Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (13): 2574-2578, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.2574.2578

A Fast Kernel-induced Fuzzy C-means Algorithm and its Application to Segmentation of Microscopic Image of Harmful Algae

[1,2]Wenjuan Zhang and [3]Jiayin Kang
[1]School of Computer Engineering,
[2]Jiangsu Institute of Marine Resources,
[3]School of Electronics Engineering, Huaihai Institute of Technology, Lianyungang, 222005, China

Abstract: Segmentation of microscopic image of harmful algae is a crucial step in harmful algae classification and recognition system. FCM is a powerful tool for image segmentation and has been applied into various applications successfully. However, traditional FCM algorithm is sensitive to the noise due to the fact that it only accounts for the pixel's value information, not takes the neighboring pixels' spatial information into consideration. Furthermore, the performance of traditional FCM becomes poor when the input data are nonlinearly. In order to overcome the shortcomings of traditional FCM, this study presented a modified Fuzzy C-means (FCM) algorithm for image segmentation. The proposed method is realized by modifying the objective function in the Szilagyi's algorithm via introducing histogram-based weight. Furthermore, the kernel method was introduced to replace the original Euclidean distance in the Szilagyi's algorithm. Experimental results on microscopic images of harmful algae show that the proposed approach has better performance compared with other methods.

Key words: Image segmentation, fuzzy C-means, histogram, harmful algae

INTRODUCTION

Fuzzy C-means (FCM) clustering algorithm, an unsupervised clustering technique, has been widely used in image segmentation since it was proposed (Zhang and Wang, 2006; Chen *et al*., 2006). Compared with hard C-means algorithm (Gorriz *et al*., 2006), FCM is able to preserve more information from the original image. However, first of all it is noise sensitive because of not taking into account the spatial information (Chuang *et al*., 2006); secondly it is suppose that each feature date has the same contribution to classifying results (Li *et al*., 2005); finally its performance becomes poor when the input data are nonlinearly. To solve the first problem, recently, many researchers proposed the algorithms accounting for spatial information via modifying the objective function of standard FCM algorithm (Chuang *et al*., 2006; Chen and Zhang, 2004). To solve the second problem, Li *et al*. (2005) proposed a modified clustering algorithm via introducing feature weight of the data. To solve the last problem, kernel method was introduced to replace the original Euclidean distance (Ma *et al*., 2007; Zhang and Chen, 2004).

Harmful algae occurs worldwide and is a major problem in the world in terms of aquatic ecosystem, human health and economy (Tran *et al*., 2010). Specifically, toxin accumulation in marine bivalves is a common phenomenon during algal blooming events that can lead to a closure of shellfish harvest for human consumption. It is reported that harmful algae blooms alone cost the United States 50 million dollars per year and sewage polluted waters adversely affect human health and the economy (LaGier *et al*., 2007). Therefore, monitoring of harmful algae automatically using microscopic image processing and recognition technology is crucial to restrict human access to contaminated waters and products. Image segmentation is one of the important steps in an image processing and analysis system of microscopic image of harmful algae.

Based on the above analysis, this study proposed an algorithm for segmentation of microscopic image of harmful algae using modified FCM combined with kernel method, KSW-FCM.

TRADITIONAL FCM ALGORITHM

The FCM algorithm assigns pixels to each category by using fuzzy memberships Let $X = \{x_i, i = 1, 2,..., n|x_i \in \mathbb{R}^d\}$ denotes an image with n pixels to be partitioned into c clusters, where x_i represents features data.

Corresponding Author: Jiayin Kang, School of Electronics Engineering, Huaihai Institute of Technology, Lianyungang, 222005, China

The algorithm is an iterative optimization that minimizes the objective function defined as follows:

$$J_m = \sum_{k=1}^{c}\sum_{i=1}^{n} u_{ki}^m \|x_i - v_k\|^2 \tag{1}$$

With the following constraints:

$$\{u_{ki} \in [0,1] | \sum_{k=1}^{c} u_{ki} = 1, \forall i, 0 < \sum_{i=1}^{n} u_{ki} < n, \forall k\} \tag{2}$$

where, u_{ki} represents the membership of pixel x_i in the k^{th} cluster, v_k is the k^{th} class center; $\|\cdot\|$ denotes the Euclidean distance, m>1 is a weighting exponent on each fuzzy membership. The parameter m controls the fuzziness of the resulting partition. The membership functions and cluster centers are updated by the following expressions:

$$u_{ki} = \sum_{l=1}^{c}\left(\frac{\|x_i - v_k\|}{\|x_i - v_l\|}\right)^{-2/(m-1)} \tag{3}$$

And:

$$v_k = \frac{\sum_{i=1}^{n} u_{ki}^m x_i}{\sum_{i=1}^{n} u_{ki}^m} \tag{4}$$

In implementation, matrix V is randomly initialize and then U and V are updated through an iterative process using Eq. 3 and 4, respectively.

SPATIALLY WEIGHTED FCM WITH KERNEL METHOD

Modified FCM algorithm: FCM clustering algorithm is a powerful tool for segmenting the images and has been applied successfully to the various applications. However, traditional FCM algorithm is sensitive to noise due to not taking spatial information into consideration. In order to overcome such drawback, many researchers have proposed the modified FCM algorithms. Szilagyi *et al.* (2003) proposed a fast FCM clustering algorithm which is called EnFCM, used for gray level image segmentation. The algorithm accounts for pixel spatial information. Before the algorithm implementation, a linearly weighted sum image ξ, composed by original image and local neighboring average of each pixel in original image, was calculated as follows:

$$\xi_i = \frac{1}{1+\alpha}\left(x_i + \frac{\alpha}{N_R}\sum_{j\in N_i} x_j\right) \tag{5}$$

where, ξ_i is the gray value of the ith pixel in the image ξ.N_i stands for the set of neighbors falling into a local window around x_i and N_R is its cardinality. The parameter α in the second term controls the effect of the penalty. In essence, the addition of the second term in Eq. 5, equivalently, formulates a spatial constraint and aims at keeping continuity on neighboring pixel values around x_i.

Accordingly, the modified objective function was described as follows:

$$J_s = \sum_{k=1}^{c}\sum_{l=1}^{q} \gamma_l u_{kl}^m \|\xi_l - v_k\|^2 \tag{6}$$

where, $\xi = \{\xi_l, l = 1, 2, k,..., q\}$ is the data set rearranging from he image ξ defined in Eq. 5 according to gray level. $v = \{v_k\}$ (k = 1, 2,..., c) represents the prototype of the k^{th} cluster, $U = \{u_{kl}\}$ (k = 1, 2, L, ..., c, l = 1, 2, L,..., q) represents the fuzzy membership of gray value l with respect to cluster k. q denotes the number of the gray level of the given image which is generally much smaller than N. γ_l is the number of the pixels having the gray value equal to l, where l = 1, 2, L,..., q. Naturally $\sum_{l=1}^{q}\gamma_l = N$

Similar to the standard FCM algorithm, under the constraints that:

$$\sum_{k=1}^{c} u_{kl} = 1$$

for any l, minimize J_s defined in Eq. 6. Specifically, taking the first derivatives of J_s with respect to u_{kl} and v_k and zeroing them, respectively, two necessary but not sufficient conditions for J_s will be obtained as follows:

$$u_{kl} = \frac{(\xi_l - v_k)^{-2/(m-1)}}{\sum_{r=1}^{c}(\xi_l - v_r)^{-2/(m-1)}} \tag{7}$$

$$v_k = \frac{\sum_{l=1}^{q}\gamma_l u_{kl}^m \xi_l}{\sum_{l=1}^{q}\gamma_l u_{kl}^m} \tag{8}$$

Obviously, in Eq. 6, gray level was viewed as the classified data. Hence, the number of classified data only depends on gray level and doesn't enlarge with the increasing of image size. However, Eq. 6 doesn't take different gray level has different influence on classifying results into consideration, i.e., Eq. 6 consider that every gray level has the same contribution to the classifying results. Actually, according to the gray level histogram of the image it is clear that the occurrence frequencies of different gray level are different. Therefore, different gray level has different contribution to clustering results. Based on above analysis, we modified the objective function in Eq. 6 as follows:

$$J_s = \sum_{k=1}^{c}\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m \|\xi_l - v_k\|^2 \quad (9)$$

where, w_l is the weighting coefficient of ξ_l, (l = 1, 2, ..., q) and can be computed via histogram as follows:

$$w_l = \frac{\gamma_l}{N}, \quad l = 0, 1, \cdots, q \quad (10)$$

where, q denotes the number of the gray level of the given image. γ_l is the number of the pixels having the gray value equal to l, where l = 1, 2,...q. Naturally:

$$\sum_{l=1}^{q} \gamma_l = N, \sum_{l=1}^{q} w_l = 1$$

i.e., w_l (l = 1, 2,..., q) can be viewed as the occurrence probability of each gray level. Hence, from Eq. 10 it is known that the weighting coefficient of each gray level can be given by the normalized image histogram.

Similarly, under the constraints that:

$$\sum_{k=1}^{c} u_{kl} = 1$$

for any l, minimize J_s defined in Eq. 9. Specifically, taking the first derivatives of J_s with respect to u_{kl} and v_k and zeroing them, respectively, two necessary but not sufficient conditions for J_s will be obtained as follows:

$$u_{kl} = \frac{(\xi_l - v_k)^{-2/(m-1)}}{\sum_{r=1}^{c} (\xi_l - v_r)^{-2/(m-1)}} \quad (11)$$

$$v_k = \frac{\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m \xi_l}{\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m} \quad (12)$$

From Eq. 12 it is known that the function of weighting coefficient w_l lies in adjusting the clustering center. Equation 12 will degenerated to Eq. 8 while $w_l = 1/q$.

The modified FCM algorithm (spatially weighting FCM clustering algorithm, called SWFCM) can be summarized as follows:

Step 1: Fix m>1 and 2≤c≤N-1; then select initial class prototypes v_k (k = 1, 2,...,c); set ε>0 to a very small value

Step 2: Compute the new image ξ in terms of Eq. 5 in advance

Repeat:

Step 3: Compute/modify μ_{kl} with μ_{kl} by Eq. 11 and 12

Step 4: Update v_k with the modified μ_{kl} by Eq. 12 Until $(|V_{new}-V_{old}|<\varepsilon)$

Modified FCM combined with kernel method: Though FCM has been applied to numerous clustering problems it still suffers from poor performance when boundaries among clusters in the input data are nonlinear. One alternative approach is to transform the input data into a feature space of a higher dimensionality using a nonlinear mapping function so that nonlinear problems in the input space can be linearly treated in the feature space. One of the most popular data transformation methods adopted in recent studies is the kernel method (Ma *et al*., 2007; Zhang and Chen, 2004). One of the advantageous features of the kernel method is that input data can be implicitly transformed into the feature space without knowledge of the mapping function. Furthermore, the dot product in the feature space can be calculated using a kernel function.

With the incorporation of the kernel method, the objective function of (1) in the feature space using the mapping function Φ can be rewritten as follow:

$$J_m^{\Phi} = \sum_{k=1}^{c}\sum_{i=1}^{N} u_{ki}^m \|\Phi(x_i) - \Phi(v_k)\|^2 \quad (13)$$

Through kernel substitution, the above objective function can be rewritten as:

$$J_m^{\Phi} = 2\sum_{k=1}^{c}\sum_{i=1}^{N} u_{ki}^m (1 - K(x_i, v_k)) \quad (14)$$

where, K (x_i, v_k) is function kernel which generally takes the Gaussian Radial Basis Function (GRBF) kernel with the following form:

$$K(x, y) = \exp\left(\frac{-\|x - y\|^2}{\sigma^2}\right) \quad (15)$$

By using the Lagrange multiplier to minimize the objective function (14), the membership functions can be updated as follow:

$$u_{ki} = \frac{(1 - K(x_i, v_k))^{-1/(m-1)}}{\sum_{l=1}^{c} (1 - K(x_i, v_l))^{-1/(m-1)}} \quad (16)$$

And the cluster centers can be updated as follow:

$$v_k = \frac{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k) x_i}{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k)} \quad (17)$$

Similar to the kernelized version of (1), we kernelized (9) and obtain the following new objective function through the kernel-induced distance measure substitution:

$$J_s^{\Phi} = \sum_{k=1}^{c}\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m (1 - K(\xi_l, v_k)) \quad (18)$$

where, K (ξ, v) is still taken as GRBF, w_l, γ_l and ξ_l are defined as (9).

Formally, the above optimization problem comes in the form:

$$\min_{U,\{v_l\}_{l=1}^{c}} J_s^{\Phi} \text{ subject to Eq. (2)} \quad (19)$$

In a similar way to the standard FCM algorithm, the objective function J^{Φ}_s can be minimized under the constraint of U as stated in (2) and then the membership functions and cluster centers are updated by the following expressions:

$$u_{kl} = \frac{(1 - K(\xi_l, v_k))^{-1/(m-1)}}{\sum_{r=1}^{c}(1 - K(\xi_l, v_r))^{-1/(m-1)}} \quad (20)$$

$$v_k = \frac{\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m (K(\xi_l, v_k)\xi_l)}{\sum_{l=1}^{q} w_l \gamma_l u_{kl}^m (K(\xi_l, v_k))} \quad (21)$$

The modified FCM, spatially weighted FCM with kernel method (KSW-FCM), can be summarized as follows:

Step 1: Fix m>1and 2≤c≤N-1; set ε>0 to a very small value. maximum iterative number n_{max}
Step 2: Initialize cluster centers $v_k^{(0)}$ (k = 1, 2, ..., c)
Step 3: Compute the new image ξ in terms of Eq. 5 in advance
Repeat:
Step 4: Compute/modify μ_{kl} with v_k by Eq. 20 and 21
Step 5: Update v_k with the modified μ_{ld} by Eq. 21 Until ($\|V^{(n+1)} - V^{(n)}\| \le \varepsilon$ or iterative number $n > n_{max}$)

EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we carried out some experiments on microscopic images of harmful algae to verify the performance of our proposed algorithm for image segmentation. All experimental data are taken from the web page: http://hab.jnu.edu.cn/index.asp.

Figure 1a shows an original microscopic image of harmful algae. Figure 1b-d display the segmentation results using traditional FCM algorithm, Szilagyi's algorithm (Szilagyi *et al*., 2003 and the proposed algorithm, respectively. The parameters' settings are c = 2, m = 2, a = 5, $\varepsilon = 10^{-5}$.

Fig. 2a shows another original microscopic image of harmful algae. Figure 2b-d display the segmentation

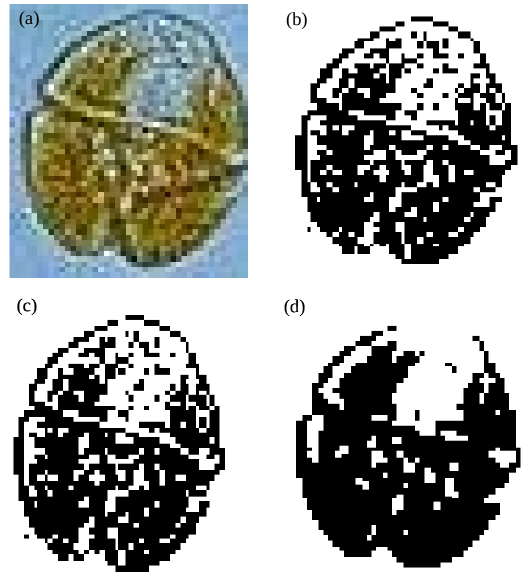

Fig. 1(a-d): Segmentation results of microscopic image of harmful algae, (a) Original image, (b) Segmentation using FCM, (c) Segmentation using Szilagyi's algorithm presented in Szilagyi *et al*. (2003) and (d) Segmentation using proposed method

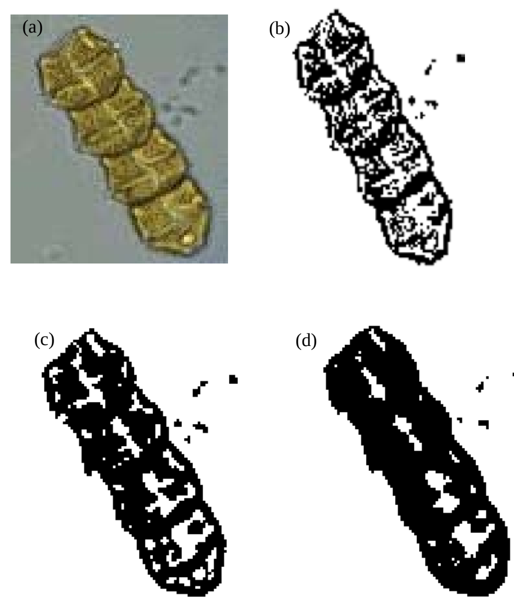

Fig. 2(a-d): Segmentation results of microscopic image of harmful algae, (a) Original image, (b) Segmentation using FCM, (c) Segmentation using szilagyi's algorithm presented in Szilagyi *et al*. (2003) and (d) Segmentation using proposed method results using traditional FCM algorithm, Szilagyi's algorithm and the proposed algorithm, respectively. The parameters' settings are c = 2, m = 2, a = 5, $\varepsilon = 10^{-5}$.

From Fig. 1 and 2, we can see that the segmentation effects of both the Szilagyi's algorithm and the KNW-FCM are better than that of traditional FCM algorithm. Furthermore, compared Fig. 1c with Fig. 1d and Fig. 2c with Fig. 2d, we know that the segmentation results obtained by the proposed method are more consistent and accurate than that of the Szilagyi's algorithm.

CONCLUSION

In this study, an automatic modified FCM clustering algorithm for image segmentation was proposed. The proposed algorithm is realized by modified the objective function in the Szilagyi's algorithm via introducing the gray histogram-based weighting and kernel method. Experimental results on microscopic images of harmful algae show that proposed method can dramatically improves the robustness to noise compared with traditional FCM and can achieves better denoising performance compared with Szilagyi's algorithm. Therefore, the proposed approach will be promising in real processing and recognition system of microscopic image of harmful algae.

ACKNOWLEDGMENT

This work is jointly funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, the open project of Jiangsu Institute of Marine Resources (No. HK13016). The experimental data are obtained from the Research Center for Harmful Algae and Aquatic Environment, Jinan University.

REFERENCES

Chen, S. and D. Zhang, 2004. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE Trans. Syst. Man Cybernet., 34: 1907-1916.

Chen, W., M.L. Giger and U. Bick, 2006. A fuzzy C-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. Acad. Radiol., 13: 63-72.

Chuang, K.S., H.L. Tzeng, S. Chen, J. Wu and T.J. Chen, 2006. Fuzzy c-means clustering with spatial information for image segmentation. Comput. Med. Imaging Graph., 30: 9-15.

Gorriz, J.M., J. Ramirez, E.W. Lang and C.G. Puntonet, 2006. Hard C-means clustering for voice activity detection. Speech Communi., 48: 1638-1649.

LaGier, M.J., J.W. Fell and K.D. Goodwin, 2007. Electrochemical detection of harmful algae and other microbial contaminants in coastal waters using hand-held biosensors. Mar. Pollut. Bull., 54: 757-770.

Li, J., X. Gao and L. Jiao, 2005. A new feature weighted fuzzy clustering algorithm. Acta Electron. Sin., 34: 89-92.

Ma, B., H.Y. Qu and H.S. Wong, 2007. Kernel clustering-based discriminant analysis. Pattern Recogn., 40: 324-327.

Szilagyi, L., Z. Benyo, S.M. Szilagyi and H.S. Adam, 2003. MR brain image segmentation using an enhanced fuzzy c-means algorithm. Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Volume 1, September 17-21, 2003, Cancun, Mexico, pp: 724-726.

Tran, D., H. Haberkorn, P. Soudant, P. Ciret and J.C. Massabuau, 2010. Behavioral responses of *Crassostrea gigas* exposed to the harmful algae *Alexandrium minutum*. Aquaculture, 298: 338-345.

Zhang, D.Q. and S.C. Chen, 2004. A novel kernelized fuzzy C-means algorithm with application in medical image segmentation. Artif. Intell. Med., 32: 37-50.

Zhang, D.B. and Y.N. Wang, 2006. Medical image segmentation based on FCM clustering and rough set. Chin. J. Sci. Instrument, 27: 1683-1687.