Artificial Intelligence in Medicine (2004) **32**, 37–50

ELSEVIER

ARTIFICIAL INTELLIGENCE IN MEDICINE

http://www.intl.elsevierhealth.com/journals/aiim

A novel kernelized fuzzy C-means algorithm with application in medical image segmentation

Dao-Qiang Zhang[a,b], **Song-Can Chen**[a,b,*]

[a]*Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China*
[b]*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, PR China*

Received 31 March 2003; received in revised form 26 October 2003; accepted 17 January 2004

KEYWORDS
Image segmentation; Fuzzy C-means; Kernel method; Kernel-induced distance; Magnetic resonance imaging

Summary Image segmentation plays a crucial role in many medical imaging applications. In this paper, we present a novel algorithm for fuzzy segmentation of magnetic resonance imaging (MRI) data. The algorithm is realized by modifying the objective function in the conventional fuzzy C-means (FCM) algorithm using a kernel-induced distance metric and a spatial penalty on the membership functions. Firstly, the original Euclidean distance in the FCM is replaced by a kernel-induced distance, and thus the corresponding algorithm is derived and called as the kernelized fuzzy C-means (KFCM) algorithm, which is shown to be more robust than FCM. Then a spatial penalty is added to the objective function in KFCM to compensate for the intensity inhomogeneities of MR image and to allow the labeling of a pixel to be influenced by its neighbors in the image. The penalty term acts as a regularizer and has a coefficient ranging from zero to one. Experimental results on both synthetic and real MR images show that the proposed algorithms have better performance when noise and other artifacts are present than the standard algorithms.

1. Introduction

With the increasing size and number of medical images, the use of computers in facilitating their processing and analyses has become necessary [1]. In particular, as a task of delineating anatomical structures and other regions of interest, image segmentation algorithms play a vital role in numerous biomedical imaging applications such as the quantification of tissue volumes, diagnosis, study of anatomical structure, and computer-integrated surgery [1–3]. Classically, image segmentation is defined as the partitioning of an image into non-overlapping, constituent regions which are homogeneous with respect to some characteristics such as intensity or texture.

Because of the advantages of magnetic resonance imaging (MRI) over other diagnostic imaging [2], the majority of researches in medical image segmentation pertains to its use for MR images, and there are a lot of methods available for MR image segmentation [2–6]. Among them, fuzzy segmentation methods are of considerable benefits, because they could retain much more information from the original image than hard segmentation methods [3]. In particular, the fuzzy C-means (FCM) algorithm [7], assign pixels to fuzzy clusters without labels. Unlike the hard clustering methods which force pixels to belong exclusively to one class, FCM allows pixels to belong to multiple clusters with varying

*Corresponding author. Tel.: +86-25-489-2805.
E-mail addresses: daoqz@mail.com, s.chen@nuaa.edu.cn (S.-C. Chen).

0933-3657/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.artmed.2004.01.012

degrees of membership. Because of the additional flexibility, FCM has been widely used in MR image segmentation applications recently. However, because of the spatial intensity inhomogeneity induced by the radio-frequency coil in MR image, conventional intensity-based FCM algorithm has proven to be problematic, even when advanced techniques such as non-parametric, multi-channel methods are used [2]. To deal with the inhomogeneity problem, many algorithms have been proposed by adding correction steps before segmenting the image [4,5] or by modeling the image as the product of the original image and a smooth varying multiplier field [2,6]. Recently, many researchers have incorporated spatial information into the original FCM algorithm to better segment the images [6,8–10]. Tolias and Panas [8] proposed a fuzzy rule-based system to impose spatial continuity on FCM, and in another paper [9], they used a small positive constant to modify the membership of the center pixel in a 3×3 window. Pham et al. [6] modified the objective function in the FCM algorithm to include a multiplier field containing the first- and second-order information of the image. Similarly, Ahmed et al. [11] proposed an algorithm to compensate for the intensity inhomogeneity and to label a pixel by considering its immediate neighborhood. A rather recent approach proposed by Pham [12] is to penalize the FCM objective function to constrain the behavior of the membership functions, similar to methods used in the regularization and Markov random field (MRF) theory.

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', e.g. SVM [13–15], KPCA [16] and KFD [17]. And this 'kernel method' has also been applied to unsupervised clustering [18–20]. However, a drawback of these kernel clustering algorithms using the dual representation for clustering prototypes (that is, each prototype is formulated as a linear sum of after-mapped dataset elements, and hence the parameters to be optimized are not original prototypes anymore but linearly-combined coefficients) is that the clustering prototypes lie in high dimensional feature space and hence clustering results lack clear and intuitive descriptions as in the original space. In this paper, a novel kernelized fuzzy C-means (KFCM) algorithm is proposed to compensate for such a lack and then applied to the MR image segmentation. It is realized by replacing the original Euclidean distance in the FCM algorithm with a kernel-induced distance and adding a novel spatial penalty also. The penalty term acts as a regularizer and a coefficient associated with the term is ranging from zero to one. It is shown that the proposed algorithm has better segmentation results on simulated or real MR images corrupted by noise and other artifacts than the standard algorithms such as FCM.

The rest of this paper is organized as follows. In Section 2, some basic concepts on the 'kernel method' are briefly introduced. In Section 3, the KFCM is derived from the original FCM based on the 'kernel method'. The KFCM with spatial constraints is presented in Section 4 to segment the MR images. Some experimental comparisons are presented in Section 5. Finally, Section 6 gives our conclusions and several issues for future works.

2. The 'kernel method'

In the last years, a number of powerful kernel-based learning machines, e.g. Support Vector Machines (SVM) [13–15], Kernel Fisher Discriminant (KFD) [17] and Kernel Principal Component Analysis (KPCA) [16] were proposed and have found successful applications such as in pattern recognition and function approximation. A common philosophy behind these algorithms is based on the following kernel (substitution) trick, that is, firstly with a (implicit) nonlinear map, from the data space to the mapped feature space, $\Phi : X \to F(x \to \Phi(x))$, a dataset $\{x_1, \ldots, x_n\} \subseteq X$ (an input data space with low dimension) is mapped into a potentially much higher dimensional feature space or inner product F, which aims at turning the original nonlinear problem in the input space into potentially a linear one in rather high dimensional feature space so as to faciliate problem solving as proved by Cover [21].

A kernel in the feature space can be represented as a function K below:

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle \tag{1}$$

where $\langle \Phi(x), \Phi(y) \rangle$ denotes the inner product operation.

An interesting point about kernel function is that the inner product can be implicitly computed in F, without explicity using or even knowing the mapping Φ. So, kernels allow computing inner products in spaces, where one could otherwise not practically perform any computations. Three commonly-used kernel functions in literature are [22]:

(1) Gaussian radial basis function (GRBF) kernel

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

(2) Polynomial kernel

$$K(x, y) = (1 + \langle x, y \rangle)^d. \tag{3}$$

(3) Sigmoid kernel

$$K(x,y) = \tanh(\alpha\langle x,y\rangle + \beta) \tag{4}$$

where σ, d, α, β are the adjustable parameters of the above kernel functions. For the sigmoid function, only a set of parameters satisfying the Mercer theorem can be used to define a kernel function.

3. Kernelized fuzzy C-means algorithm

The standard FCM objective function for partitioning a dataset $\{x_k\}_{k=1}^N$ into c clusters is given by

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m ||x_k - v_i||^2 \tag{5}$$

where $\{v_i\}_{i=1}^c$ are the centers or prototypes of the clusters and the array $\{u_{ik}\}(=U)$ represents a partition matrix satisfying

$$U \in \left\{ u_{ik} \in [0,1] \middle| \sum_{i=1}^{c} u_{ik} = 1, \ \forall k \quad \text{and} \quad 0 < \sum_{k=1}^{N} u_{ik} < N, \ \forall i \right\} \tag{6}$$

The parameter m is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. In image clustering, the most commonly used feature is the gray-level value, or intensity of image pixel. Thus, the FCM objective function is minimized when high membership values are assigned to pixels whose intensities are close to the centroid of its particular class, and low membership values are assigned when the point is far from the centroid.

From the discussion in Section 2, we know every algorithm that only uses inner products can implicitly be executed in the feature space F. This trick can also be used in clustering, as shown in support vector clustering [18] and kernel (fuzzy) c-means algorithms [19,20]. A common ground of these algorithms is to represent the clustering center as a linearly-combined sum of all $\Phi(x_k)$, i.e. the clustering centers lie in feature space. In this section, we construct a novel kernelized FCM algorithm with objective function as follows:

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m ||\Phi(x_k) - \Phi(v_i)||^2 \tag{7}$$

where Φ is an implicit nonlinear map. Unlike [19,20], $\Phi(v_i)$ here is not expressed as a linearly-combined sum of all $\Phi(x_k)$ anymore, a so-called dual representation, but still reviewed as an mapped point (image) of v_i in the original space, then with the kernel substitution trick, we have

$$\begin{aligned} ||\Phi(x_k) - \Phi(v_i)||^2 &= (\Phi(x_k) - \Phi(v_i))^{\mathrm{T}}(\Phi(x_k) - \Phi(v_i)) \\ &= \Phi(x_k)^{\mathrm{T}}\Phi(x_k) - \Phi(v_i)^{\mathrm{T}}\Phi(x_k) \\ &\quad - \Phi(x_k)^{\mathrm{T}}\Phi(v_i) + \Phi(v_i)^{\mathrm{T}}\Phi(v_i) \\ &= K(x_k,x_k) + K(v_i,v_i) - 2K(x_k,v_i) \end{aligned} \tag{8}$$

Below we confine ourselves to the Gaussian RBF kernel, so $K(x,x) = 1$. From Eqs. (7) and (8), can be simplified to

$$J_m = 2\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m (1 - K(x_k, v_i)) \tag{9}$$

Formally, the above optimization problem comes in the form

$$\min_{U,\{v_i\}_{i=1}^c} J_m, \quad \text{subjects to Eq.(6)} \tag{10}$$

In a similar way to the standard FCM algorithm, the objective function J_m can be minimized under the constraint of U. Specifically, taking the first derivatives of J_m with respect to u_{ik} and v_i, and zeroing them, respectively, two necessary but not sufficient conditions for J_m to be at its local extrema will be obtained as follows:

$$u_{ik} = \frac{(1 - K(x_k, v_i))^{-1/(m-1)}}{\sum_{j=1}^{c}(1 - K(x_k, v_j))^{-1/(m-1)}} \tag{11}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m K(x_k, v_i) x_k}{\sum_{k=1}^{n} u_{ik}^m K(x_k, v_i)} \tag{12}$$

Here we use only the Gaussian RBF kernel for the simplicity of derivation of Eqs. (11) and (12) and hence the algorithm in [23] is just a special case of our algorithm. For other kernel functions, the corresponding equations are a little more complex, because their derivatives are not as simple as the Gaussian RBF kernel function.

The proposed kernelized fuzzy C-means algorithm can be summarized in the following steps:

Step 1: Fix c, $t_{\max}$, $m > 1$ and $\varepsilon > 0$ for some positive constant.
Step 2: Initialize the memberships u_{ik}^0.
Step 3: For $t = 1, 2, \ldots, t_{\max}$, do:
(a) update all prototypes v_i^t with Eq. (12);
(b) update all memberships u_{ik}^t with Eq. (11);
(c) compute $E^t = \max_{i,k}|u_{ik}^t - u_{ik}^{t-1}|$, if $E^t \le \varepsilon$, stop;
end;

According to Huber [24], a robust procedure should have the following properties: (1) it should have a reasonably good accuracy at the assumed model; (2) small deviations from the model assumptions should impair the performance only by a small amount; (3) larger deviations from the model assumptions should not cause a catastrophe. It has been shown that FCM is not a robust estimator from the robust statistical point of view [25]. In the literature, there are many robust estimators [26,25]. In Appendix A, we show that the above mentioned KFCM using the kernel in Eq. (2) is a robust estimator. Here we can give an intuitive explanation for the robustness of KFCM. By Eq. (12), the data point x_k is endowed with an additional weight $K(x_k, v_i)$, which measures the similarity between x_k and v_i. When x_k is an outlier, i.e. x_k is far from the other data points, $K(x_k, v_i)$ will be very small, thus the weighted sum of data points shall be suppressed and hence more robust. Note that when σ in Eq. (2) tends to zero, $K(x_k, v_i)$ turns into an impulse function with the value of 1 only at $x_k = v_i$ and 0 elsewhere. In this extreme case, each given data point will have no longer neighborhood but become a ''soliton'' and the distance between any two points in the feature space approaches a common value of 1, leading to difficulties in clustering them. On the other hand, when sigma tends to infinity, the distance between any two points in the feature space will approaches zero and thus all data will cluster together, leading to a difficulty in separating them. In short, we should avoid such extreme cases in practice and choose an appropriate value for sigma neither too large nor too small by trial-and-error technique or experience or prior knowledge. In this paper, we choose the sigma by the trial-and-error technique.

4. Spatially constrained KFCM for image segmentation

Although KFCM can be directly applied to image segmentation like FCM, it would be helpful to consider some spatial constraints on the objective function.

We propose a modification to Eq. (7) by introducing a penalty term containing spatial neighborhood information. As mentioned before, this penalty term acts as a regularizer and biases the solution toward piecewise-homogeneous labeling. Such regularization is helpful in segmenting images corrupted by noise. The modified objective function is given by

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m (1 - K(x_k, v_i)) + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m \sum_{r\in N_k} (1 - u_{ir})^m \tag{13}$$

where N_k stands for the set of neighbors that exist in a window around x_k (do not include x_k itself) and N_R is the cardinality of N_k. The parameter α controls the effect of the penalty term and lies between zero and one inclusive. The relative importance of the regularizing term is inversely proportional to the signal-to-noise ratio (SNR) of the MRI data. Low SNR would require a higher value of α, and vice versa. The new penalty term is minimized when membership value for a particular class of a pixel is large and the membership values for that cluster should be also large at neighboring pixels, and vice versa. In other words, a pixel's membership value is correlated to those of the other pixels at its neighborhood. It is interesting to note that Eq. (13) is reminiscent of Rajesh Dave's approach to robust fuzzy C-means [25]. However, there are at least two differences between Eq. (13) and Dave's robust fuzzy C-means (see Eq. (12) in [25]). Firstly, the membership u_{ik} in Eq. (13) is still constrained by Eq. (6), while in Dave's robust fuzzy C-means, there are no constraints on the memberships other than the requirement that they should be in [0, 1]. Another difference is that in Eq. (13) we emphasize more the effect of the neighbors on the memberships of a pixel, while in Dave's robust fuzzy C-means no such neighbors exists.

An iterative algorithm for minimizing Eq. (13) can be derived by evaluating the centroids and membership functions that satisfy a zero gradient condition like the KFCM. A necessary condition on u_{ik} for Eq. (13) to be at a local minimum or an saddle point is

$$u_{ik} = \frac{\left((1 - K(x_k, v_i)) + \left(\alpha\sum_{r\in N_k}(1 - u_{ir})^m / N_R\right)\right)^{-1/(m-1)}}{\sum_{j=1}^{c}\left((1 - K(x_k, v_j)) + \left(\alpha\sum_{r\in N_k}(1 - u_{jr})^m / N_R\right)\right)^{-1/(m-1)}} \tag{14}$$

Because the penalty function does not depend on v_i, the necessary conditions under which Eq. (13) attains its minima is identical to that of KFCM, i.e. Eq. (12). Alternating iterations between the two necessary conditions will result in convergence of the algorithm to the minima of Eq. (14), which is almost identical to the KFCM, except in Step 3(b),

using Eq. (14) instead of Eq. (11) to update the memberships.

In the above discussion, we confine ourselves to only Gaussian RBF kernel. In fact, in terms of Chapelle et al. [27], we can also use more general RBF kernels as below:

$$K(x,y) = \exp(-\rho d(x,y)) \tag{15}$$

$d(x, y)$ can be chosen to be the following general form

$$d(x,y) = \sum_i |x_i^a - y_i^a|^b (0 < b \leq 2) \tag{16}$$

Obviously, the generalized RBF kernels satisfy $K(x,x) = 1$, and when they are used in Eq. (13), the iterative equations are similar as Eqs. (12) and (14). Appendix B gives the detailed derivations of the two necessary conditions (Eqs. (12) and (14)) for Eq. (13) to be at a local minimum or a saddle point.

5. Results and discussions

In this section, we describe some experimental results to compare the segmentation performance

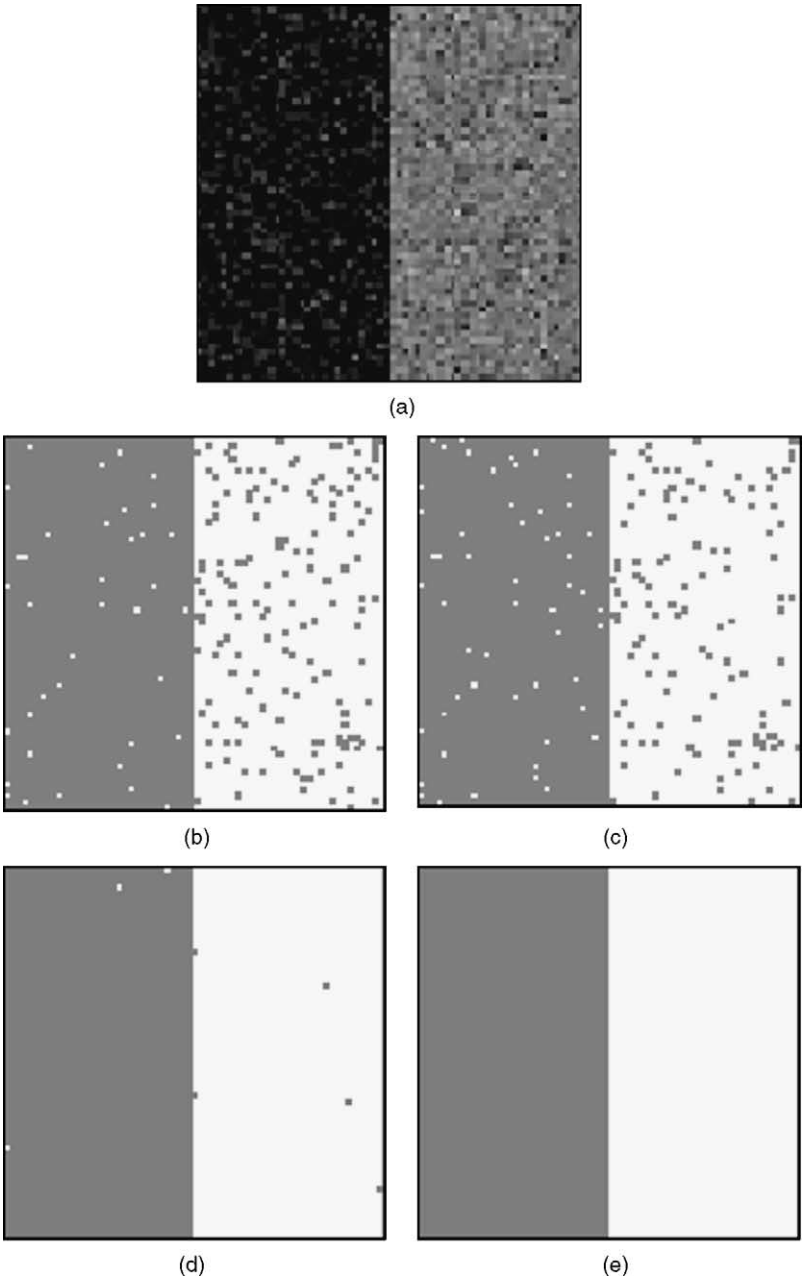

Figure 1 Comparison of segmentation results on a synthetic image corrupted by 5% Gaussian noise. (a) The original image, (b) using FCM, (c) using KFCM, (d) using SFCM, and (e) using SKFCM.

of the following algorithms, i.e. FCM, FCM with spatial constraints [12] (SFCM), KFCM, and KFCM with spatial constraints (SKFCM). We test the four methods on three different datasets. One is a simple synthetic image, another is the classical simulated brain database of McGill University [28], and the last one is real MR slices. Only the Gaussian RBF kernel is used for KFCM and SKFCM, and because of the large dynamic range of the parameter value in SFCM, we use the value recommended in [12].

The synthetic image is shown in Fig. 1a. It contains a two-class pattern corrupted by 5% Gaussian noise. The intensity values of the two classes are 0 and 90, respectively, and the image size is 64×64 pixels. Fig. 1b–e show the segmentation results of FCM, KFCM, SFCM and SKFCM, respectively. Here we set the parameter $\sigma = 150$ (Gaussian RBF kernel width), $\alpha = 0.7$, $m = 2$, $N_R = 8$ (a 3×3 window centered around each pixel, except the central pixel itself). These values will be used in the rest of this paper if no specific value is explicitly stated. As shown in Fig. 1b and c, without spatial constraints, neither FCM nor KFCM can separate the two classes, while SFCM nearly and SKFCM completely succeed in correcting and classifying the data as shown in Fig. 1d and e. Note that because of the

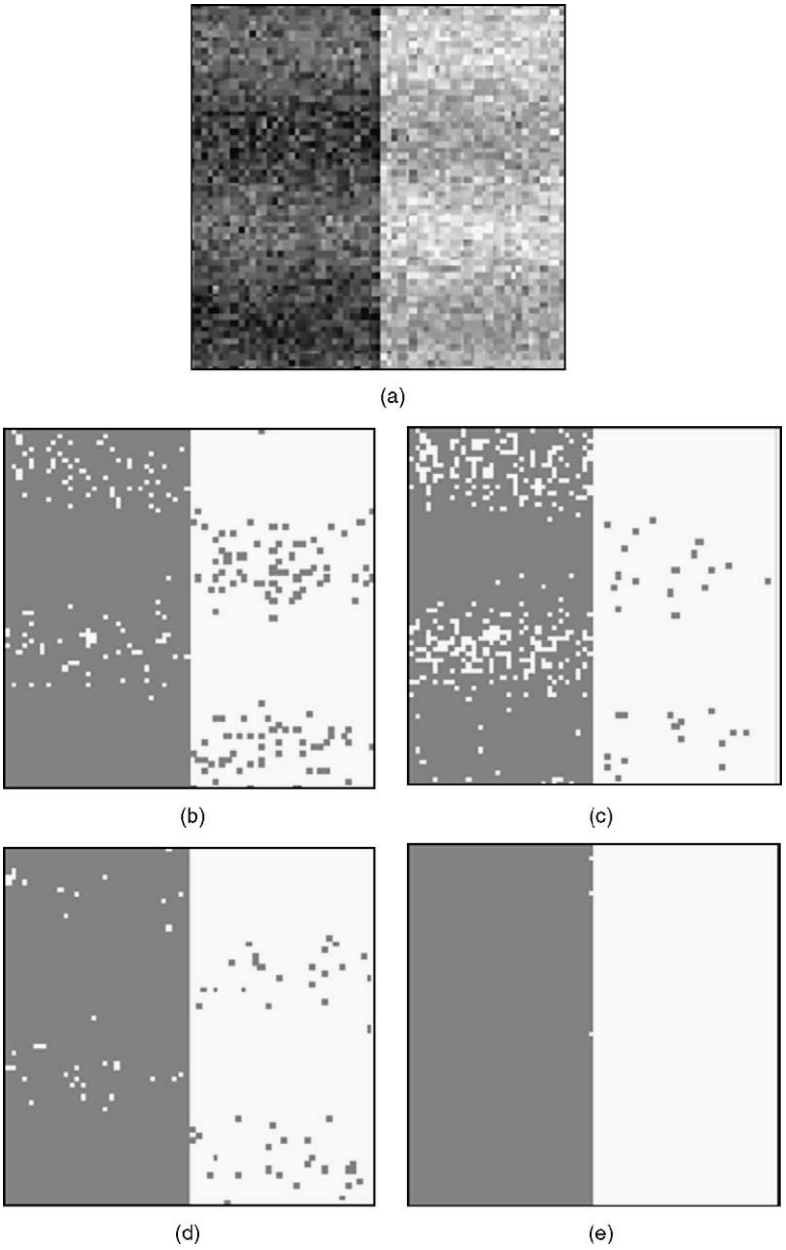

Figure 2 Comparison of segmentation results on a synthetic image corrupted by 5% Gaussian noise and sinusoidal intensity inhomogeneity. (a) The original image, (b) using FCM, (c) using KFCM, (d) using SFCM, and (e) using SKFCM.

injection of the kernel, KFCM need more execution time than FCM, and correspondingly, SKFCM is slower than SFCM. Typically, the algorithms without kernel are several times faster than those with injection of kernels.

Fig. 2a is the synthetic test image corrupted by 5% Gaussian noise and intensity inhomogeneity which is simulated by a sinusoid function. Fig. 2b–e show the results of FCM, KFCM, SFCM, and SKFCM, respectively. As in Fig. 1, SKFCM acquires the best segmentation performance. In order to obtain a quantitative comparison, Table 1 gives the segmentation accuracy (SA) of four methods in Figs. 1a and 2a, where the segmentation accuracy is defined as the sum of the total number of pixels divided by the sum of number of correctly classified pixels [11].

Table 1 Segmentation accuracy (SA %) of four methods on synthetic image

	FCM	KFCM	SFCM	SKFCM
Fig. 1a	96.02	96.51	99.34	100
Fig. 2a	94.41	91.11	98.41	99.88

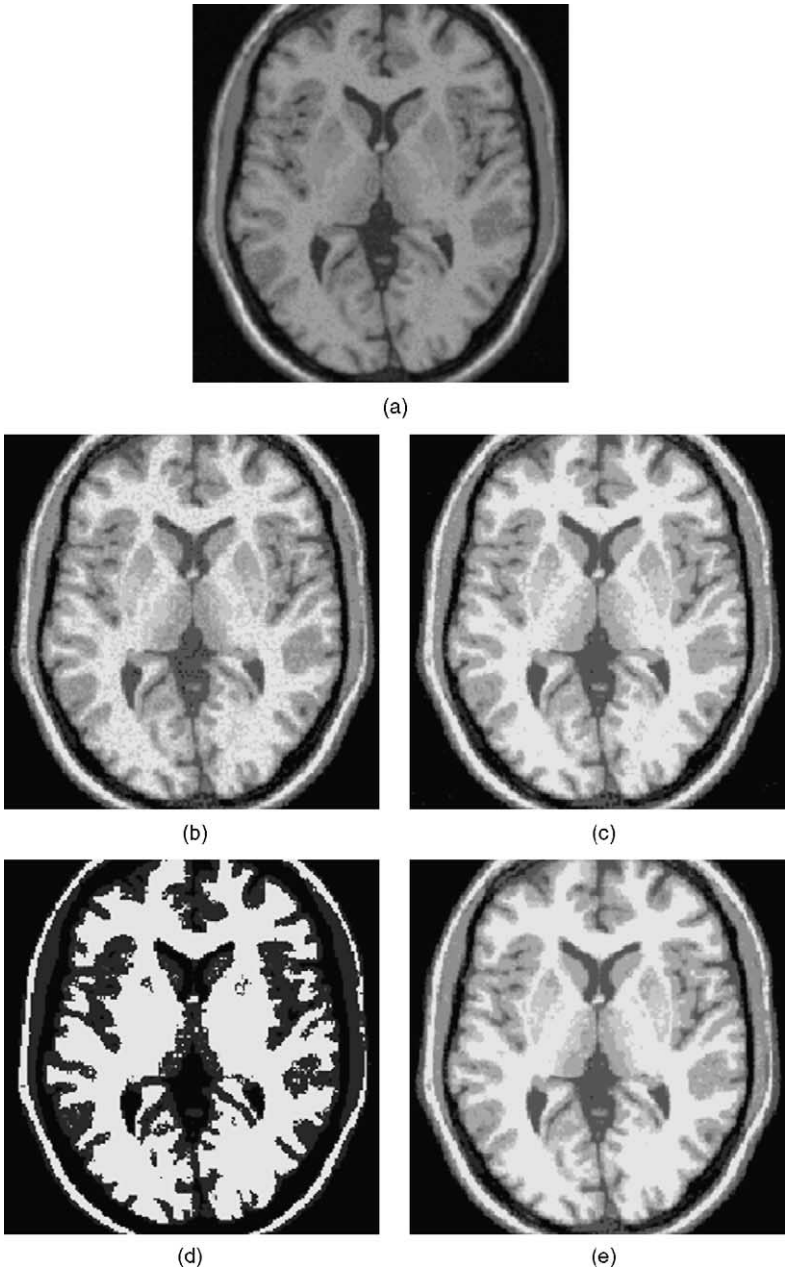

Figure 3 Comparison of segmentation results on a simulated brain MR image corrupted by 3% noise. (a) Original T1-weighted image, (b) using FCM, (c) using KFCM, (d) using SFCM, and (e) using SKFCM.

Figs. 3 and 4 present a comparison of segmentation results between FCM, KFCM, SFCM and SKFCM, when applied on T1-weighted MR phantom [28]. The advantages for using digital phantoms rather than real image data for validating segmentation methods include prior knowledge of the true tissue types and control over image parameters such as modality, slice thickness, noise and intensity inhomogeneities. Here in our experiments, we use a high-resolution T1-weighted phantom with slice thickness of 1mm, 3% noise and no intensity inhomogeneities. Two slices in the axial plane with the sequence of 91 and 121 are shown in Figs. 3a and 4a, respectively. The segmentation results on two slices using the four methods with eight classes are shown in Figs. 3b–e and 4b–e, respectively. Table 2 gives the quantitative comparison scores corresponding to Fig. 3a using four methods with eight classes. Note that SFCM in fact finds three classes, but because Class 5 is too small compared with the other classes, its score is rounded to 0.00 in Table 2. The comparison scores (also named as degree of equality in [29]) for each algorithm and for each class are calculated by the following equation [30]:

$$s_{ij} = \frac{A_{ij} \cap A_{\mathrm{ref}j}}{A_{ij} \cup A_{\mathrm{ref}j}} \tag{17}$$

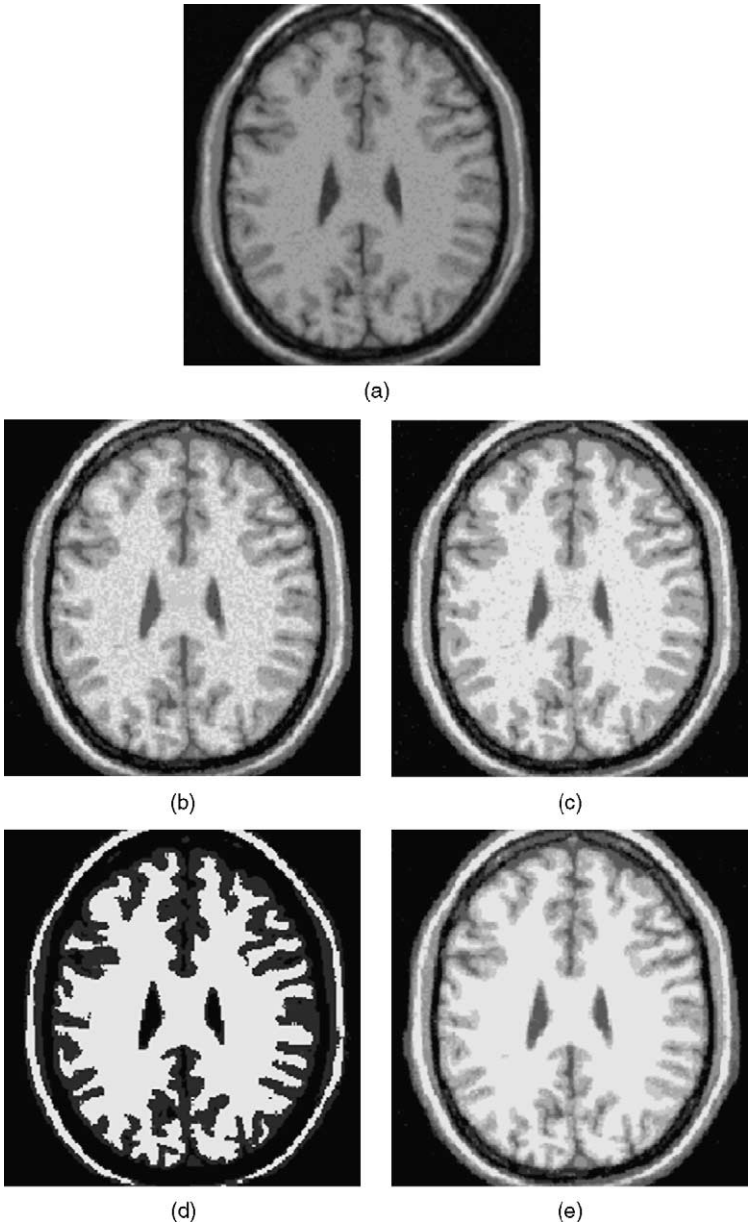

Figure 4 Another simulated brain MR example. (a) Original T1-weighted image corrupted by 3% noise, (b) using FCM, (c) using KFCM, (d) using SFCM, and (e) using SKFCM.

Table 2 Comparison scores for Fig. 3a using four methods with eight classes

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
FCM	0.46	0.75	0.51	0.39	0.47	0.35	0.65	0.17
KFCM	0.63	0.86	0.73	0.66	0.75	0.60	0.88	0.26
SFCM	0	0	0	0	0.00	0	0.59	0.31
SKFCM	0.66	0.86	0.72	0.72	0.69	0.47	0.88	0.29

where A_{ij} represents the set of pixels belonging to the jth class found by the ith algorithm and A_{refj} represent the set of pixels belonging to the jth class in the reference segmented image. Here we choose $\alpha = 0.1$, because the noise is relatively small. In this case, FCM and SFCM cannot correctly classify the images, while KFCM and SKFCM acquire satisfying segmentation results.

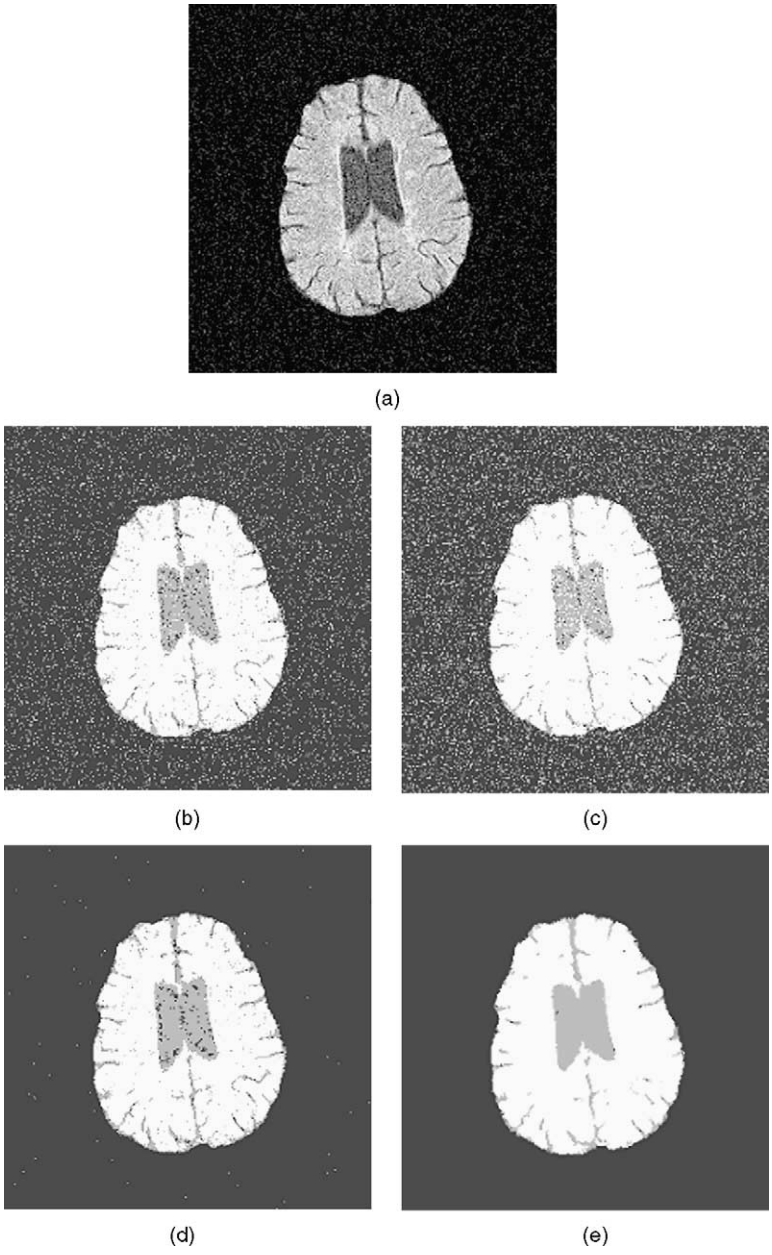

Figure 5 Brain MRI example. (a) Original MR image corrupted by 5% Gaussian noise, (b) FCM result, (c) KFCM result, (d) SFCM result, and (e) SKFCM result.

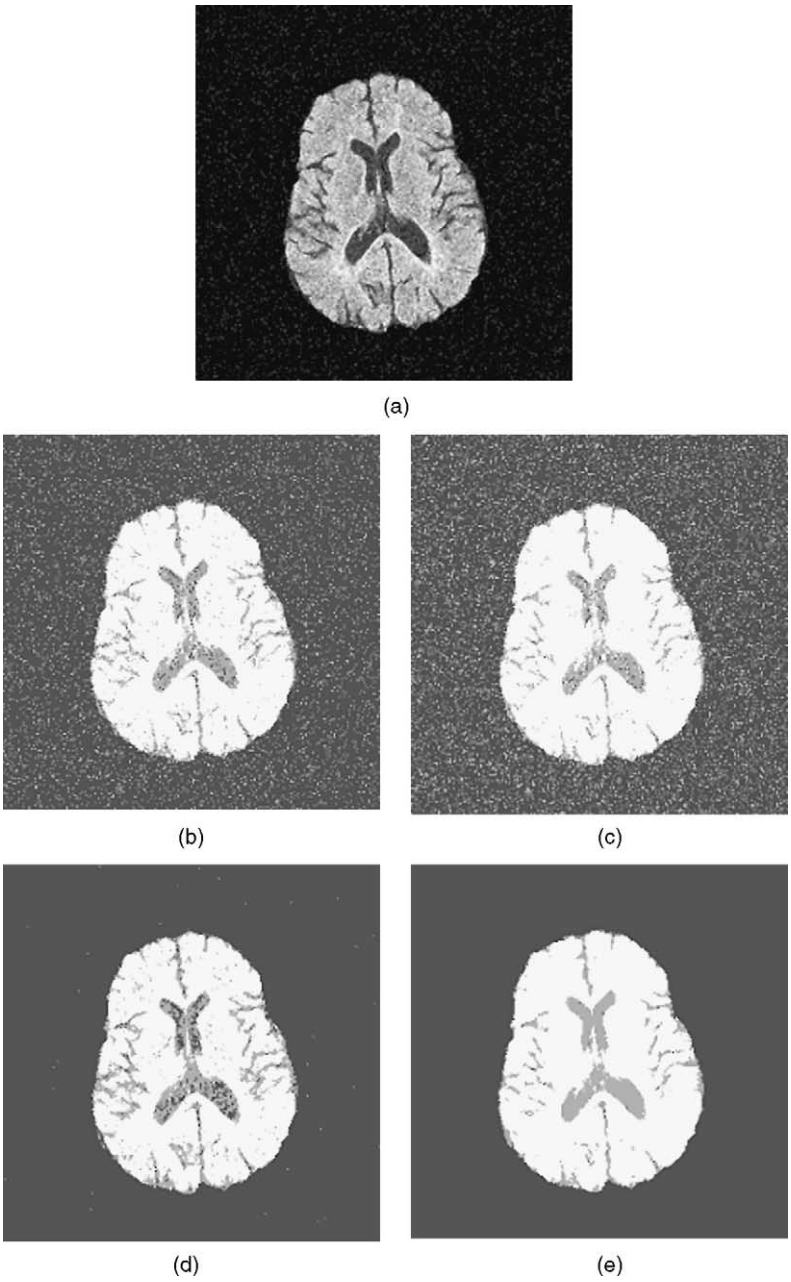

Figure 6 Another Brain MRI example. (a) Original MR image corrupted by 5% Gaussian noise, (b) FCM result, (c) KFCM result, (d) SFCM result, and (e) SKFCM result.

Figs. 5 and 6 present a comparison of segmentation results among FCM, KFCM, SFCM and SKFCM when applied on real MR slices corrupted by 5% Gaussian noise. These T1-weighted MR images are obtained using a General Electric Signa 1.5-Tesla clinical MR imager with in-plane resolution of 0.94 mm^2 [11]. Figs. 5a and 6a show the artificially corrupted images, and Figs. 5b—e and 6b—e are the results using FCM, KFCM, SFCM, and SKFCM with three classes, respectively. Table 3 gives the corresponding comparison scores for Figs. 5a and 6a using four methods with three classes. From the images and Table 3, we can see that without spatial constraints, both FCM and KFCM are affected by the noise badly, while SFCM partially and SKFCM nearly completely eliminate the effect of noise.

Finally, Fig. 7 shows a comparison of segmentation results of the four methods on a T2-weighted MR image corrupted by slight intensity inhomogeneities. Fig. 7a is the original image, and Fig. 7b—e are the results using FCM, KFCM, SFCM and SKFCM, respectively. Note that SKFCM is much less fragmented than other algorithms and the incorporation

Table 3 Comparison scores for Figs. 5a and 6a using four methods with three classes

	Fig. 5a			Fig. 6a		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
FCM	0.88	0.32	0.93	0.87	0.29	0.92
KFCM	0.79	0.19	0.91	0.79	0.17	0.91
SFCM	0.99	0.75	0.94	0.98	0.62	0.92
SKFCM	1.00	0.78	0.95	0.99	0.68	0.94

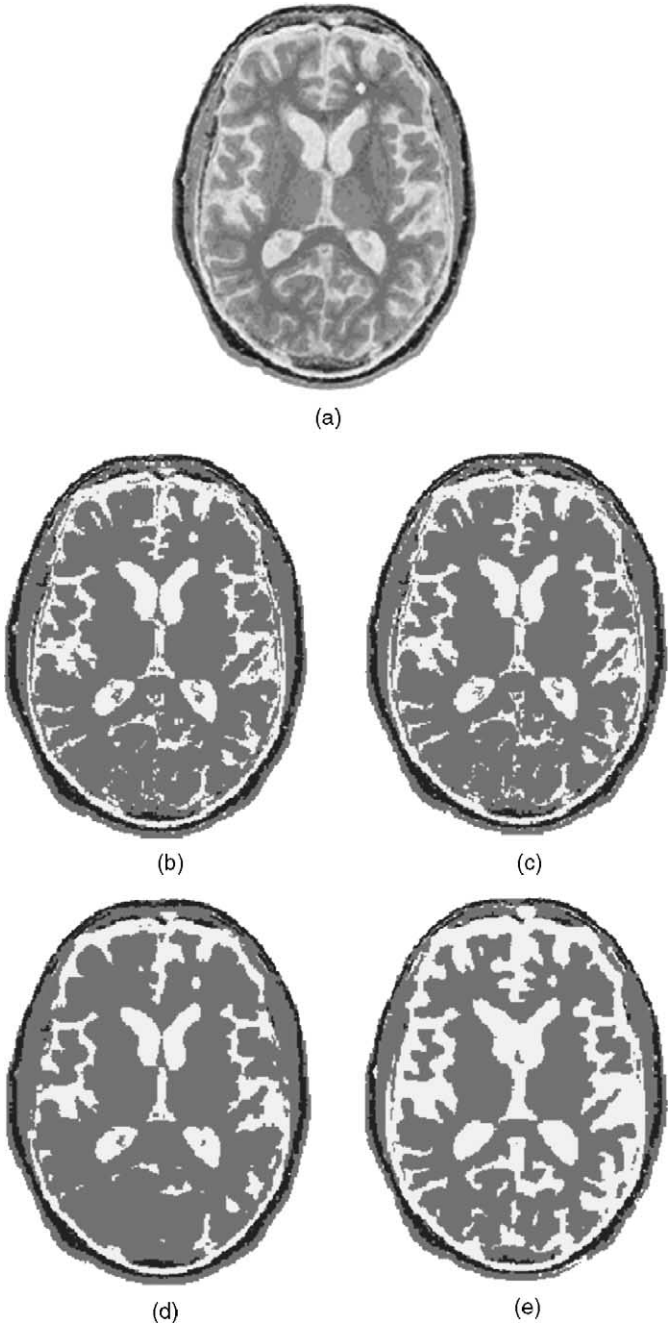

Figure 7 A T1-weighted Brain MRI example. (a) Original MR image corrupted by slight intensity inhomogeneities, (b) FCM result, (c) KFCM result, (d) SFCM result, and (e) SKFCM result.

of spatial constraints into the classification has somewhat the disadvantage of blurring of some fine details, but SKFCM had better result than SFCM, as shown in Fig. 7d and e.

6. Conclusion

In this paper, a novel kernelized fuzzy C-means algorithm is proposed and applied to MR image segmentation. KFCM adopts a new kernel-induced metric in the data space to replace the original Euclidean norm metric in FCM and the clustered prototypes still lie in the data space so that the clustering results can directly be reformulated and interpreted in the original space. It has been proved that KFCM is a robust clustering approach in Appendix A.

Furthermore, we added a spatial constraint on the objective function of KFCM to effectively segment MR images corrupted by noise. Although the spatial constraint used in KFCM is similar to that used in [12], ours is simpler and computationally inexpensive. What's more, since the distance induced by a Gaussian RBF is constrained between zero and one, exactly consistent with the range of the membership value; we only need to adjust the regularizer coefficient between zero and one to control the relative importance of the regularizing term.

The results presented in this paper are preliminary and further clinical evaluation is required. Because nearly all modified FCM algorithms for image segmentation are based on adding some type of penalty terms to the original FCM objective function, the KFCM can be applied in a reasonably straightforward manner to improve the performance of these algorithms.

Acknowledgements

The authors are grateful to the anonymous reviewers for their comments and suggestions which greatly improve the presentation of this paper. This work was supported in part by the National Science Foundations of China and of Jiangsu under Grant Nos. 60271017 and BK2002092, respectively.

Appendix A

Proof that KFCM using the kernel in Eq. (2) is a robust estimator.

Proof. According to Huber [24], there are many robust estimators, e.g. M-, L-, and R-estimator. In this section, we are only interested in M-estimator and follow the process of proof in [23]. Let $\{x_1, x_2, \ldots, x_n\}$ be an observed dataset and θ an unknown parameter to be estimated. A M-estimator for the location estimation can be generated by minimizing the following objective function:

$$J(\theta) = \sum_{i=1}^{n} \rho(x_i - \theta) \tag{A.1}$$

where ρ is a function and taken as $1 - K(x, \theta)$ in this proof and only dependent on $(x_i - \theta)$. Then the M-estimator is generated by solving the equation

$$J'(\theta) = \sum_{i=1}^{n} \phi(x_i - \theta) = 0 \tag{A.2}$$

where $\phi(x - \theta) = (\partial/\partial\theta)\rho(x - \theta)$. If we take $\rho(x - \theta) = (x - \theta)^2$ and $\rho(x - \theta) = |x - \theta|$, respectively, their M-estimators are the corresponding mean and median of the sample dataset.Eq. (A.2) can be solved by rewriting as

$$\sum_{i=1}^{n} w_i(x_i - \theta) = 0 \tag{A.3}$$

where $w_i = \phi(x_i - \theta)/(x_i - \theta)$, called the weighted function. This gives the estimator as below:

$$\hat{\theta} = \sum_{i=1}^{n} \frac{w_i x_i}{\sum_{i=1}^{n} w_i} \tag{A.4}$$

It is the weighted mean of the sample dataset. Note the result by solving Eq. (A.4) may not be a closed form for $\hat{\theta}$. We can apply the fixed-point iteration or alternate optimization to obtain a solution of Eq. (A.4) iteratively.

The influence function (IF) can help us to assess the relative influence of individual observations toward the value of an estimate. The M-estimator has been shown that its influence is proportional to its ϕ function [24]. Now we have IF of an M-estimator with

$$\mathrm{IF}(x; F, \theta) = \frac{\phi(x - \theta)}{\int \phi'(x - \theta)\, \mathrm{d}F_X(x)} \tag{A.5}$$

where $F_X(x)$ stands for a distribution function of X. If an IF of an estimator is unbounded, an outlier might cause trouble. There are many measures of robustness derived from IF, one of which is the gross error sensitivity (GES) defined below:

$$\gamma^* = \sup_x |\mathrm{IF}(x; F, \theta)| \tag{A.6}$$

This quantity can interpret the worst approximate influence that the addition of an infinitesimal point

mass can have on the value of the associated estimator. For our case, the resulting solutions from Eq. (9) is an M-estimator if $\rho(x-\theta) = 1 - K(x,\theta)$ and $K(x,\theta)$ is taken as Eq. (2). For simplicity, we only consider Eq. (2) with single variable, i.e.

$$K(x,\theta) = \exp\left(\frac{-(x-\theta)^2}{\sigma^2}\right) \tag{A.7}$$

Thus we have

$$\phi(x-\theta) = \exp\left(\frac{-(x-\theta)^2}{\sigma^2}\right)\left(\frac{-2}{\sigma^2}\right)(x-\theta) \tag{A.8}$$

By applying the L'Hospital's rule, the following limitations for Eq. (A.8) can be derived:

$$\lim_{x\to\pm\infty}\phi(x-\theta) = 0 \tag{A.9}$$

And at the same time, we can get their bounded maximum and minimum by zeroing the derivative of the function in Eq. (A.8).

According to the above, the function in Eq. (A.8) is bounded and continuous and thus the corresponding IF is also bounded and continuous, resulting in a finite gross error sensitivity. Hence our estimator based on Eq. (2) is robust.

Appendix B

Derivations of Eqs. (12) and (14).

Proof. The constraint optimization problem in Eq. (10), where J_m is defined in Eq. (13), can be solved by using the Lagrange multiplier method. Now define new objective function as follows:

$$L_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m (1 - K(x_k, v_i)) + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^m \sum_{r\in N_k}(1-u_{ir})^m + \lambda\left(1 - \sum_{i=1}^{c} u_{ik}\right) \tag{B.1}$$

where the kernel is define in Eq. (2), N_k the set of neighbors that exist in a window around x_k, but do not include x_k itself, and N_R is the cardinality of N_k. Taking the derivative of L_m with respect to u_{ik} and setting the result to zero, we have, for $m > 1$

$$\left\{\frac{\partial L_m}{\partial u_{ik}} = m u_{ik}^{m-1}\left[(1 - K(x_k, v_i)) + \frac{\alpha}{N_R}\sum_{r\in N_k}(1-u_{ir})^m\right] - \lambda\right\}_{u_{ik}=u_{ik}^*} = 0 \tag{B.2}$$

Solving for u_{ik}^*, we have

$$u_{ik}^* = \left(\frac{\lambda}{m\left((1 - K(x_k, v_i)) + \left(\alpha\sum_{r\in N_k}(1-u_{ir})^m/N_R\right)\right)}\right)^{1/(m-1)} \tag{B.3}$$

Since $\sum_{j=1}^{c} u_{jk} = 1, \quad \forall k$

$$\sum_{j=1}^{c}\left(\frac{\lambda}{m\left((1 - K(x_k, v_j)) + \left(\alpha\sum_{r\in N_k}(1-u_{jr})^m/N_R\right)\right)}\right)^{1/(m-1)} = 1 \tag{B.4}$$

Solving λ from Eq. (B.4) and substituting into Eq. (B.3), we can get the zero-gradient condition for the membership as shown in Eq. (14).

Similarly, Substituting Eq. (2) into Eq. (B.1) and zeroing the derivative of L_m with respect to v_i, we have

$$\left\{\frac{\partial L_m}{\partial v_i} = \sum_k u_{ik}^m(-K(x_k, v_i))2(x_k - v_i)\right\}_{v_i=v_i^*} = 0 \tag{B.5}$$

According to Eq. (B.5), Eq. (12) can be derived.

References

[1] Pham DL, Xu CY, Prince JL. A survey of current methods in medical image segmentation. Ann. Rev. Biomed. Eng. 2000;2:315–37 [Technical report version, JHU/ECE 99–01, Johns Hopkins University].

[2] Wells WM, Grimson WEL, Kikinis R, Arrdrige SR. Adaptive segmentation of MRI data. IEEE Trans Med Imaging 1996;15:429–42.

[3] Bezdek JC, Hall LO, Clarke LP. Review of MR image segmentation techniques using pattern recognition. Med Phys 1993;20:1033–48.

[4] Dawant BM, Zijidenbos AP, Margolin RA. Correction of intensity variations in MR image for computer-aided tissue classification. IEEE Trans Med Imaging 1993;12:770–81.

[5] Johnson B, Atkins MS, Mackiewich B, Andson M. Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI. IEEE Trans Med Imaging 1996;15:154–69.

[6] Pham DL, Prince JL. An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities. Pattern Recognit Lett 1999;20(1):57—68.
[7] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press; 1981.
[8] Tolias YA, Panas SM. On applying spatial constraints in fuzzy image clustering using a fuzzy rule-based system. IEEE Signal Proc Lett 1998;5(10):245—7.
[9] Tolias YA, Panas SM. Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership fucntions. IEEE Trans Syst, Man, Cybernet Part A 1998;28(3):359—69.
[10] Liew AWC, Leung SH, Lau WH. Fuzzy image clustering incorporating spatial continuity. IEE Proc Vis Image Signal Proc 2000;147(2):185—92.
[11] Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imaging 2002;21(3):193—9.
[12] Pham DL. Fuzzy clustering with spatial constraints. In: Proceedings of the IEEE International Conference on Image Processing, New York, USA, August, 2002.
[13] Cristianini N, Taylor JS. An introduction to SVMs and other kernel-based learning methods. Cambridge University Press; 2000.
[14] Vapnik VN. Statistical learning theory. New York: Wiley; 1998.
[15] Scholkopf B. Support vector learning. R. Oldenbourg Verlay; 1997.
[16] Scholkopf B, Smola AJ, Muller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 1998;10(5):1299—319.
[17] Roth V, Steinhage. Nonlinear discriminant analysis using kernel functions. In: Advances in neural information processing systems, vol. 12. 2000, pp. 568—74.
[18] Hur AB, Horn D, Siegelmann HT, Vapnik V. Support vector clustering. J Mach Learn Res 2001;2:125—37.
[19] Girolami M. Mercer kernel-based clustering in feature space. IEEE Trans Neural Networks 2002;13(3):780—4.
[20] Zhang DQ, Chen SC. Fuzzy clustering using kernel methods. In: Proceedings of the International Conference on Control and Automation, Xiamen, China, June, 2002.
[21] Cover TM. Geometrical and statistical properties of systems of linear inequalities in pattern recognition. IEEE Trans Electron Comput 1965;14:326—34.
[22] Muller KR, Mika S et al. An introduction to kernel-based learning algorithms. IEEE Trans Neural Networks 2001; 12(2):181—202.
[23] Wu KL, Yang MS. Alternative c-means clustering algorithms. Pattern Recognit 2002;35:2267—78.
[24] Huber PJ. Robust statistics. New York: Wiley; 1981.
[25] Dave RN, Krishnapuram R. Robust clustering method: a unified view. IEEE Trans Fuzzy Syst 1997;5(2):270—93.
[26] Leski J. Towards a robust fuzzy clustering. Fuzzy sets and systems. Corrected proof, available online, 8 February 2003, in press.
[27] Chapelle O, Haffner P, Vapnik V. Syms for histogram-based image classification. IEEE Trans Neural Networks 1999; 10(5):1055—65.
[28] Kwan RKS, Evans AC, Pike GB. An extensible MRI simulator for post-processing evaluation. Visualization in biomedical computing (VBC'96). Lecture notes in computer science, vol. 1131. Springer-Verlag; 1996. pp. 135—40.
[29] Lin C.T, Lee C.S.G. Real-time supervised structure/parameter learning for fuzzy neural network. In: Proceedings of the 1992 IEEE International Conference on Fuzzy Systems, San Diego, CA, pp. 1283—90.
[30] Masulli F, Schenone A. A fuzzy clustering based segmentation system as support to diagnosis in medical imaging. Artif Intell Med 1999;16(2):129—47.