2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Fuzzy C-Means Clustering with Weighted Energy Function in MRF for Image Segmentation

Chi Wang, Jia Liu, Maoguo Gong, Licheng Jiao, and Jing Liu

***Abstract*—In this paper, we present a new Markov Random Field based FCM image segmentation algorithm. A new energy function is proposed to utilize the spatial and contextual information simultaneously. In the proposed energy function, we use a weighted distance to reflect the different effects of neighborhood pixels. By using the new energy function, the new algorithm has a better performance in noise-corrupted images. Experimental results on real and synthetic images show our method is effective.**

I. Introduction

Image segmentation is a traditional topic in image processing and is of great importance. After decades of development, many powerful algorithms have been proposed. Segmentation algorithms divide the given image into several parts by their intensity, color, texture or other property. A great number of them are based on the Fuzzy c-means algorithm and proved to be one of the main classes of algorithm on image segmentation. On noise-free images, the conventional FCM converges fast and outputs in an excellent precision. But if noise is added, the segmentation result would deteriorate. This is mainly because that the original FCM does not consider context information. Another reason for the noise susceptibility is the using of the Euclidean measure which is sensitive to outliers.

Many algorithms introduced regularization terms to include contextual information. Some of them use kernel measure instead of the Euclidean one and are proved to be more robust to noise. To name a few, FCM_S [1] first introduced spatial neighborhood terms. Chen et al. [2] proposed two variants of the FCM_S. Zhang et al. [3] proposed kernel distance measure for the FCM data space. Fast Generated FCM (FGFCM) [4] uses a sum image based on a non-linearly weight which consists of both spatial and intensity information. But the FGFCM have two defects. The first one is that the algorithm does not apply on the original image so the original information is not fully used. The second one is that it needs complicated parameter tuning. The Fuzzy Local Information C-Means algorithm, proposed by Krinidis et al. [5], solved these problems by introducing a regularize term called fuzzy factor. This factor considers both spatial distance between pixels and intensity differences with the clustering center.

This work was supported by the National Natural Science Foundation of China (Grant no. 61273317), the National Top Youth Talents Program of China, the Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20130203110011) and the Fundamental Research Fund for the Central Universities (Grant no. K5051202053).

The authors are with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an, 710071, China. Email: gong@ieee.org.

Recently, we proposed two powerful derivatives of the original FLICM: the RFLICM [6] and the KWFLICM [7]. They are proved to have a better performance than their predecessors. The first one we proposed, the RFLICM, improves the noise robustness of FLICM by introducing a coefficient concerning the intensity property of the context. The KWFLICM improves the RFLICM by simultaneously concerning the contextual dependencies and the kernel distance of a specific neighbor.

We noticed that Markov Random Field (MRF) can utilize neighborhood information effectively. Geman and Geman [8] first introduced MRF in image processing by making an analogy between images and statistical mechanics systems. They also mentioned the equivalence between MRF and Gibbs Random Field which makes MRF computable. A comprehensive introduction to MRF in image processing can be found in [9].

Because of the effectiveness of MRF in utilizing contextual information, some powerful MRF-based segmentation methods have been proposed. Using the idea from the mean field approximation principle in statistical physics [10], Celeux et al. [11] proposed an EM-like algorithm to infer parameters in Hidden Markov Random Field. By using the Kullback-Leibler divergence information, Chatzis [12] introduced the MRF prior into fuzzy objective function called HMRF-FCM. This method has a good noise robustness and segmentation precision.

In this study, in order to further exploit the MRF-based FCM algorithm, a new energy function is proposed which introduces a weighted distance between central pixel and the neighbor of it. Theoretical analysis is given and experimental results verified its excellent performance.

The remainder of this paper is organized as follows: Section II introduces the MRF-based FCM framework and our improvement on it. Section III considers the details in our novel energy function. Section IV presents experimental results and comparisons between our methods and competitive ones. Section V summarizes the whole paper.

II. Motivation and method

Let X denotes an image which is hoped to be divided into clusters given by set C. The standard FCM algorithm is the minimization process of the following objective function

$$J_{ij} = \sum_{i \in C} \sum_{j \in X} u_{ij}^{m} d_{ij} \tag{1}$$

where d_{ij} is the dissimilarity function.m is a parameter called fuzzifier.$U = \{u_{ij}\}$ is called membership matrix satisfying

$$U \in \left\{ u_{ij} \in [0,1] \mid \sum_{i=1} u_{ij} = 1 \, and \, 0 < \sum_{j=1}^{N} u_{ij} < N, \forall i, j \right\} \tag{2}$$

where N is the number of pixels. Using Lagrange multipliers, the objective function can be minimized iteratively. The FLICM algorithm introduces a fuzzy factor to regularize the above objective function. Many later algorithms are based on the FLICM and are proved to have good performance. However, there are some disadvantages of the FLICM-based segmentation algorithms. These algorithms are time consuming because of the complicated regularization term. Moreover, a recent comment on FLICM [13] reveals that it cannot converge to a local minima and the energy function actually cannot be minimize iteratively. As mentioned in Section I, MRF based FCM serves as a good way to model the grey level relation between pixels. Here comes a brief introduction to MRF.

An image I is considered as a field and each pixel j is an element. According to the Hammersley-Clifford theorem [8], a field is a Markov Random Field if and only if the following equation is satisfied

$$p(x_j \mid x_{I-\{j\}}) = p(x_j \mid x_{\partial_j}), \forall j \in I \tag{3}$$

where ∂_j denotes the neighborhood of the pixel j. And $I - \{j\}$means the pixels other than j. This equation is to indicate that the distribution of a pixel depend only on its neighbor. The theorem shows the equivalence between MRF and Gibbs Random Field. So the joint probability distribution can be given by

$$p(x \mid \beta) \triangleq Z(\beta)^{-1} \exp[-E(x \mid \beta)] \tag{4}$$

where $Z(\beta)$ is the partition function given by

$$Z(\beta) = \sum_{x \in C} \exp[-E(x \mid \beta)] \tag{5}$$

where E is the energy function and β is a parameter called the inverse supercritical temperature.

Here we have a computable expression of MRF. The definition and utilization of the energy function will be discussed in Section III.

The FCM objective function is modified to introduce probabilistic distributions. The FCM derivative proposed in [14] is satisfying. The K-L divergence is used to compute the distance between two distributions. Thus, the objective function is as follows

$$J_{ij} = \sum_{i \in C} \sum_{j \in X} u_{ij} d_{ij}(x_j - v_i) + \lambda \sum_{i \in C} \sum_{j \in X} u_{ij} \log(\frac{u_{ij}}{\pi_{ij}}) \tag{6}$$

where $d_{ij} = -log[p_i^k(x_j \mid \mu_i^k, \sigma_i^k)]$, λ= 0.1. The other parameters will be introduced in the following algorithm procedure.

The MRF-based FCM image segmentation algorithm:

1. Using the standard FCM algorithm, i.e. the equation (1) to initialize our membership matrix. Let $k = 1$.

2. Compute the energy function of each pixel. (see Section III)

3. Minimize the following function (7) to get the parameter β^k.

$$\beta^k = \arg\max_{\beta} \sum_{i=1}^{|C|} \sum_{j=1}^{N} \log p(x_j = i \mid x_{\partial_j}; \beta^k) \tag{7}$$

4. Compute the point-wise prior given by

$$\pi_{ij}^k = \frac{\exp(-E_{ij}^k(x \mid \beta^k))}{\sum_h^{|C|} \exp(E_{hj}^k(x \mid \beta^k))} \tag{8}$$

5. Compute the conditional probability by (9) and the distance matrix given by (10), respectively

$$p_i^k(x_j \mid \mu_j^k, \sigma_i^k) = \frac{1}{\sigma_i^k \sqrt{2\pi}} \exp[-\frac{(x_j - \mu_j^k)^2}{(2\sigma_i^k)^2}] \tag{9}$$

$$d_{ij}^k = -\log[p_i^k(x_j \mid \mu_i^k, \sigma_i^k)] \tag{10}$$

6. Compute the membership matrix by

$$u_{ij}^{k+1} = \frac{\pi_{ij}^k \exp(-d_{ij}^k)}{\sum_{h=1}^{|C|} \pi_{hj} \exp(-d_{hj}^k)} \tag{11}$$

7. Compute the objective function and compare the difference of its value given by

$$\left| J_{ij}^{k+1} - J_{ij}^k \right| < T_J \tag{12}$$

before changing the loop ending flag. T_J is the converge threshold.

8. If the loop continues, compute the mean and standard deviation of the conditional distribution which is Gaussian distribution given by

$$\begin{cases} \mu_i^{k+1} = \frac{\sum_{j \in X} u_{ij}^k x_j}{\sum_{j \in X} u_{ij}^k} \\ \sigma_i^{k+1} = \sqrt{\frac{\sum_{j \in X} u_{ij}^k (x_j - \mu_i^{k+1})^2}{\sum_{j \in X} u_{ij}^k}} \end{cases} \tag{13}$$

Then, $k = k + 1$, return to 2.

III. Description of the proposed energy function

As mentioned in the previous sections, the energy function is critical to our algorithm. In order to improve the noise resistance ability, there are two important points. The first one is the way that the neighborhood information is derived. The algorithm should compute the data that represents the correlation between pixels in a certain neighborhood system. The second one is the way the function utilize the neighborhood information which reflects the effectiveness of the algorithm in dealing with outliers. So the description of our energy function is divided into two parts. Namely, the weighted distance and the introduction of the distance into our energy function. The weighted distance consists of two different parts. The first part concerns the Euclidean distance between each pixel in the neighborhood system and the central pixel. So the first part can be obtained by

$$W_s = \frac{1}{d_{ij}+1} \tag{14}$$

Note that this d_{ij} is Euclidean distance, which is different from the one in (10). W_s denotes the spatial correlation between the central pixel and the other neighborhood pixels which have the same class with the central one. The first part does not consider the intensity values around the central pixel. So we have the second part of our weighted distance.

W_g denotes the property of the entire neighborhood. In order to compute W_g, we first compute the deviation mean ratio C_j.

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

where x denote the set of pixel in the entire neighborhood. The ratio, also called index of dispersion, measures the dispersion of a certain dataset. Then we project it into kernel space in order to penalize the distance more effectively.

$$\xi_{ij} = \exp[-(C_j - \bar{C})] \tag{16}$$

We finally normalize the ξ_{ij} by $\eta_{ij} = \frac{\xi_{ij}}{\sum_{k\in\partial_j}\xi_{ik}}$and guarantee the weight to be non-negative by

$$W_g = \begin{cases} 2+\eta_{ij} & C_j < \bar{C} \\ 2-\eta_{ij} & C_j \geq \bar{C} \end{cases} \tag{17}$$

The weighted distance is obtained by

$$W = W_s \cdot W_g \tag{18}$$

We then introduce the weighted distance into our energy function. Mathematically, an energy function is given by:

$$E = \sum_{\{i\}\in C_1} V_1 + \sum_{\{i,i'\}\in C_2} V_2 + \sum_{\{i,i',i''\}\in C_3} V_3 + ... \tag{19}$$

where C_1, C_2, C_3 are different sizes of pixel batches called clique.V_1, V_2, V_3 describe the relationship between pixels in a clique called clique potential. However, it seems useless to model the relationship among three or more elements in an image. So we only consider the mutual dependency in cliques whose size is 1 or 2. We can simply consider clique potential as a mutual dependency between the pixel and its neighbor. Consequently, how to model the dependency is vital for the performance of our algorithm. There are several ways in designing energy function. In [12], the author proposed that the energy of a certain pixel equals the number of neighborhood pixels which have the same class with it. In [15], the energy function of a pixel depends on not only the above number of the same class, but also the specific membership of the entire neighbor together with a presupposed parameter. In the proposed energy function, we use the weighted distance which has been mentioned before. The energy of a pixel is equal to the summation of weighted membership of all the pixels in the neighborhood which share the same class with the central pixel

$$E_{ij}(x \mid \beta) = -\beta \sum_{k\in\partial_j} \frac{u_{ij}}{W_j}\delta(x_j - x_k) \tag{20}$$

where $\delta(\cdot)$ is denoted by

$$\delta(x_j - x_k) = \begin{cases} 1 & \text{if } x_j = x_k \\ 0 & \text{otherwise.} \end{cases}$$

We eventually yield the energy expression. Using (20), the equation (8) can be computed after substituting E_{ij} into it.

IV. Experimental settings

In the experimental part, we test our algorithm in various scenarios to show the effectiveness of it. We use 5×5 window in all the test images.

A. Segmentation on synthetic images

The weighted energy function significantly improves the robustness to noise and is suitable for segmentation in noisy images. In order to show the superiority of our method, we compare our algorithm with several competitors.

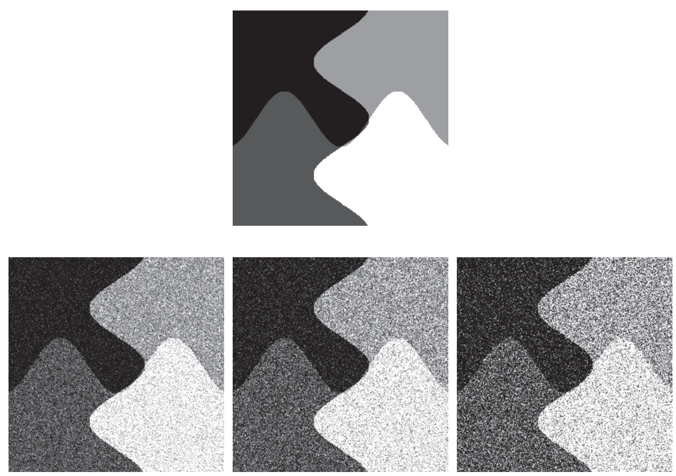

Fig. 1. Upper panel is the original image. Lower panel are images corrupted by Gaussian noise ($\mu = 0$, $\sigma = 0.03, 0.05, 0.10$ respectively).

The segmentation result is given in Fig.2. We can see the proposed method has a better sementation result. In addition, these methods is compared with respect to the segmentation accuracy defined as

$$SA = \frac{N_{correct}}{N} \times 100\%$$

where $N_{correct}$ is the number of correctly segmented pixels. N is the total number of the pixels. The outcome is given in Table I.

Second we use the coins picture to give a comparison between our method with HMRF-FCM and KWFLICM. Because of the background simplicity of this picture, we use it as synthetic image.

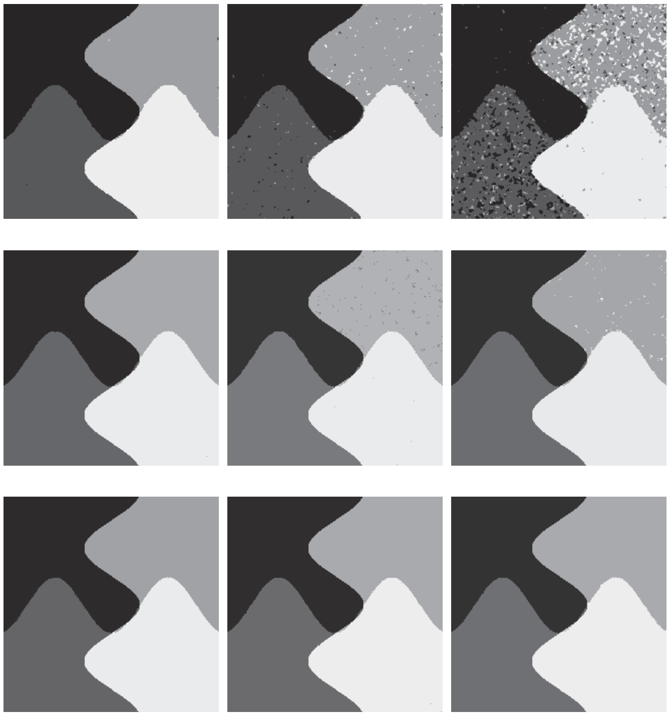

Fig. 2. Segmentation results of the three corrupted images in Fig.1. The outcomes are KWFLICM, HMRF-FCM, and the proposed method by row respectively.

TABLE I
SEGMENTATION ACCURACY OF DIFFERENT METHODS

Noise	0.03	0.05	0.10
KWFLICM	99.84	76.03	75.95
HMRF-FCM	99.97	99.96	99.71
Proposed	**99.98**	**99.98**	**99.93**

In this scenario the noise is significant heavier. But due to the using of contextual information, the outcome of the proposed method is quite stable.

In the comparison between these three algorithms, we can see that because of using the new energy function the proposed method has a better noise resistance than others. When the noise is heavy ($\sigma = 5$ in Fig. 4) our method can still have a good segmentation result.

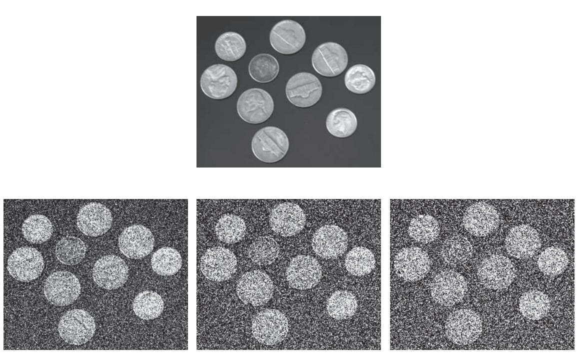

Fig. 3. Upper panel is the original image. Lower panel are images corrupted by Gaussian noise ($\mu = 0$, $\sigma = 0.03, 0.05, 0.10$ respectively).

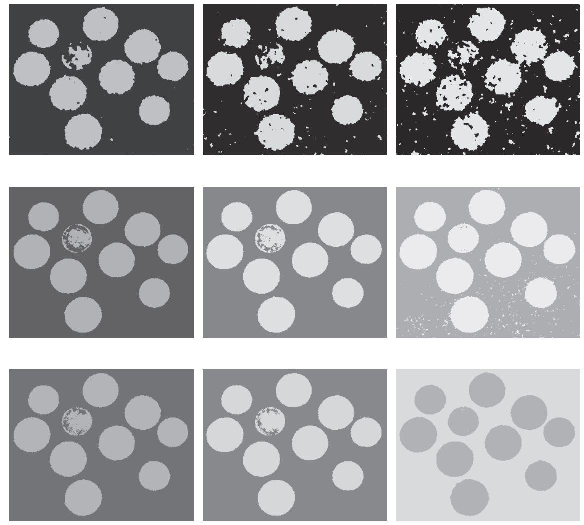

Fig. 4. Segmentation results of the three corrupted images in Fig.3. The outcomes are KWFLICM, HMRF-FCM, and the proposed method by row respectively.

B. Segmentation on natural images

Natural images have more delicate details than the synthetic ones and therefore these natural images are harder to be segmented. We evaluate our algorithm using part of the Berkeley Image Segmentation Dataset. In this experiment, we first test our method on natural images and then the corresponding noise-corrupted ones. The results of noise-free ones are shown in Fig. 5.

Then we test our algorithm on noise-corrupted ones. The original image is corrupted by Gaussian noise ($\mu = 0$, $\sigma = 0.30$). The noise is heavy, so we can fully evaluate the improvement of our energy function in using context information. Results are in Fig. 6.

We can see that our method largely preserves the original segmentation results and the contour of each result is still distinct. The speckles and noisy points are mostly in low-contrast area.

Finally we compare all the three algorithms: KWFLICM, HMRF-FCM and the proposed method on the above ostrich image. The noise is also Gaussian ($\mu = 0$, $\sigma = 0.30$).

Fig. 7. Segmentation results on the noisy ostrich image. From left to right is the result of KWFLICM, HMRF-FCM and the proposed method.

In Fig. 7, we can see that the proposed method has fewer speckles in the segmentation result than the HMRF-FCM's. In the detailed images (see Fig. 8), the ostrich's beak is

Fig. 5. Segmentation results on part of Berkeley Segmentation Dataset.

Fig. 6. Segmentation results on part of Berkeley Segmentation Dataset corrupted by Gaussian noise ($\mu = 0$, $\sigma = 0.30$).

Fig. 8. Details of Fig. 7.

clearer in our result. The KWFLICM's performance is worse because the ostrich's left eye becomes indistinguishable and the background is blurred too much.

V. Conclusion

In this paper, we first discuss the effectiveness of introducing Markov Random Field in FCM-based segmentation algorithms. Then the fusion of MRF and FCM is elucidated. A new energy function is proposed in order to further utilize the contextual information. Theoretical analysis is presented and experimental results show that our algorithm outperforms the comparison algorithms in segmentation on both noise-free and noisy image.

References

[1] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of mri data," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 3, pp. 193–199, 2002.

[2] S. Chen and D. Zhang, "Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, no. 4, pp. 1907–1916, 2004.

[3] D.-Q. Zhang, S. Chen, Z.-S. Pan, and K.-R. Tan, "Kernel-based fuzzy clustering incorporating spatial constraints for image segmentation," in *Machine Learning and Cybernetics, 2003 International Conference on*, vol. 4. IEEE, 2003, pp. 2189–2192.

[4] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy< i> c</i>-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognition*, vol. 40, no. 3, pp. 825–838, 2007.

[5] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *Image Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1328–1337, 2010.

[6] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 2141–2151, 2012.
[7] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy c-means clustering with local information and kernel metric for image segmentation," *Image Processing, IEEE Transactions on*, vol. 22, no. 2, pp. 573–584, 2013.
[8] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 721–741, 1984.
[9] S. Z. Li and S. Singh, *Markov random field modeling in image analysis*. Springer, 2009, vol. 3.
[10] D. Chandler, "Introduction to modern statistical mechanics," *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771*, vol. 1, 1987.
[11] G. Celeux, F. Forbes, and N. Peyrard, "EM procedures using mean field-like approximations for markov model-based image segmentation," *Pattern recognition*, vol. 36, no. 1, pp. 131–144, 2003.
[12] S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 5, pp. 1351–1361, 2008.
[13] T. Celik and H. K. Lee, "Comments on "a robust fuzzy local information c-means clustering algorithm"," *Image Processing, IEEE Transactions on*, vol. 22, no. 3, pp. 1258–1261, 2013.
[14] H. Ichihashi, K. Miyagishi, and K. Honda, "Fuzzy c-means clustering with regularization by kl information," in *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 924–927.
[15] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified mrf energy function for change detection in synthetic aperture radar images," *Fuzzy Systems, IEEE Transactions on*, vol. 22, no. 1, pp. 98–109, Feb 2014.