Computer Vision and Image Understanding **84,** 285–297 (2001)
doi:10.1006/cviu.2001.0951, available online at http://www.idealibrary.com on IDEAL®

Spatial Models for Fuzzy Clustering

Dzung L. Pham

Laboratory of Personality and Cognition, Gerontology Research Center, NIA/NIH, 5600 Nathan Shock Drive, Baltimore, Maryland 21224
E-mail: dzung.pham@nih.gov

Received May 25, 2001; accepted November 20, 2001

A novel approach to fuzzy clustering for image segmentation is described. The fuzzy C-means objective function is generalized to include a spatial penalty on the membership functions. The penalty term leads to an iterative algorithm that is only slightly different from the original fuzzy C-means algorithm and allows the estimation of spatially smooth membership functions. To determine the strength of the penalty function, a criterion based on cross-validation is employed. The new algorithm is applied to simulated and real magnetic resonance images and is shown to be more robust to noise and other artifacts than competing approaches. © 2001 Elsevier Science (USA)

Key Words: fuzzy clustering; fuzzy c-means; image segmentation; Markov random fields; cross-validation.

1. INTRODUCTION

The fuzzy C-means algorithm (FCM) has been utilized in a wide variety of image processing applications such as medical imaging [1, 2] and remote sensing [3, 4]. Its advantages include a straightforward implementation, fairly robust behavior, applicability to multichannel data, and the ability to model uncertainty within the data. A major disadvantage of its use in imaging applications, however, is that FCM does not incorporate information about spatial context, causing it to be sensitive to noise and other imaging artifacts.

Mathematically, FCM is derived to minimize the following objective function with respect to the membership functions u_{jk} and the centroids v_k

$$J_{\mathrm{FCM}} = \sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}\|\mathbf{y}_j - \mathbf{v}_k\|^2, \tag{1}$$

where y_j is the observation at pixel j, C is the number of clusters or classes, and Ω is the

1077-3142/01 $35.00

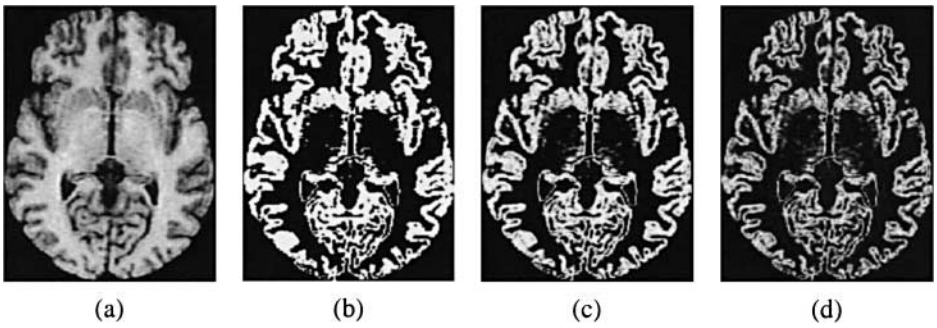


FIG. 1. Fuzzy c-means results for different values of the q parameter: (a) slice from MR image of the brain, (b) FCM membership function corresponding to gray matter tissue for $q = 1.2$, (c) FCM membership function for $q = 2$, (d) FCM membership function for $q = 3$.

image domain. The membership functions are constrained to be positive and to satisfy

$$\sum_{k=1}^{C} u_{jk} = 1. \tag{2}$$

Here, the objective function is minimized when high membership values are obtained in areas where the observations are close to the centroid, and low membership values are obtained where observations are distant from the centroid. The parameter q is a weighting exponent that satisfies $q > 1$ and controls the degree of "fuzziness" in the resulting membership functions. As q approaches unity, the membership functions become more crisp, and approach binary functions. As q increases, the membership functions become increasingly fuzzy. Figure 1 shows the results of applying FCM to a magnetic resonance (MR) image of the brain as the parameter q is increased.

It is apparent from (1) that the FCM objective function does not take into consideration any spatial dependence between observations. Thus, the computed membership functions can exhibit a sensitivity to noise in the observed image. One obvious approach to compensate for this sensitivity is to smooth the image before applying FCM. However, standard smoothing filters can result in a loss of important image details. More importantly, there is no way to rigorously control the trade-off between the smoothing and the clustering result that is obtained. Other approaches for obtaining spatially smooth membership functions include postprocessing the membership functions [5, 6] or incorporating multiscale information to enforce spatial constraints [4, 7, 8]. Recently, an approach was proposed for increasing the robustness of FCM to noise by directly modifying the objective function [9]. The normed distance computed at each pixel within the image was replaced with a weighted sum of distances within a neighborhood of the pixel. Of the aforementioned approaches, this latter one is the most similar to this work. The formulation of [9], however, leads to a rather counterintuitive equation for computing class centroids and the overall methodology may be difficult to incorporate into some FCM-based applications.

A more classical approach to incorporating spatial information is to penalize the FCM objective function to constrain the behavior of the membership functions, similar to methods used in regularization and Markov random field (MRF) theory [10]. This penalty can be used to discourage unlikely or undesirable configurations in the membership functions, such as a high membership value immediately surrounded by low values of the same class. An

advantage of using a penalty function is that only the computations involving membership values is changed—centroid computations are the same as in standard FCM. This kind of approach was adopted in [11], which used standard first order differences as a penalty to force membership values to be similar to neighboring values. The main problem with such a penalty function, however, is that it can drastically alter the characteristics of the membership function in an undesirable fashion. For example, first order differences will cause membership functions to be nearly piecewise constant. Second order differences will cause membership functions to be more smooth. However, depending on the value of the q parameter, this may contradict the desired characteristics of the membership functions.

In this paper, a new objective function is proposed for incorporating spatial context into FCM. The objective function includes a penalty term that is reminiscent of MRF priors but is consistent with the desired behavior of the membership functions dictated by the value of the q parameter. The algorithm for minimizing the new objective function, called the Robust Fuzzy C-means Algorithm (RFCM), possesses improved robustness to noise over FCM. An advantage of RFCM is that its equations are only slightly different from the original FCM equations. Furthermore, other variants of FCM (cf. [12, 13]) should also benefit from the proposed spatial model. It is shown that the strength of the penalty term can be selected using a criterion based on cross-validation. Results are shown applying the new algorithm to real and simulated magnetic resonance images. Throughout this paper, it will be assumed that the total number of clusters, C, is known. Methods based on FCM for determining the number of clusters [14, 15], however, can be applied to this work in a straightforward manner.

2. THEORY

In this section, a new objective function is proposed for fuzzy clustering with spatially constrained membership functions and an algorithm is derived for minimizing this objective function.

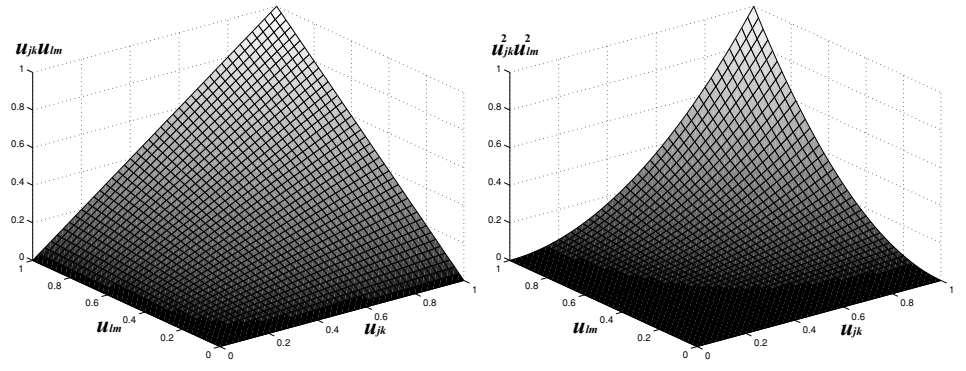
2.1. Penalized FCM

In order to restrict the membership functions in FCM to be spatially smooth, the following objective function is proposed

$$J_{\mathrm{RFCM}} = \sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}\|\mathbf{y}_j - \mathbf{v}_k\|^2 + \frac{\beta}{2}\sum_{j\in\Omega}\sum_{k=1}^{C} u_{jk}^{q}\sum_{l\in N_j}\sum_{m\in M_k} u_{lm}^{q}, \tag{3}$$

where N_j is the set of neighbors of pixel j, and $M_k = \{1, \ldots, C\}\backslash\{k\}$. The parameter β controls the trade-off between minimizing the standard FCM objective function and obtaining smooth membership functions. Appropriate selection of this parameter is described in Section 2.2.

The new penalty term is minimized when the membership value for a particular class is large and the membership values for the other classes at neighboring pixels is small (and vice versa). In other words, it constrains the membership value of a class to be negatively correlated with the membership values of the other classes at neighboring pixels. Figure 2 illustrates the role of the penalty function for different values of q. In both plots, the penalty of a single pairwise interaction is minimized when one or both membership functions are



(a) (b)

FIG. 2. Magnitude of the penalty function for different values of q: (a) plot of $u_{jk}u_{lm}$, (b) plot of $u_{jk}^2u_{lm}^2$.

zero, and maximized when both values are unity. Thus, configurations where adjacent pixels have high membership values in different classes are discouraged. When $q = 2$ there is a larger region of the plot where a relatively low penalty is obtained. This allows for the existence of membership values which are not necessarily binary, and is consistent with the role of q in the standard FCM formulation. Note that the penalty term is similar to the energy terms used in MRF prior probability models [16]. The major differences are that here, u_{jk} are continuously valued variables and the parameter q is included to regulate the "fuzziness" of the neighborhood interactions.

An iterative algorithm for minimizing (3) can be derived by evaluating the centroids and membership functions that satisfy a zero gradient condition. Using Lagrange multipliers to enforce the constraint in (2) and taking the partial derivative of (3) with respect to u_{jk} yields

$$\frac{\partial}{\partial u_{jk}}\left(\sum_{j\in\Omega}\sum_{k=1}^{C}u_{jk}^q\left(\|\mathbf{y}_j-\mathbf{v}_k\|^2+\frac{\beta}{2}\sum_{l\in N_j}\sum_{m\in M_k}u_{lm}^q\right)+\sum_{j\in\Omega}\lambda_j\left(1-\sum_{k=1}^{C}u_{jk}\right)\right)$$
$$=qu_{jk}^{q-1}\left(\|\mathbf{y}_j-\mathbf{v}_k\|^2+\beta\sum_{l\in N_j}\sum_{m\in M_k}u_{lm}^q\right)-\lambda_j, \tag{4}$$

where the factor of $\frac{1}{2}$ on β vanishes because the derivative operator results in a term corresponding to the product of u_{jk} and its neighbors, plus a term corresponding to the reverse product of the neighbors and u_{jk}. Setting this partial derivative to zero yields

$$u_{jk}=\left(\frac{q\left(\|\mathbf{y}_j-\mathbf{v}_k\|^2+\beta\sum_{l\in N_j}\sum_{m\in M_k}u_{lm}^q\right)}{\lambda_j}\right)^{-1/(q-1)}. \tag{5}$$

To solve for λ_j, the constraint equation (2) is employed, yielding

$$\sum_{k=1}^{K}\left(\frac{q\left(\|\mathbf{y}_j-\mathbf{v}_k\|^2+\beta\sum_{l\in N_j}\sum_{m\in M_k}u_{lm}^q\right)}{\lambda_j}\right)^{-1/(q-1)}=1. \tag{6}$$

Because λ_j does not depend on k, it can be factored out of the summation, leading to

$$\lambda_j^{-1/(q-1)} = q \sum_{k=1}^{C} \left(\|\mathbf{y}_j - \mathbf{v}_k\|^2 + \beta \sum_{l \in N_j} \sum_{m \in M_k} u_{lm}^q \right)^{-1/(q-1)}. \tag{7}$$

Combining Eqs. (5) and (7) results in the following necessary condition for u_{jk} to be at a local minimum of J_{RFCM}

$$u_{jk} = \frac{\left(\|\mathbf{y}_j - \mathbf{v}_k\|^2 + \beta \sum_{l \in N_j} \sum_{m \in M_k} u_{lm}^q \right)^{-1/(q-1)}}{\sum_{i=1}^{C} \left(\|\mathbf{y}_j - \mathbf{v}_i\|^2 + \beta \sum_{l \in N_j} \sum_{m \in M_i} u^q{}_{lm} \right)^{-1/(q-1)}}. \tag{8}$$

The presence of the q parameter within the penalty function allows for an elegant expression to be obtained for u_{jk}. Using a different exponent value would yield a considerably more complicated equation. When $\beta = 0$, Eq. (8) is identical to the standard FCM equation for computing membership values. When $\beta > 0$, the dependency on the neighbors causes u_{jk} to be large when the neighboring membership values of other classes is small. Conversely, if the neighboring membership values are small, the membership value at u_{jk} is increased. The result is a smoothing effect that causes neighboring membership values in the same class to be similar to one another. Note that the positivity condition on all membership values is satisfied by (8).

Because the penalty function does not depend on v_k, the necessary conditions on the centroids for (3) to be minimized is identical to that of FCM [17]. Iteration through these necessary conditions results in an algorithm for minimizing the objective function that is almost identical to standard FCM. The steps are as follows:

ALGORITHM 1. Robust Fuzzy C-means Algorithm

1. Obtain initial estimates of the centroids, v_k.
2. Compute membership functions using (8).
3. Compute centroids:

$$\mathbf{v}_k = \frac{\sum_{j \in \Omega} u_{jk}^q \mathbf{y}_j}{\sum_{j \in \Omega} u_{jk}^q}, \quad k = 1, \ldots, C \tag{9}$$

4. Go to step 2 and repeat until convergence.

For scalar imaging data, Step 1 can be performed automatically using the histogram smoothing method described in [18]. Convergence can be considered to be obtained when the change in the objective function is less than some threshold, or when the change in the membership values is less than some threshold. The latter criterion is used in this paper. Figure 3 shows the results of applying RFCM to the same image shown in Fig. 1a for different values of q and β. Compared to the standard FCM results in Fig. 1, the spatial constraint clearly causes a smoothing effect. However, the effect is consistent with the selection of q. Figures 3b and 3c show that for $q = 2$, even when the β parameter is significantly increased, the membership function does not become binarized.

Based on the results of [17] and because the penalty function is independent of the centroids, Step 3 of RFCM is guaranteed to decrease the objective function (3). To show that the computation in Step 2 will also decrease the objective function, it is sufficient to

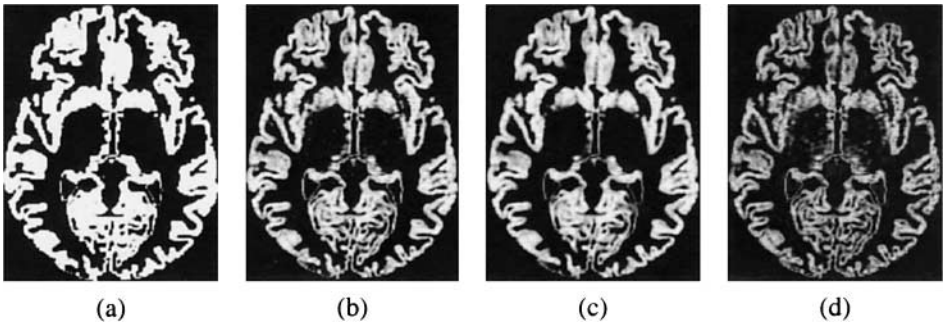


FIG. 3. RFCM membership function corresponding to gray matter for different values of β and q: (a) $\beta = 100$ and $q = 1.2$, (b) $\beta = 100$ and $q = 2$, (c) $\beta = 250$ and $q = 2$, (d) $\beta = 100$ and $q = 3$.

show that the Hessian matrix with respect to $u_{j1}, \ldots, u_{jC}$ is positive definite for each pixel j. This $C \times C$ matrix will be diagonal, with diagonal elements d_k that satisfy

$$d_k = q(q-1)u_{jk}^{q-2}\|\mathbf{y}_j - \mathbf{v}_k\|^2 + q(q-1)\beta u_{jk}^{q-2}\sum_{l\in N_j}\sum_{m\neq k} u_{lm}^q. \tag{10}$$

By substituting (8) for u_{jk} and assuming that $\|\mathbf{y}_j - \mathbf{v}_k\| > 0$, each d_k must be strictly positive and therefore the Hessian matrix will be positive definite. Together, these properties guarantee that the RFCM objective function must decrease with each iteration of the algorithm.

Straightforward generalizations to the penalty function in (3) are possible. For example, interactions between certain classes may be stronger than others. Furthermore, when pixels or voxels are not isotropic in size, it maybe desirable to vary the strength of interactions along different coordinate axes. A more general penalty function is

$$\sum_{j\in\Omega}\sum_{l\in N_j}\beta_l \mathbf{u}_j \mathbf{V}\mathbf{u}_l,$$

where $\mathbf{u}_j = [u_{j1}^q, \ldots, u_{jC}^q]$, V is a symmetric matrix with zero entries along the diagonal and positive entries elsewhere, and β_l now varies with respect to the direction of the neighbor being considered.

Note that in this work, the value of β does not vary with the local characteristics of the image. It may, in some applications, be desirable to adjust β according to certain image features, such as edges, to prevent smoothing of those features. Currently, pixels at step edges will incur a slight penalty in favor of other classes. Unlike typical noisy pixels, however, there will still be some neighbors of the same class. In addition, natural images tend to have softer edges and thus, the magnitude of the penalty will be smaller since pixels in between classes will have low membership values for all classes. The effects of edges on RFCM is demonstrated in Section 3, where it is shown that edges are generally still preserved even with a fixed β, when appropriately selected.

2.2. Selection of β

Proper selection of the β parameter is important for obtaining optimal or near-optimal performance. Determination of an appropriate value of β is dependent on the image being clustered. It is apparent from (3) and (8) that the value depends on the brightness of the

image as well as the variation of the intensity values relative to the centroid values of each class. In this section, a criterion based on cross-validation is described for choosing an appropriate value of β.

Cross-validation is a well-studied method for selection of a regularization parameter in solving discrete ill-posed problems [19]. The basic premise of cross-validation is that the data is split into a validation set and a data set. The algorithm under study is applied to the data set for a fixed value of the unknown parameter, where the data points corresponding to the validation set are assumed to be missing. The validation set is then used to test the appropriateness of the parameter selection by computing a cross-validation error. The process is repeated for different values of the parameter, and the parameter value that yields the lowest cross-validation error is considered to be optimal. In a sense, cross-validation selects the parameter that allows the data set to best estimate or interpolate the validation set.

In true cross-validation schemes, the validation set is chosen to be one pixel, the error is computed for that one pixel, and this process is repeated for every pixel in the image. The total cross-validation error is defined to be the sum of the errors resulting from using each pixel as a validation set. When the model is nonlinear and the size of the data is large, true cross-validation can be computationally intensive since the algorithm must be reapplied N times, where N is the total number of pixels in the image. As an alternative, the *holdout method* [20] is used in this work, where the validation set consists of a fixed subsampling of pixels, and the algorithm is applied only once to compute the cross-validation error over that set of pixels.

In the case of RFCM, the estimated membership values within the validation set as well as the centroids should satisfy the same criteria as the values within the data set. The cross-validation error can therefore be defined to be

$$E_{cv} = \sum_{j \in \Omega_v} \sum_{k=1}^{C} u_{jk}^{q} \|\mathbf{y}_j - \mathbf{v}_k\|^2, \tag{11}$$

where Ω_v is the validation set using the holdout method. The RFCM algorithm using cross-validation is as follows:

ALGORITHM 2. RFCM using cross-validation

1. Apply FCM (or RFCM with $\beta = 0$) to the image.
2. Compute $\beta_{inc} = 0.1 * \frac{J_{FCM}}{(J_{RFCM} - J_{FCM})/\beta}$.
3. Set $\beta = \beta_{inc}$.
4. Apply RFCM to the data set.
5. Compute E_{cv} on validation set.
6. Set $\beta = \beta + \beta_{inc}$, and go to Step 4, using the current result as an initialization of the next application of RFCM.
7. Apply RFCM to entire image using the value of β that minimizes E_{cv} .

Step 2 is a rule of thumb that simply increments β by 10% of the ratio of the data term to the penalty term in (3). In Step 3, the validation set is assumed to be missing, so the membership values at those locations are computed using the values of the neighboring membership values. In most situations, the cross-validation error is very well behaved for different values of β and the algorithm can be terminated when the first local minimum is found. Figure 4 shows a typical plot of the cross-validation error computed for various values of β when applying RFCM to an image. In this example, the error is minimized for

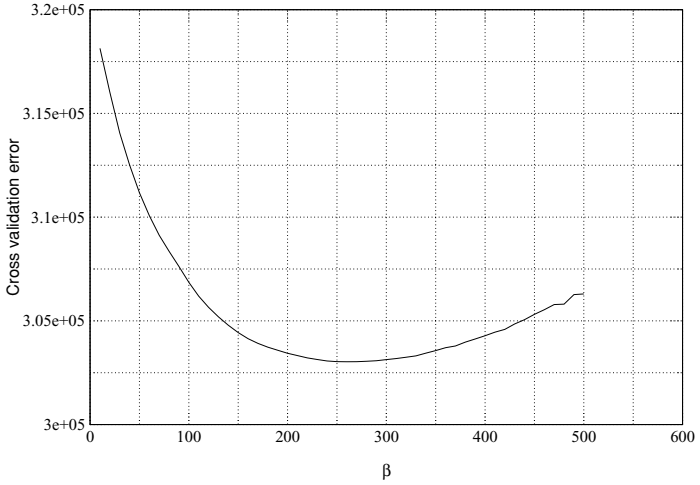


FIG. 4. Plot of cross-validation error for different values of β.

$\beta \approx 260$ and the curve is nearly convex, although an additional local minimum is present where β is very large. Note that using 10% of the ratio of the energy terms in Step 2 was determined empirically. In some cases, the step size may be too large and local minima might still be missed by the algorithm.

3. RESULTS

In this section, the results of applying RFCM to real and simulated images are shown. For the results on magnetic resonance imaging data, the algorithms were applied to three-dimensional images and the first-order, six voxel neighborhood system was used for the spatial penalty term in RFCM. In these experiments, $q = 2$ and $C = 3$, where the three classes correspond to gray matter, white matter, and cerebrospinal fluid in the brain. All images were eight bit and background pixels were ignored. The algorithms were implemented in C on a 1 GHz PC system running Linux. Execution times varied according to the size of the image and amount of noise. Noisier images increased the amount of time required for both estimating β and for convergence of the clustering algorithm. For the two-dimensional results, total execution time was approximately 30 s. For the three-dimensional scalar images (approximately $180 \times 180 \times 180$ voxels), execution times varied from 4 to 12 min to estimate β, and 40 s to 3.5 min to perform the final clustering. As a comparison, standard FCM required approximately 20 s when applied to the same images.

Figure 5 displays the results of FCM and RFCM when applied to a two-dimensional synthetic test image. The test image possesses four classes with intensity values 0, 25, 50, and 75, and the image size is 96×96 pixels. White Gaussian noise with a standard deviation of 10 was added to the image. Figure 5a shows the test image and Figs. 5b and 5c show the results of the maximum membership classification produced by standard FCM and RFCM, respectively. A maximum membership classification is simply the assignment of a pixel to the class of highest membership. The value of β was determined by the cross-validation method using the first found local minimum and was found to be approximately 435. The RFCM classification is clearly much more robust to the presence of noise than the FCM

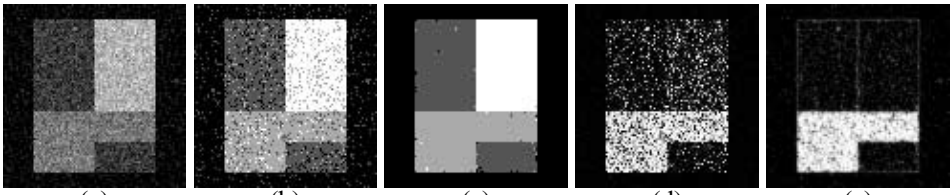


FIG. 5. FCM and RFCM applied to a two-dimensional test image: (a) test image, (b) FCM maximum membership classification, (c) RFCM classification, (d) FCM membership function, (e) RFCM membership function.

classification. The misclassification rate (MCR), computed by

$$MCR = \frac{number\ of\ pixels\ misclassified}{total\ number\ of\ pixels}, \tag{12}$$

was 14.14 and 0.52% for FCM and RFCM, respectively.

Figures 5d and 5e show the membership function generated by FCM and RFCM corresponding to the class with the second brightest intensity. These figures illustrate the effect of the neighborhood interactions on the computation of membership values. The RFCM result is obviously smoother and less speckled. It is also possible to see that there is a faint artifact induced at edges within the image. This is caused by the influence of the penalty function in neighborhoods containing pixels from different classes. Because edges do not conform to the assumption that neighboring pixels should be similar, the membership value of the actual class is lowered slightly, thereby slightly increasing the membership values for other classes.

Figure 6 shows the application of RFCM to a real multichannel MR image. Figures 6a and 6b show one slice taken from a dual spin–echo MR image of the brain. The images have been preprocessed to extract the intracranial volume. Figures 6c and 6d are the maximum membership classification obtained using FCM and RFCM, respectively. The cross-validation criterion was used to automatically select the amount of smoothing performed in RFCM, resulting in a value of β approximately equal to 80. The RFCM classification exhibits less speckling and the segmented tissues have smoother boundaries. Figures 6e–6h show the membership functions computed by FCM and RFCM for gray matter and white matter tissue types. Again, the smoother appearance of the RFCM results is evident, and fine details of the anatomy are still preserved.

Figure 7 shows some results of applying RFCM to a simulated MR image taken from the Brainweb database [21]. An MR image was simulated with 1-mm cubic voxels, 5% noise, and no inhomogeneity. For selection of β, every tenth voxel along each dimension was used as the validation set. Figure 7a shows the true hard classification used to simulate the image. The RFCM maximum membership classification shown in Fig. 7b is very close to the truth model. Figure 7c is the gray matter partial volume model used to simulate the MR image and Fig. 7d is the RFCM gray matter membership function. As has been reported previously in the literature [22], fuzzy clustering algorithms can be used to estimate partial volumes and Fig. 7d shows that RFCM obtains a reasonable reconstruction of the true partial volume fractions. Figures 7e and 7f show surfaces computed using FCM and RFCM that roughly correspond to the gray matter–white matter boundary. The surfaces were constructed using an isosurface algorithm computed on the white matter membership function at a value of

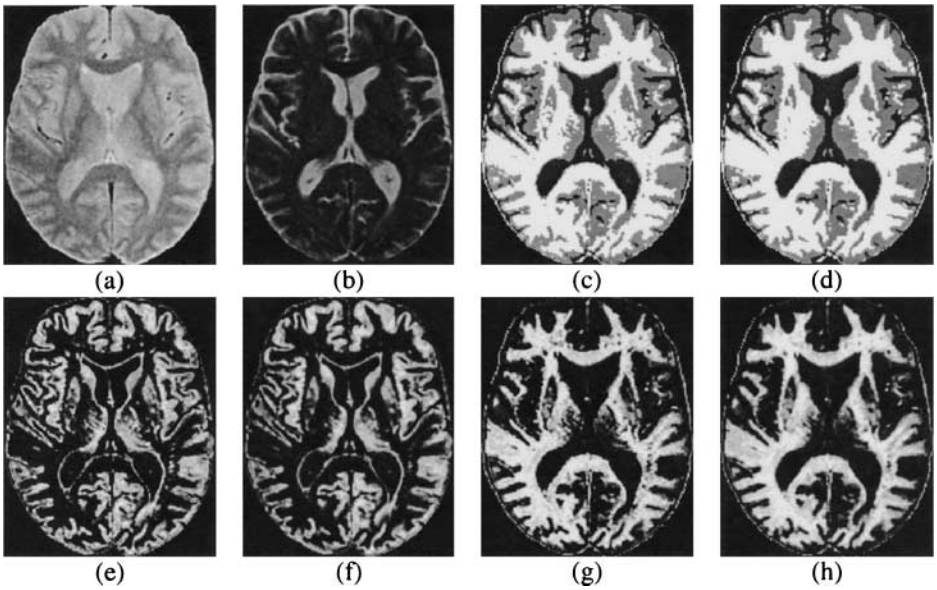


FIG. 6. FCM and RFCM applied to dual spin–echo MR images: (a) Proton density weighted image, (b) T2-weighted image, (c) FCM maximum membership classification, (d) RFCM maximum membership classification, (e) FCM gray matter membership, (f) RFCM gray matter membership, (g) FCM white matter membership, (h) RFCM white matter membership.

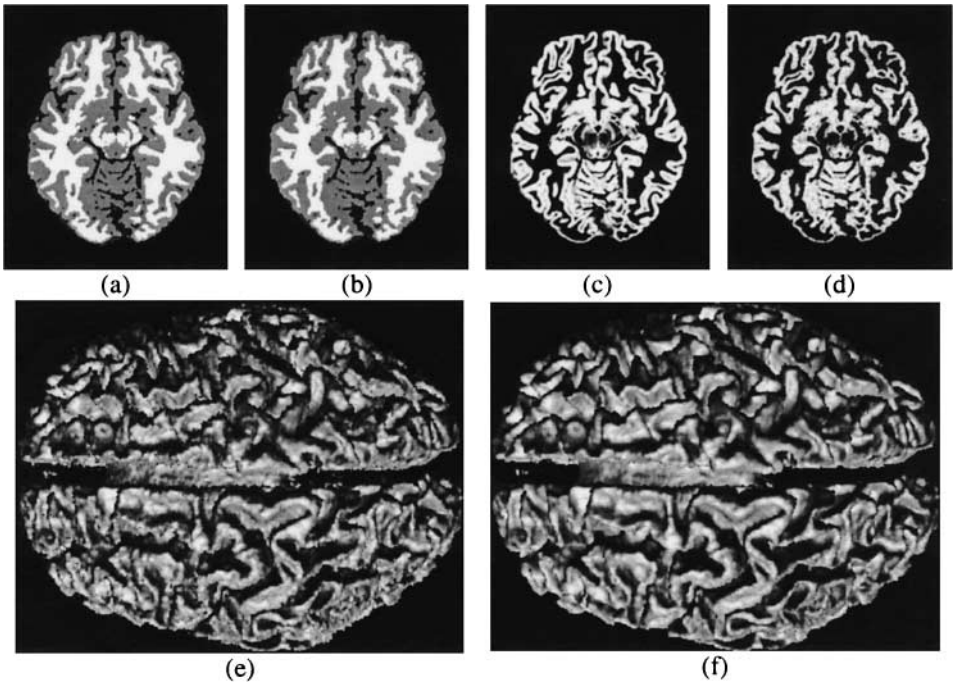


FIG. 7. Results on a simulated T1-weighted MR image: (a) true hard classification, (b) RFCM maximum membership classification, (c) true gray matter partial volume image, (d) RFCM gray matter membership, (e) surface of gray matter/white matter interface computed from FCM white matter membership, (f) surface computed from RFCM white matter membership.

TABLE 1
Error Measures from Simulated MR Results

	Noise level/Error					
	Misclassification Rate			Average RMS Error		
Method-Error	3%Noise	5%Noise	7%Noise	3%Noise	5%Noise	7%Noise
Optimal RFCM	3.761%	4.956%	6.047%	0.1188	0.1388	0.1587
CV RFCM	3.785%	5.017%	6.067%	0.1200	0.1394	0.1586
Standard FCM	3.988%	6.601%	10.249%	0.1302	0.1728	0.2208
FCM w/Med. filt.	6.930%	7.344%	7.881%	0.1825	0.1868	0.1930
FCM w/Opt. AD filt.	3.760%	5.188%	6.516%	0.1277	0.1524	0.1726

0.5. These figures clearly illustrate how RFCM is more robust than FCM, yielding smoother and more realistic boundaries.

Table 1 shows the error values obtained when applying several different algorithms to the simulated images for different levels of noise. Two error measures were used: MCR (see Eq. (12)) and the average root mean squared (RMS) error of the membership functions. The RMS error is computed between the estimated membership value and the true partial volume fractions. The sum of the errors over the three different classes divided by three yields the average RMS error. *Optimal RFCM* refers to RFCM applied to the images multiple times using different values of β, and selecting the result that yielded the lowest misclassification rate. In other words, β was chosen with prior knowledge of the truth models. *CV RFCM* refers to RFCM using the cross-validation criterion to select β. *Standard FCM* refers to FCM with no smoothing applied. *FCM w/ Med. Filt.* refers to FCM applied to images that have been prefiltered with a $3 \times 3 \times 3$ median filter. Finally, *FCM w/ Opt. AD filt.* refers to FCM applied to images that have been prefiltered with anisotropic diffusion, where "Opt." is an abbreviation for "Optimal." The smoothing parameter and the number of iterations of the anisotropic diffusion were selected by running the diffusion and FCM with multiple values and selecting the values that yielded the lowest misclassification rate.

As expected, an increase in the level of noise always led to an increase in the errors for all methods. The RFCM methods were most robust to increased noise, however. When comparing the optimal β to the β selected using cross-validation for RFCM, it was found that cross-validation consistently underestimated β. The optimal values of β were empirically found to be 75, 200, and 300 for the 3, 5, and 7% noise cases, respectively. The values of β determined using GCV were approximately 55, 150, and 260, respectively. This problem with cross-validation has been noted under other models as well [23, 24]. Despite this underestimation, the final errors are extremely close. Standard FCM performs quite well for low levels of noise but its performance rapidly degrades as noise is increased. The median filter results demonstrate what can occur when blindly applying filtering to remove noise. At the 3 and 5% noise levels, the median filter oversmoothed the image, increasing the errors over even the standard FCM method. The use of FCM in tandem with anisotropic diffusion results in substantial improvements over FCM and is comparable to RFCM in terms of misclassification rate. Because anisotropic diffusion tends to converge to images that are nearly piecewise constant however, its average RMS are all higher than the RFCM errors. In addition, appropriate selection of the number of iterations and smoothing parameter

was difficult. Small changes in those values sometimes led to large increases in error. Nevertheless, one might interpret the errors as saying that RFCM is operating in a manner somewhat similar to anisotropic diffusion.

4. CONCLUSIONS

This work describes a natural extension of FCM to incorporate spatial constraints on the membership functions. The new formulation should prove extremely beneficial to improving the performance of fuzzy clustering algorithms in image processing applications. This is especially true because the penalty function can be applied in a straightforward manner to nearly all variants of FCM, such as the adaptive fuzzy C-means algorithm [13] and the possibilistic C-means algorithm [12], with minimal changes to the resulting implementation.

ACKNOWLEDGMENTS

The author thanks Jerry Prince, Chenyang Xu, Xiao Han, Maryam Rettmann, Susan Resnick, and Alan Zonderman for their thoughtful comments and suggestions. The author also thanks Dr. Xu for his implementation of the three-dimensional anisotropic diffusion and median filtering algorithms.

REFERENCES

1. D. L. Pham, J. L. Prince, A. P. Dagher, and C. Xu, An automated technique for statistical characterization of brain tissues in magnetic resonance imaging, *Int. J. Pattern Recognit. Artificial Intell.*, to appear (Tech. Rep. JHU/ECE 96–13).
2. J. C. Bezdek, L. O. Hall, and L. P. Clarke, Review of MR image segmentation techniques using pattern recognition, *Med. Phys.* **20**, 1993, 1033–1048.
3. E. Rignot, R. Chellappa, and P. Dubois, Unsupervised segmentation of polarimetric SAR data using the covariance matrix, *IEEE Trans. Geosci. Remote Sensing* **30**(4), 1992, 697–705.
4. W. Chumsamrong, P. Thitimajshima, and Y. Rangsanseri, Syntetic aperture radar (SAR) image segmentation using a new modified fuzzy c-means algorithm, in *Proceedings of Geoscience and Remote Sensing Symposium*, Vol. 2, pp. 624–626, 2000.
5. Y. A. Tolias and S. M. Panas, On applying spatial constraints in fuzzy image clustering using a fuzzy rule-based system, *IEEE Signal Process. Lett.* **5**(10), 1998, 245–247.
6. A. Pardo and G. Sapiro, Vector probability diffusion, in *Proceedings of the 2000 Int'l Conf. on Image Processing*, Vol. 1, pp. 884–887, 2000.
7. Y. A. Tolias and S. M. Panas, Image segmentation by a fuzzy clustering algorithm using adaptive spatially constrained membership functions, *IEEE Trans. Systems, Man, Cybernet. A* **28**(3), 1998, 359–369.
8. S. T. Acton and D. P. Mukherjee, Scale space classification using area morphology, *IEEE Trans. Image Process.* **9**(4), 2000, 623–635.
9. A. W. C. Liew, S. H. Leung, and W. H. Lau, Fuzzy image clustering incorporating spatial continuity, *IEE Proc. Visual Image Signal Process.* **147**(2), 2000, 185–192.
10. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.* **6**, 1984, 721–741.
11. R. Castagno, T. Ebrahimi, and M. Kunt, Video segmentation based on multiple features for interactive multimedia applications, *IEEE Trans. Circuits Systems Video Technol.* **8**(5), 1998, 562–571.
12. R. Krishnapuram and J. M. Keller, A possibilistic approach to clustering, *IEEE Trans. Fuzzy Systems* **1**(2), 1993, 98–110.
13. D. L. Pham and J. L. Prince, Adaptive fuzzy segmentation of magnetic resonance images, *IEEE Trans. Med. Imaging* **18**(9), 1999, 737–752.

14. I. Gath and A. B. Geva, Unsupervised optimal fuzzy clustering, *IEEE Trans. Pattern Anal. Machine Intell.* **11**(7), 1987, 773–781.
15. X. L. Xie and G. Beni, A validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Machine Intell.* **13**(8), 1991, 841–847.
16. J. Zhang, The mean field theory in EM procedures for Markov random fields, *IEEE Trans. Signal Process.* **40**(10), 1992, 2570–2583.
17. J. C. Bezdek, A convergence theorem for the fuzzy ISODATA clustering algorithms, *IEEE Trans. Pattern Anal. Machine Intell.* **2**, 1980.
18. D. L. Pham and J. L. Prince, An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, *Pattern Recognit. Lett.* **20**(1), 1999, 57–68.
19. G. Wahba, *Spline Models for Observational Data*, Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, PA, 1990.
20. S. J. Reeves, A cross-validation framework for solving image restoration problems, *J. Visual Comm. Image Represent.* **3**, 1992, 433–445.
21. D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, and N. J. Kabani, *et al.*, Design and construction of a realistic digital brain phantom, *IEEE Trans. Med. Imaging* **17**, 1998, 463–468. [See http://www.bic.mni.mcgill.ca/brainweb]
22. M. E. Brandt, T. P. Bohan, L. A. Kramer, and J. M. Fletcher, Estimation of CSF, white and gray matter volumes in hydrocephalic children using fuzzy clustering of MR images, *Comput. Med. Imaging Graphics* **18**, 1994, 25–34.
23. A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, A study of methods of choosing the smoothing paramter in image restoration by regularization, *IEEE Trans. Pattern Anal. Machine Intell.* **13**(4), 1991, 326–339.
24. N. P. Galatsanos and A. K. Katsaggelos, Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation, *IEEE Trans. Image Process.* **1**(3), 1992, 322–336.