796, t = 28.09 s (d) HMRF-FCM, PR = 0.85, t = 159.09 s (e) GHFCM, PR = 0.851, t = 8.43 s (f) FRFCM, PR = 0.6664, t = 10.09 s.

time of the different methods is presented on the last line of Table 3. It is noted that the computation of our method, GHFCM, is slower than FCM_S and FRFCM, but is still much faster than other methods. It is noted that MFCM is invalid for this RGB image segmentation application. Compared to other methods, our algorithm can be calculated more quickly and achieves the best segmentation result.

6.3. Multidimensional images

In this experiment, we try to segment the multidimensional RGB color image into three classes: the blue sky, the red roof and the white wall. The original image (481×321) shown in Fig. 8(a) is corrupted by heavy Gaussian noise with mean = 0 and covariance = 0.15, and by heavy Speckle noise with noise density = 0.15. The Gaussian noised image is shown in Fig. 8(b) and the segmentation results of GGFCM, FCM_S, FLICM, HMRF-FCM, FRFCM and our proposed methods are shown in Fig. 9(a) through (f), respectively. The accuracy of segmentation for GGFCM and FRFCM is quite poor. This is expected that no spatial constraints are taken into account in GGFCM. FCM_S, FLICM and HMRF-FCM obtain better results, but they are still sensitive to heavy noise. The accuracy of the segmentation results from GHFCM, as shown in Fig. 9(e), is better than that of other methods, obtaining the highest PR values. It is worth pointing out that the PR value of HMRF-FCM (0.843) is just a little lower than the PR value of GHFCM (0.851); however, the segmentation result of HMRF-FCM, shown in Fig. 9(d), is much poorer than GHFCM (Fig. 9(e)). We also segment speckle noised image, shown in Fig. 8(c), to evaluate the performance of various methods. Similar results are obtained and shown in Fig. 10(a) through (f). We can observe that the proposed GHFCM again obtains the best segmentation results.

We also evaluate the computation time for all methods in the previous experiment. The computation time t of the different methods is also presented in Figs. 9 and 10. It is noted that the computation of our methods is much faster than that of other methods except for FCM_S methods. Compared to other methods, our models can be calculated more quickly and achieve the best segmentation results.

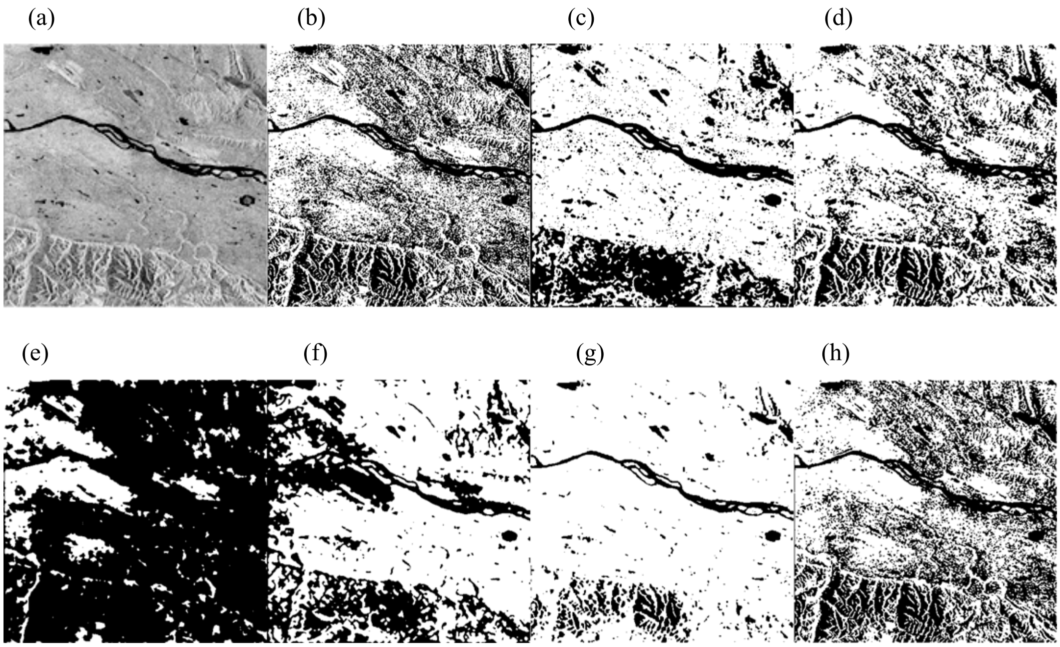

Fig. 11. SAR Image Segmentation. (a) Original SAR image (b) GGFCM (c) MFCM (d) FCM_S (e) FLICM (f) HMRF-FCM (g) GHFCM (h) FRFCM.

6.4. SAR images

Synthetic aperture radar (SAR) data are often affected by speckle noise, which originates in the SAR system's coherent nature. In this experiment, we evaluate various methods based on real RADARSAT-1 SAR image which is shown in Fig. 11(a). RADARSAT-1 is a sophisticated Earth observation satellite developed by Canada to monitor environmental changes and the planet's natural resources.

Our purpose is to distinguish the Mountains and the River. Thus we set the component number $J = 2$. The segmentation results of GGFCM, MFCM, FCM_S, FLICM, HMRF-FCM, FRFCM and our proposed methods are shown in Fig. 11(b) through (h), respectively. Existence of noise has led to a "spotty" result using GGFCM, FCM_S, and FRFCM shown in Fig. 11(b), (d) and (h), respectively. FLICM and HMRF-FCM are failed to segment the SAR image, shown in Fig. 11(e) and (f), respectively. FLICM mixed the mountain and the river together and HMRF-FCM misclassifies some river parts. The MFCM shown in Fig. 11(c) gives an improved result; however, it loses some image details and is still not robust enough to image speckle. From Fig. 11(g), it can be seen that the proposed GHFCM obviously overcomes these shortcomings and preserve more details of mountain's ridges and river's tributaries. Through discussion above, we can conclude that the proposed GHFCM is more robust to image speckle and can preserve more image details simultaneously.

7. Conclusions

In this paper, we propose a novel simple and effective fuzzy clustering approach for image segmentation. The GHFCM is introduced by incorporating the hierarchical distance function and spatial constraints into the fuzzy objective function. Thus, our model can be considered as an extension of standard FCM and can degrade to FCM by setting the proper parameters. Although the idea of generalized mean and hierarchical model is adopted for standard FCM, it can be easily extended to many other modified FCM models. Compared to the standard FCM, HFCM improves data representation in sub-distance relevant to classification. Moreover, HFCM exhibits robustness with respect to the subcluster number K at the second level, and this constitutes a great advantage over the standard FCM. In GFCM, the distance function of an image pixel is influenced by the distance function of pixels in its immediate neighborhood with the help of the generalized mean.

Acknowledgments

This work was supported in part by the Canada Chair Research Program and the Natural Sciences and Engineering Research Council of Canada. This work was supported in part by the National Natural Science Foundation of China under Grant 61402235, 61105007 and 61103141. This work was supported in part by the National Research Foundation of Korea (Grant NRF-2013K2A2A2000777). This work was supported in part by the Natural Science Foundation of Jiangsu Province BK2012858 and BK2011825. This work was supported in part by PAPD (A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions). This work was supported in part by the Innovation Fund of China Academy of Space Technology: CAST201302. H. Zhang is the corresponding author of this paper.

References

[1] J. Yu, General C-means clustering model, *IEEE Trans Pattern Anal Mach Intell* **27**(8) (2005), 1197–1211.

[2] N.B. Karayiannis, Weighted fuzzy learning vector quantization and weighted generalized fuzzy c-means algorithms, *Proc IEEE Conf Fuzzy Systems* (1996), 773–779.

[3] S. Krinidis and V. Chatzis, A robust fuzzy local information C-means clustering algorithm, *IEEE Trans Image Process* **5**(19) (2010), 1328–1337.

[4] 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* **21** (2002), 193–199.

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

[6] S. Chen and D. Zhang, Robust image segmentation using FCM with spatial constraints based on new Kernel-induced distance measure, *IEEE Transactions on Systems, Man and Cybernetics* **34**(4) (2004), 1907–1916.

[7] D. Pham and J.L. Prince, An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, *Pattern Recognit Lett* **20** (1999), 57–68.

[8] P. Kersten, Fuzzy order statistics and their application to fuzzy clustering, *IEEE Trans Fuzzy Syst* **7**(6) (1999), 708–712.

[9] L. Bobrowski and J. Bezdek, C-means clustering with the l1 and l_∞ norms, *IEEE Trans Syst, Man, Cybern* **21**(3) (1991), 545–554.

[10] R.J. Hathaway, J.C. Bezdek and Y. Hu, Generalized fuzzy c-means clustering strategies using Lp norm distances, *IEEE Trans Fuzzy Syst* **8**(5) (2000), 576–582.

[11] F. Klawonn and A. Keller, Fuzzy clustering based on modified distance measures, *in Proc 3rd Int Symp Adv Intell Data Anal* (1999), 291–302.

[12] S.P. Chatzis and T.A. Varvarigou, A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation, *IEEE Trans Fuzzy Syst* **16**(5) (2008), 1351–1361.

[13] P. Clifford, Markov Random Fields in Statistics, in Disorder in Physical Systems. A Volume in Honour of John M. Hammersley on the Occasion of His 70th Birthday, G. Grimmett and D. Welsh, Eds. Oxford, U.K.: Clarendon Press, Oxford Science Publication, 1990.

[14] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, *Proc IEEE* **77**(2) (1989), 257–286.

[15] Y. Zhang, M. Brady and S. Smith, Segmentation of brain MR images through a hidden markov random field model and the expectation maximization algorithm, *IEEE Trans Med Imag* **20**(1) (2001), 45–57.

[16] J. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters, *Journal of Cybernetics* **3** (1974), 32–57.

[17] K.S. Chuang, H.L. Tzeng, S. Chen, J. Wu and T.J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, *Comput Med Imag Graph* **30**(1) (2006), 9–15.

[18] W. Pedrycz, Conditional fuzzy clustering in the design of radial basis function neural network, *IEEE Trans Neural Netw* **9** (1998), 601–612.

[19] C. Wei and C. Fahn, The multisynapse neural network and its application to fuzzy clustering, *IEEE Trans Neural Netw* **13**(3) (2002), 600–618.

[20] M.S. Yang, On a class of fuzzy classification maximum likelihood procedures, *Fuzzy Sets Syst* **57** (1993), 365–375.

[21] M. Jordan and R. Jacobs, Hierarchical mixtures of experts and the EM algorithm, *in Proc Int Joint Conf Neural Netw* **2** (1993), 1339–1344.

[22] M.K. Titsias and A. Likas, Mixture of experts classification using a hierarchical mixture model, *Neural Comput* **14**(9) (2002), 2221–2244.

[23] P. Orbanz and J.M. Buhmann, SAR images as mixtures of gaussian mixtures, *IEEE Int'l Conf, Image Processing* **2** (2005), 209–212.

[24] M. Di Zio, U. Guarnera and R. Rocci, A mixture of mixture models for a classification problem: The unity measure error, *Computational Statistics and Data Analysis* **51**(5) (2007), 2573–2585.

[25] D. Martin, C. Fowlkes, D. Tal and J. Malik, A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics, *in Proc 8th IEEE Int Conf Comput Vis*, Vancouver, BC, Canada, **2** (2001), 416–423.

[26] R. Unnikrishnan and M. Hebert, Measures of similarity, *in Proc IEEE Workshop Comput Vis Appl* (2005), 394–400.

[27] U. Qamar, A dissimilarity measure based Fuzzy c-means (FCM) clustering algorithm, *Journal of Intelligent and Fuzzy Systems* **26**(1) (2014), 229–238.

[28] J.-Y. Leu, K.-K. Huang and Y.-M. Huang, Two novel fuzzy clustering methods for solving data clustering problems, *Journal of Intelligent and Fuzzy Systems* **26**(2) (2014), 705–719.

[29] W.l. Cai, S.C. Chen and D.Q. Zhang, Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation, *Pattern Recognit* **40**(3) (2007), 825–838.