Segmentation of M-FISH Images for Improved Classification of Chromosomes With an Adaptive Fuzzy C-means Clustering Algorithm

Hongbao Cao, Hong-Wen Deng, and Yu-Ping Wang, *Senior Member, IEEE*

***Abstract*—An adaptive fuzzy c-means algorithm was developed and applied to the segmentation and classification of multicolor fluorescence *in situ* hybridization (M-FISH) images, which can be used to detect chromosomal abnormalities for cancer and genetic disease diagnosis. The algorithm improves the classical fuzzy c-means algorithm (FCM) by the use of a gain field, which models and corrects intensity inhomogeneities caused by a microscope imaging system, flairs of targets (chromosomes), and uneven hybridization of DNA. Other than directly simulating the inhomogeneously distributed intensities over the image, the gain field regulates centers of each intensity cluster. The algorithm has been tested on an M-FISH database that we have established, which demonstrates improved performance in both segmentation and classification. When compared with other FCM clustering-based algorithms and a recently reported region-based segmentation and classification algorithm, our method gave the lowest segmentation and classification error, which will contribute to improved diagnosis of genetic diseases and cancers.**

***Index Terms*—Adaptive fuzzy c-means (AFCM) clustering, background correction, image segmentation, multicolor fluorescence *in situ* hybridization (M-FISH) image classification.**

I. Introduction

MULTICOLOR fluorescence *in situ* hybridization (M-FISH) is a combinatorial labeling technique that is developed for the analysis of human chromosomes [1], [2]. The technique has been used for the characterization of chromosomal translocations, to search for cryptic rearrangements, and to study mutagenesis, tumors, and radiobiology [3]. In this technology, chromosomes are labeled with five dyes and a DNA stain known as 4'-6-diamidino-2-phenylindole (DAPI) that attaches to DNA and labels all chromosomes. A fluorescent microscope that is equipped with a filter wheel is used to capture the chromosome images. Each dye is visible in a particular wavelength and can be captured by the use of a specific filter. Therefore, M-FISH signals can be obtained as multispectral or multichannel images, in which a chromosome was stained to be visible (signed as "1") or not visible (signed as "0"). For a number n, the number of Boolean combination is 2^n. Hence, five spectra are sufficient to distinguish the 24 classes of chromosomes in human genome. In addition to that, DAPI is used to counterstain each chromosome such that all of the chromosomes are visible in a DAPI channel. By simultaneously viewing six different channel images, pixel-wise classification of human chromosome is possible. This technique is also called color karyotyping in cytogenetics [1]. Fig. 1 shows an example of M-FISH images of a male cell, where 22 autosomes and both sex chromosomes are classified from a five-channel spectral image data and are displayed by the use of 24 pseudocolors. For a normal cell, each chromosome should be painted with the same color. Otherwise, it indicates that chromosomal abnormalities might exist, which are associated with certain genetic diseases and cancers.

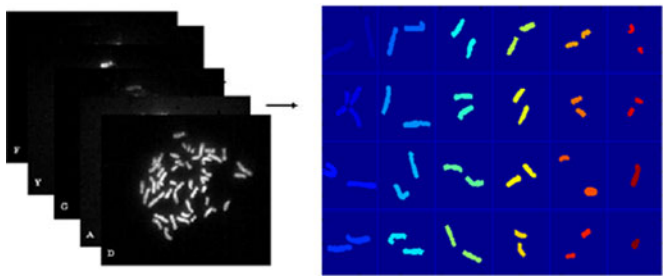

Fig. 1. Twenty-four classes of chromosomes are classified from the five-channel spectral images; each class of chromosome is displayed with a different pseudocolor. This pixel-wise classification technique is called color karyotyping.

The detection of chromosomal abnormalities depends on accurate pixel-wise classification techniques. Even though many attempts have been made to automate image analysis procedure [4]–[9], the reliability of the diagnosis technique has not reached the level for clinical use due to a number of factors that include nonhomogeneity of staining, variations of intensity levels within and between image sets, and emission spectral overlaps between fluorophores [8]–[11]. The sizes of the misclassified regions are often larger than the actual chromosomal rearrangements or lost, which often lead to incorrect interpretation by cytogeneticists. To improve the detection of chromosomal abnormalities for clinical diagnosis, accurate segmentation and classification algorithms have to be developed.

Manuscript received October 11, 2011; revised February 3, 2011 and May 5, 2011; accepted May 17, 2011. Date of publication June 20, 2011; date of current version February 7, 2012. This work was supported by the National Institutes of Health under Award 1R15GM088802-01 and by the Shanghai Eastern Scholarship Program.

H. Cao is with the Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118 USA (e-mail: hcao3@tulane.edu).

H. W. Deng is with the Department of Biostatistics, Tulane University, New Orleans, LA 70118 USA (e-mail: hdeng2@tulane.edu).

Y.-P. Wang is with the Department of Biomedical Engineering and Department of Biostatistics, Tulane University, New Orleans, LA 70118 USA, and also with the Shanghai University for Science and Technology, Shanghai 200444, China (e-mail: wyp@tulane.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2011.2160025

In this paper, an adaptive fuzzy c-means (AFCM) algorithm was developed and applied to the classification of M-FISH images by considering intensity inhomogeneities, which often exist in the images. Different from the previous AFCM algorithm that is proposed for MRI image analysis [12], [13], we proposed an improved AFCM classification method (IAFCM) with a new objective function, which yields better background compensation and results in improved chromosome segmentation and classification. This is also different from other FCM-based algorithms, which directly simulate the bias field [14]–[17].

II. Related Work

The algorithms for classification of M-FISH images can be categorized into two groups: the pixel-by-pixel classification [18]–[22] and the region-based classification [7], [8], [23]–[26]. In the pixel-by-pixel classification algorithms, even with preprocessing and postprocessing, the classification accuracy is still not high enough (less than 90%) [4], [7], [9], [22], [26]. It was shown in [7] that the pixel-by-pixel classification was dominated by image inhomogeneities, and the average classification accuracy was only 68% with a standard deviation of 17.5%. In a recently proposed multichannel region-based method by Karvelis *et al*., the overall chromosome classification accuracy reaches 82.4% $\pm$ 14% for the database built by us [27]. However, these existing methods still cannot have sufficient accuracy for clinical use [7], [8], [23]–[26].

Among many factors that cause uneven distributed intensities in M-FISH images, there are three important ones.

1) *System error caused by the microscope system:* Because of the optical imaging, a microscope image always has a much brighter center than its surroundings. As a result, the intensities of chromosomes at the surroundings are much lower than those at the center.
2) *The flair effects of the chromosomes:* Background intensity near the chromosome cluster is usually higher than that of the areas far away from the chromosome cluster [4]. Because of this reason, the background near a chromosome may be clustered as "target" rather than background.
3) *The uneven hybridization within a chromosome:* Because of this reason, the intensity of the same chromosome may vary greatly. These undesired intensity inhomogeneities would affect subsequent classification accuracy.

In order to avoid the influences of background inhomogeneity on the segmentation and classification of chromosomes, several region-based segmentation and classification algorithms were developed [23], [24], which require the proper setting of complicated parameters. Recently, Karvelis *et al*. proposed a multichannel-region-based segmentation and classification method [8], [25], [26]. Different from the pixel-by-pixel classification, this method took the spatial context into consideration and achieved an overall accuracy as high as 89% on nonoverlapping chromosome images and 82.4% for the whole dataset by the use of an M-FISH chromosome image database established by us [27]. However, their methods still depend on preprocessing or postprocessing steps to overcome the oversegmentation problems. In addition, they used Otsu's method for the segmentation of DAPI channel image to find the chromosome region, which has been proven to not be reliable in this paper.

Our recent work has shown that fuzzy c-means (FCM) clustering-based algorithms can provide classification accuracy as high as over 89% [19]–[21]. The FCM can be obtained by the minimization of the following objective function [28]:

$$J_{\mathrm{FCM}} = \sum\nolimits_{i\in D} \sum\nolimits_{k=1}^{\mathrm{NC}} u_{ik}^{q} \left\| y_i - c_k \right\|^2 \tag{1}$$

subject to

$$\sum\nolimits_{k=1}^{\mathrm{NC}} u_{ik} = 1 \tag{2}$$

where u_{ik} is the membership function with values between 0 and 1; c_k is the cluster center; q is a weighting exponent on each fuzzy membership and determines the amount of fuzziness; D is the area of image; NC is the number of clusters; $\|*\| = \sqrt{\langle *, * \rangle}$; and $\langle a, b \rangle$ represents the inner product of vectors a and b.

There have been many improvements over the classical FCM algorithm [12]–[17]. Most of those improvements focused on simulation and correction of the slowly changing bias field of an image and were mostly applied to MRI processing, which produced improved image segmentation results.

Among improved FCM methods, Pham and Prince [12], [13] proposed an AFCM method that used a gain field to modify the centers of each cluster and to compensate the slowly changing inhomogeneities effects. In their method, they employed the energy of the first and second derivatives of the gain field to control the smoothness of the gain field. The objective function that they proposed could be expressed by

$$J_{\mathrm{AFCM}} = \sum_{i\in D} \sum_{k=1}^{\mathrm{NC}} u_{ik}^{q} \left\| g_i y_i - C_k \right\|^2 + \lambda_1 \sum\nolimits_{i\in D} \left(G_i'\right)^2 + \lambda_2 \sum\nolimits_{i\in D} \left(G_i''\right)^2 \tag{3}$$

where $G = \{g_i\}$ is the gain field, and $\left(G_i'\right)^2$ and $\left(G_i''\right)^2$ are the energies of the first and second derivatives of the gain field, respectively, which are used to control the smoothness of the gain field. For the definition of other parameters in (3), see [12] and [13].

The AFCM algorithm that is proposed by Pham and Prince took the spatial context of the image into consideration, which is desirable for compensating background inhomogeneities. The work has shown that AFCM segmentation yields lower error rates than that of the classical FCM algorithm when segmenting MR brain images with intensity inhomogeneities [12], [13]. However, when applying this algorithm to an M-FISH image, we found that the shape of a gain field is not always effective to compensate the intensity inhomogeneity. When the gain field is too sharp, it fails to compensate the slow changes (intensity inhomogeneities that are caused by uneven illumination); when the gain field is too smooth, it fails to correct the local changes (intensity inhomogeneities that are caused by the flair effects of the chromosomes and the uneven hybridization within a chromosome). Thus, a gain field that can compensate both the local intensity changes and the slow intensity changes is desirable. In

addition, the method that they proposed takes very long computational time in each of its iteration (90 s by the use of a laptop with dual CUP T3400 at 2.16 GHz and RAM of 4 GB, compared with 14 s by the use of our proposed method), because it involves the solution of large-scale differential equations.

In this paper, an IAFCM segmentation algorithm was introduced and applied to the classification of M-FISH images. Different from the existing AFCM algorithm that is discussed earlier, the proposed IAFCM algorithm used a new objective function with a different regulation term, which appears to be more effective in controlling the shape of the gain field. Both AFCM and the proposed IAFCM in this paper are seeking an optimum gain field that can compensate the background intensity inhomogeneity. The difference is that the IAFCM we proposed in this paper employed the variances of a given point to its local area as a regularization term in controlling the gain field, while AFCM used the energy of first and second derivatives of a given point for the regularization, as is shown in (3). In our IAFCM algorithm, the regularization term uses the approximation of the first-order derivative with a filter, which can preserve the shape of the gain filed while suppressing noise. In addition, the proposed algorithm avoids solving large differential equations and gives much faster computational speed. In order to evaluate the performance of the algorithm, we compared it with FCM, AFCM, Otsu's method, and a recently reported region-based method for M-FISH image segmentation and classification algorithm using the same database established by us [12], [13]. Results from the testing of our database have shown that IAFCM increased both the segmentation and classification accuracy.

III. Improved Adaptive Fuzzy C-Means-Based Multicolor Fluorescence *in situ* Hybridization Image Segmentation and Classification

A. Formulation of Improved Adaptive Fuzzy C-Means

The objective function of the proposed IAFCM is introduced as follows:

$$J_{\mathrm{IAFCM}} = \sum_{i\in D}\sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\|y_i - g_i c_k\right\|^2 + \lambda \sum_{i\in D}\left(g_i - (H * g)_i\right)^2 \tag{4}$$

where u_{ik} is the membership function with positive values between 0 and 1; y_i is the observed image intensity at location i; c_k is the cluster centers; q is a weighting exponent on each fuzzy membership, which determines the amount of fuzziness; D is the whole area of image; NC is the number of clusters; $\{g_i | i \in D\}$ is the gain field to be found; and H is a $(2r+1)*(2r+1)$ average convolution kernel given by

$$H = 1/N \begin{bmatrix} 1 & \dots & 1 \\ \vdots & 1 & \vdots \\ 1 & \dots & 1 \end{bmatrix} \tag{5}$$

where $N = (2r+1)*(2r+1)$ is the number of pixels within area D_{ir}. Let us write the regulation term in (4) as J_G in

$$J_G = \sum_{i\in D}\left(g_i - (H * g)_i\right)^2 \tag{6}$$

and define the modified FCM objective function J_Y as

$$J_Y = \sum_{i\in D}\sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\|y_i - g_i c_k\right\|^2. \tag{7}$$

Thus, (4) could be rewritten as follows:

$$J_{\mathrm{IAFCM}} = J_Y + \lambda J_G. \tag{8}$$

The regulation term J_G is different from that was used by Pham and Prince's work [12], [13], as explained earlier. Controlled by the coefficient λ, J_G regularize the shape of the gain field.

B. Algorithm of Improved Adaptive Fuzzy C-Means

In Pham and Prince's work, large coefficients λ for regulation term J_G were selected, and then, the optimum algorithm was developed to find g_i, c_k, and u_{ik} such that J_{IAFCM} is minimized. In this paper, the objective function that is given by (8) is taken as a conditional minimization problem with the constraint function, which can be formulated as Lagrange multipliers form as J_G. To solve this conditional minimization problem, a necessary condition is that the gradient is zero

$$\begin{cases} \dfrac{\partial J_{\mathrm{IAFCM}}}{\partial u_{ik}} = 0 \\ \dfrac{\partial J_{\mathrm{IAFCM}}}{\partial c_k} = 0 \\ \dfrac{\partial J_{\mathrm{IAFCM}}}{\partial g_i} = 0 \\ \dfrac{\partial J_{\mathrm{IAFCM}}}{\partial \lambda} = 0, \quad i \in D \end{cases} \tag{9}$$

which lead to the following equations:

$$u_{ik} = \frac{\left\|y_i - g_i C_k\right\|^{-\frac{2}{q-1}}}{\sum_{l=1}^{\mathrm{NC}}\left\|y_i - g_i C_l\right\|^{-\frac{2}{q-1}}} \tag{10}$$

$$c_k = \frac{\sum_{i\in D} u_{ik}^{q} G_i y_i}{\sum_{i\in D} u_{ik}^{q} G_i^2} \tag{11}$$

$$\times\, g_i \sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\langle C_k, C_k\right\rangle - \sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\langle y_i, C_k\right\rangle + \frac{\lambda}{N}\sum_{j\in D_{ir}}\left\{\left(g_i - (H * g)\, i\right) - \left(g_j - (H * g_j)\right)\right\} = 0$$

$$i \in D \tag{12}$$

$$g_i = (H * g)_i, \qquad i \in D. \tag{13}$$

By substitution of (13) into (12), (12) can be rewritten as follows:

$$g_i = \frac{\sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\langle y_i, C_k\right\rangle}{\sum_{k=1}^{\mathrm{NC}} u_{ik}^{q}\left\langle C_k, C_k\right\rangle}. \tag{14}$$

The solution of (10), (11), (13), and (14) gives the optimum value of (u_{ik}, g_i, c_k), which lead to the algorithm described as IAFCM.

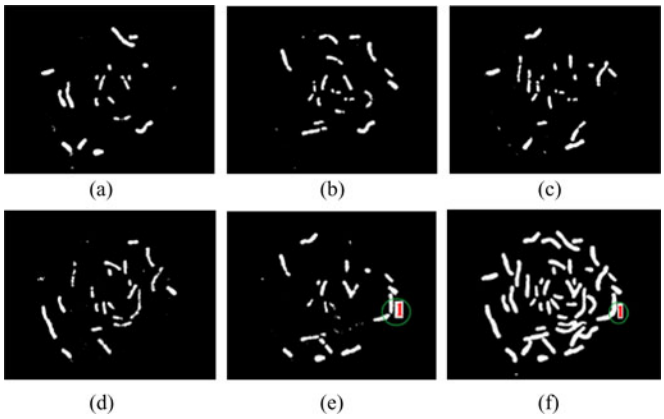

Fig. 2. Segmentation of M-FISH images at each channel. (a)–(e) Segmentation results for channels (1)–(5); (f) segmentation result of the DAPI channel.

IAFCM Algorithm:

1) Initialize g_i with 1 ($i = 1, \ldots, N$) and cluster centers c_k ($k = 1, \ldots,$ NC) with random values within the image intensity, where NC is the number of clusters.
2) Update the membership function u_{ik} by using (10).
3) Update the cluster centers C_k by using (11).
4) Calculate the gain field g_i by using (14).
5) Update the gain field g_i by using (13).

If the maximum change of $u_{ik} <$ tolerance U and the maximum change of $g_i <$ tolerance G, break. Otherwise, go to step 2).

In our paper, the classification of the 24 classes of chromosomes was realized by two steps: first, we segmented each of the five channels into two clusters (background and foreground); second, we employed the combinatorial labeling technique to assign the class labels to each pixel (see the following section for more details). Thus, for the segmentation stage, we took NC $= 2$, tolerance $U = 0.01$, and tolerance $G = 0.1$. The initialization of this improved FCM algorithm does not need special treatment, as is stated in step 1).

C. Image Classification

According to the combinatorial labeling technique that is developed for the analysis of human chromosomes [1], [2], once images of each channel were correctly segmented, the classification can be easily performed by the use of the binary combination. In this paper, the DAPI channel image was first segmented with the IAFCM algorithm to generate a chromosome mask. The mask was then applied to all other five channels so that the same background (i.e., nonchromosome regions) were identified. After image segmentation, each pixel in an M-FISH image set is labeled as $x_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}]$, where $x_{ij} \in \{0,1\}$, $i = 1, 2, \ldots, N$; $j = 1, \ldots, 5$; and N is the number of pixels in the image. Class label will be assigned to each pixel according to the binary table. For example, pixels that are labeled as [0, 0, 0, 0, 1] will be set as class 1; pixels that are labeled as [0, 0, 0, 1, 0] will be set as class 2, etc. Fig. 2 gives an example of the segmentation stage of M-FISH images for each channel. In Fig. 2, pixels of the green circled chromosome [see Fig. 2(e) and (f)] will be assigned as class 1 since they are labeled as [0, 0, 0, 0, 1].

TABLE I
SEGMENTATION RESULTS FROM IAFCM, AFCM, FCM, OTSU'S METHODS AND THE REGIONAL SEGMENTATION METHOD

Methods (# of images)	IAFCM (120)	AFCM (120)	FCM (120)	Otsu (120)	Regional (15)	Regional (183)
CR(%)	89.5±10.5	96.5±4.6	92.0±8.4	89.1±9.2	83.6±9.9	82±12
FR(%)	3.6±2.8	20.9±12.9	9.7±7.7	11.7±8.7	NA	NA

IV. RESULTS

M-FISH images from M-FISH database [27] of 20 cells (9 males and 11 females) with 120 images were tested, and the results of both image segmentation and classification were compared over the proposed IAFCM algorithm and two existing algorithms, i.e., AFCM and FCM methods. In addition, Otsu's segmentation [29] results and the segmentation and classification results that use the same database reported by Karvelis *et al*. were also listed for the purpose of comparison.

A. Image Segmentation Results

The segmentation of M-FISH images by the use of our proposed IAFCM followed the steps that are described in Section III-C, which generated the mask using DAPI channel image first followed by the segmentation of five other channels. The performance of the segmentation was evaluated with the correct detection rate (CR) and false detection rate (FR), which are defined by the following equations:

$$\text{CR} = \frac{\#\text{ chromosome pixels correctly segmented}}{\#\text{ total chromosome pixels}} \tag{15}$$

$$\text{FR} = \frac{\#\text{ background pixels segmented as chromosome}}{\#\text{ total chromosome pixels}}. \tag{16}$$

From the definition of CR and FR, it can be seen that a good segmentation should give a higher CR but a lower FR. We calculated these values from 20 cells with 120 with four different methods, and the results are listed in Table I. In Table I, the segmentation results of Otsu's method and the results recently reported by Karvelis *et al*. [8], [25], [26] were also listed, which tested on the same database that we established by the use of a region-based segmentation method.

It was reported by Karvelis *et al*. [8], [26] that the CR was 83.59% ± 9.89% for 15 none overlapping M-FISH images, and 82% ± 12% when the number of M-FISH images is 183 (excluding 17 images from the 200 image sets, which were reported as "difficult to karyotype"). In addition, the IAFCM outperformed AFCM and FCM methods by giving lowest FRs. FR was not reported in the Karvelis *et al*. work.

Fig. 3 gives an example of the segmentation results from the four methods. In Fig. 3, we could see that the result of using IAFCM [see Fig. 3(c)] is much better than those of AFCM, FCM, and Otsu's methods as in Fig. 3(d)–(f), since IAFCM gives lowest FR with a relative high CR. Furthermore, by comparing

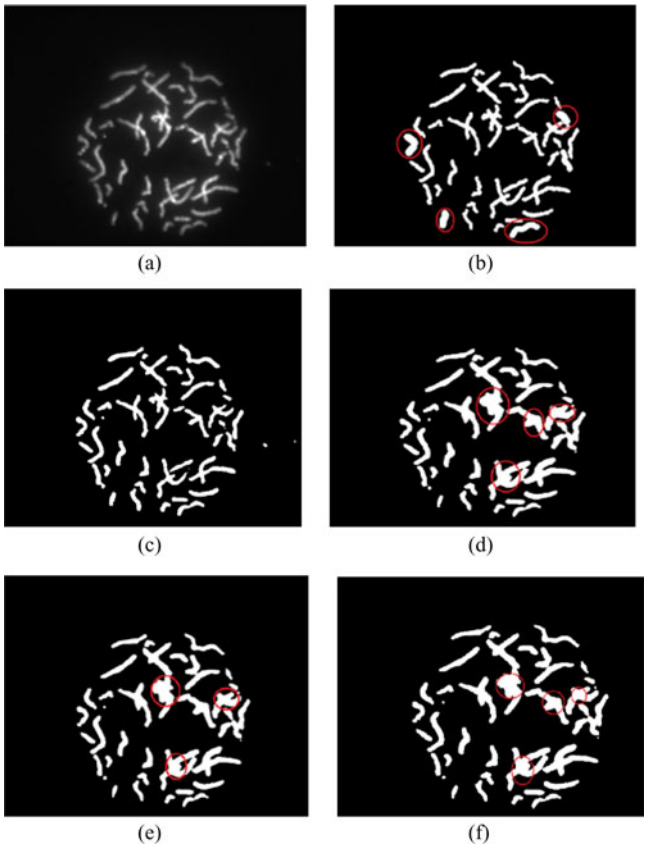

Fig. 3. Comparison of FR and CR among four methods. (a) DAPI channel. (b) Ground truth. (c) IAFCM with CR = 87.20%, FR = 8.25%. (d) AFCM with CR = 92.73%, FR = 27.03%. (e) FCM with CR = 92.20%, FR = 24.58%. (f) Otsu's method with CR = 92.73%, FR = 25.38%.

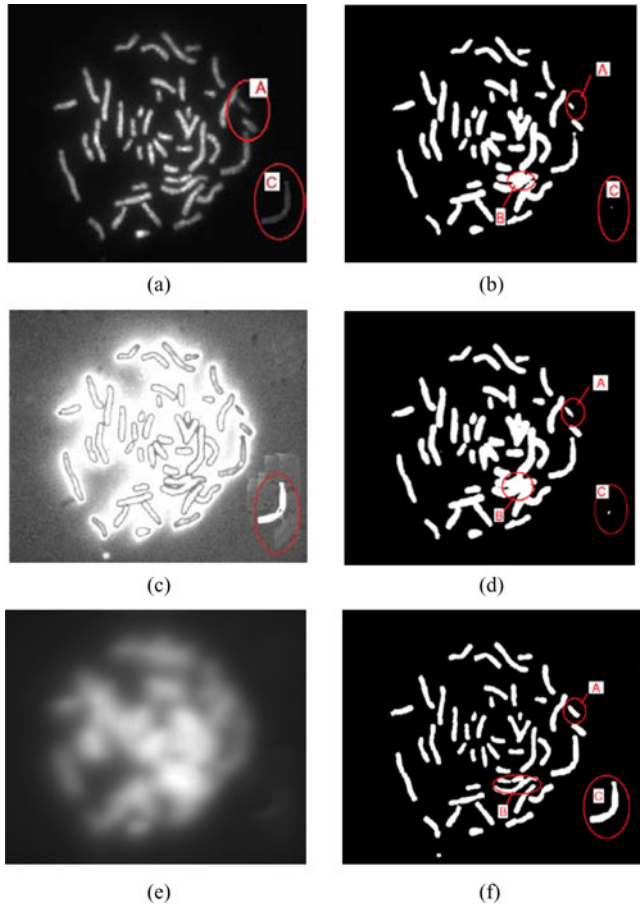

Fig. 4. Segmentation results from methods of FCM, AFCM, and IAFCM. (a) DAPI channel. (b) FCM segmentation result. (c) Gain field G of AFCM. (d) AFCM segmentation result. (e) Gain field G of IAFCM. (f) IAFCM segmentation result.

the circled areas in Fig. 3(a) and (c), we can see that IAFCM is not oversegmented. The AFCM, FCM, and Otsu's methods achieved relative higher segmentation CR by less segmentation (high FR), which will not work in the following cases that are shown in Fig. 3 [see the red circled area in Fig. 3(d)–(f)].

To further compare the difference between each segmentation method, we designed an experiment, in which a chromosome with low intensity (around 60) was added into the original DAPI channel (indicated in Fig. 4(a), red-circled area C) in such a way that its intensity contrast to its local background was about the same as that of area A. Therefore, when the chromosomes at area A were segmented, the chromosome at C should also be segmented because they have similar local contrasts.

As shown in Fig. 4(a), although the intensity of the chromosome at location C was low, it should be clearly identified by a trained cytogeneticist. Fig. 4(b) shows the segmentation results from the FCM method, in which the chromosomes in both area C and area A were almost lost. This is due to the fact that FCM-based segmentations are dependent on intensity at a single pixel. The AFCM-based segmentation result [see Fig. 4(d)] is relatively better than FCM-based method. It covers the chromosome at area A by taking spatial contextual information into consideration. Fig. 4(c) shows the gain field generated by the AFCM method, which corrected part of the inhomogeneities. However, it failed to find the whole chromosome at area C [see Fig. 4(d)]. The IAFCM segmentation result [see Fig. 4(f)] not only found the chromosomes at area C and area A but "cleaned up the mess" that existed in two other methods as well [see the red circle in the area of B in Fig. 4(f)]. This might be because the gain field [see Fig. 4(e)] in the IAFCM method is better shaped than that of AFCM, which compensates both macroscopical intensity variations and local intensity changes.

B. Multicolor Fluorescence in situ Hybridization Images Classification Results

After the image segmentation, each pixel at position i was assigned with a five-dimensional feature vector $x_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}]$ from the five color channels. Then, class labels were assigned according to the combinatorial labeling table [1], [2].

The chromosome classification accuracy of IAFCM was compared with that of AFCM and FCM, which showed a significant difference with p-values of 0.018 and 0.069, respectively. Table II gives the classification ratios of three methods for the 20 tested cells with 120 images, which are representative in our database. Considering that most of the pixels in the image are in

TABLE II
CHROMOSOME CLASSIFICATION ACCURACY BY THE USE OF IAFCM-, AFCM-, AND FCM-BASED CLASSIFICATION ALGORITHMS

Results / N.O.	IAFCM classification ratio (%)	AFCM classification ratio (%)	FCM classification ratio (%)
1	95.5	93.9	91.7
2	91.5	87.5	88.0
3	97.0	93.8	93.4
4	85.9	83.0	84.2
5	96.7	90.5	94.5
6	84.2	76.6	83.9
7	82.7	81.9	59.7
8	93.1	81.5	74.1
9	93.1	84.6	78.5
10	90.4	85.6	88.9
11	89.4	83.7	81.4
12	93.2	89.5	96.6
13	87.2	77.9	87.7
14	83.1	91.7	88.8
15	85.9	82.3	78.6
16	79.4	79.8	79.5
17	80.4	82.3	86.5
18	91.7	84.1	90.5
19	81.7	73.6	79.2
20	87.9	79.7	77.0
Ave.	88.5	84.2	84.1
Std.	5.5	5.6	8.5

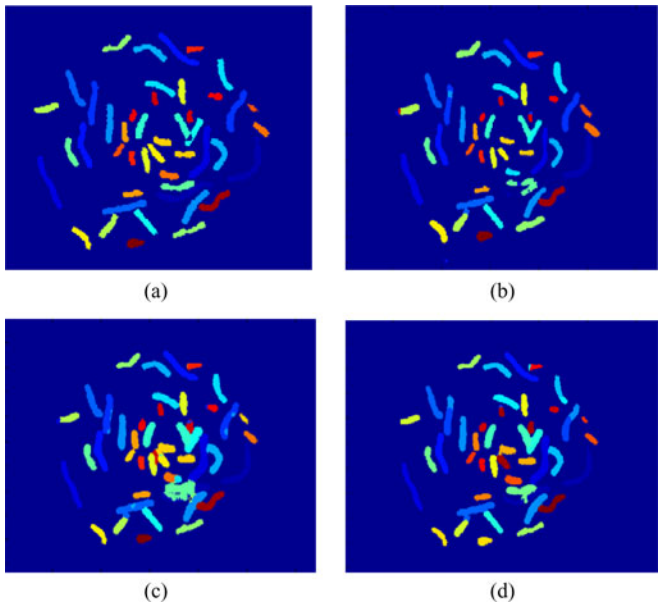

Fig. 5. Classification results for the one M-FISH data tested. (a) Ground truth. (b) IAFCM result. (c) AFCM result. (d) FCM result.

the background (which can be defined as a separate class), we only classify pixels in the chromosome region (24 classes) to evaluate the performance of each algorithm. The classification ratio is defined by

$$\text{Classification ratio} = \frac{\#\text{ chromosome pixels correctly classified}}{\#\text{ total chromosome pixels}}. \quad (17)$$

We have tested and compared these approaches with our established M-FISH image datasets [27]. Fig. 5 gives an example of classification results for one M-FISH dataset tested.

V. DISCUSSION AND CONCLUSION

In chromosome classification with M-FISH imaging, image segmentation is one of the most important steps. In order to increase the classification accuracy, image segmentation has to be improved, which would otherwise significantly affect the subsequent classification accuracy.

An important factor that influences the accuracy of image segmentation is the intensity inhomogeneity or the so-called shading artifacts. In microscope chromosome images, these shading artifacts mainly come from the image acquirement, uneven hybridization, and chromosome flairs. Many background correction methods have been developed and applied to M-FISH chromosome segmentation [4], [9], [22]. Recently, Karvelis *et al*. developed a region-based image segmentation algorithm to avoid the influence of shading artifacts [8], [26]. There are also algorithms developed in MRI processing that can perform background correction and image segmentation simultaneously [12]–[17]. However, those algorithms mainly focus on the correction of inhomogeneous background that smoothly and slowly vary through the image space. For the local uneven intensity variations within and around the chromosomes, which are caused by uneven hybridization and chromosome flairs, we need a better shaped gain field to simulate and compensate intensity inhomogeneity during the image segmentation process. Instead of simulating the background directly, the gain field introduced in [12] and [13] compensated the intensity variation by modifying the centers of each cluster, which showed advantage in image segmentation. However, it used only the information of nearest neighbors of a pixel in the regularization term to control the shape of the gain field. Our proposed IAFCM clustering algorithm also used the gain field but with an improved regularization term for the gain field [see (6)]. Instead of using the energy of first and second directives of the gain field to control the gain field shape, we used the approximation of the first derivative with a filter, which can preserve the shape of the gain filed while suppressing the noise.

The selection of filter size r (the radius of local area D_{ir}) is dependent on the desired target size for which the intensity variation is supposed to be compensated. When r is small, the shape of the gain field will be sharp, and a more detailed background will be compensated. On the other hand, when r is big, the shape of the gain field will be smooth, and slow change of background through large regions will be better corrected. When $r = r_{\max}$ that makes $D_{ir} = D$, the gain field will be the whole image plane, and the IAFCM method will become the FCM method. In this paper, we have found that $r = 30$ works well for all the images in our database [27].

Our experiment showed that the proposed IAFCM algorithm has the advantages over AFCM, FCM, and Otsu's methods in image segmentation with a lowest false detection ratio. In addition, for the 20 cells with 120 images tested, it gave higher chromosome segmentation accuracy than that of the region-based segmentation method, which is recently proposed by Karvelis

et al. [8], [26] (the CR was 83.59% $\pm$ 9.89% for 15 none overlapping M-FISH images, and 82% $\pm$ 12% when the number of M-FISH images is 183).

Although the proposed IAFCM method gave the highest classification accuracy among the existing classifiers tested, it has not employed any preprocessing and postprocessing steps. Some postprocessing methods such as the joint segmentation-classification that is proposed by Schwartzkopf *et al*. [7] and preprocessing methods such as the color compensation that is proposed by Choi *et al*. [9] can be incorporated to further increase the classification accuracy.

REFERENCES

[1] M. R. Speicher, S. G. Ballard, D. C. Ward, and Karyotyping, "Human chromosomes by combinatorial multi-fluor FISH," *Nat. Genet*., vol. 12, pp. 368–375, 1996.

[2] E. Schrock, S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, M. A. Ferguson-Smith, Y. Ning, D. H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini, and T. Ried, "Multicolor spectral karyotyping of human chromosomes," *Science*, vol. 273, pp. 494–497, 1996.

[3] T. Liehr and U. Claussen, "Multicolor-fish approaches for the charaterization of human chromosomes in clinical genetics and tumor cytogenetics," *Curr. Genom*., vol. 3, pp. 213–235, 2002.

[4] H. Choi, K. R. Castleman, and A. C. Bovik, "Joint segmentation and classification of M-FISH chromosome images," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., San Francisco, CA, Sep. 2004, pp. 1636–1639.

[5] M. P. Sampat, A. C. Bovik, J. K. Aggarwal, and K. R. Castleman, "Supervised parametric and non-parametric classification of chromosome images," *Pattern Recog*., vol. 38, pp. 1209–1223, Aug. 2005.

[6] Y. Wang and K. R. Castleman, "Normalization of multicolor fluorescence in situ hybridization (M-FISH) images for improving color karyotyping," *Cytometry*, vol. 64, pp. 101–109, Apr. 2005.

[7] W. C. Schwartzkopf, A. C. Bovik, and B. L. Evans, "Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images," *IEEE Trans. Med. Imag*., vol. 24, no. 12, pp. 1593–1610, Dec. 2005.

[8] P. S. Karvelis, A. T. Tzallas, D. I. Fotiadis, and I. Georgiou, "A multichannel watershed-based segmentation method for multispectral chromosome classification," *IEEE Trans. Med. Imag*., vol. 27, no. 5, pp. 697–708, May 2008.

[9] H. Choi, K. R. Castleman, and A. C. Bovik, "Color compensation of multicolor FISH images," *IEEE Trans. Med. Imag*., vol. 28, no. 1, pp. 129–135, Jan. 2009.

[10] C. Lee, D. Gisselsson, C. Jin, A. Nordgren, D. O. Ferguson, E. Blennow, J. A. Fletcher, and C. C. Morton, "Limitations of chromosome classification by multicolor karyotyping," *Amer. J. Hum. Genet*., vol. 68, pp. 1043–1047, 2001.

[11] H. Choi, K. R. Castleman, and A. C. Bovik, "Segmentation and fuzzy-logic classification of M-FISH chromosome images," in *Proc. IEEE Int. Conf. Image Process*., Atlanta, GA, Oct. 2006, pp. 69–72.

[12] D. L. Pham and J. L. Prince, "An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recog. Lett*., vol. 20, pp. 57–68, 1998.

[13] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag*., vol. 18, no. 9, pp. 737–752, Sep. 1999.

[14] 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," *IEEE Trans. Med. Imag*., vol. 21, no. 3, pp. 193–199, Mar. 2002.

[15] L. Jiang and W. Yang, "A modified fuzzy c-means algorithm for segmentation of magnetic resonance images," in *Proc. 7th Int. Conf. Digital Image Comput.: Tech. Appl*., 2003, pp. 225–232.

[16] A. W.-C. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Trans. Med. Imag*., vol. 22, no. 9, pp. 1063–1075, Sep. 2003.

[17] R. He, S. Datta, B. R. Sajja, and P. A. Narayana, "Generalized fuzzy clustering for segmentation of multi-spectral magnetic resonance images," *Comput. Med. Imag. Graph*, vol. 32, no. 5, pp. 353–366, 2008.

[18] M. P. Sampat, A. C. Bovik, J. K. Aggarwal, and K. R. Castleman, "Pixel-by-pixel classification of MFISH images," in *Proc. 24th IEEE Ann. Int. Conf. Eng. Med. Biol. Soc*., Houston, TX, 2002, pp. 999–1000.

[19] Y. Wang, "Classification of M-FISH images using fuzzy C-means clustering algorithm and normalization approaches," in *Proc. 38th Asilomar Conf. Signals, Syst. Comput*., Nov. 2004, vol. 1, no. 7–10, pp. 41–44.

[20] Y. Wang and A. K. Dandpat, "Classification of multi-spectral fluorescence in situ hybridization images with fuzzy clustering and multiscale feature selection," in *Proc. IEEE Int. Workshop Genom. Signal Process. Statist*., May 2006, pp. 95–96.

[21] Y. Wang, "Detection of chromosomal abnormalities with multi-color fluorescence in situ hybridization (M-FISH) imaging and multi-spectral wavelet analysis," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., Vancouver, BC, Canada, Aug. 2008, pp. 1222–1225.

[22] H. Choi, A. C. Bovik, and K. R. Castleman, "Feature normalization via expectation maximization and unsupervised nonparametric classification for M-FISH chromosome images," *IEEE Trans. Med. Imag*., vol. 27, no. 8, pp. 1107–1119, Aug. 2008.

[23] R. Eils, S. Uhrig, K. Saracoglu, K. Satzler, A. Bolzer, I. Petersen, J. Chassery, M. Ganser, and M. R. Speicher, "An optimized fully automated system for fast and accurate identification of chromosomal rearrangements by multiplex-FISH (M-FISH)," *Cytogenet. Cell Genet*., vol. 82, no. 3–4, pp. 160–171, 1998.

[24] K. Saracoglu, J. Brown, L. Kearney, S. Uhrig, J. Azofeifa, C. Fauth, M. Speicher, and R. Eils, "New concepts to improve resolution and sensitivity of molecular cytogenetic diagnostics by multicolor fluorescence in situ hybridization," *Cytometry*, vol. 44, no. 1, pp. 7–15, May 2001.

[25] P. S. Karvelis, D. I. Fotiadis, I. Georgiou, and M. Syrrou, "A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*., New York, Aug. 30–Sep. 3, 2006, pp. 3009–3012.

[26] P. S. Karvelis, D. I. Fotiadis, and A. Tzallas, "Region based segmentation and classification of multispectral chromosome images," in *Proc. 20th IEEE Int. Symp. Comput.-Based Med. Syst*., pp. 251–256.

[27] (available until 2020). [Online]. Available: http://sites.google.com/site/xiaobaocao006/database-for-download

[28] J. C. Bezdek, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Mach. Intel*., vol. PAMI-2, no. 1, pp. 1–8, Jan. 1980.

[29] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man., Cybern*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

Hongbao Cao received the B.E. and M.S. degrees in biomedical engineering from the College of Precision Instrument and Optoelectronics Engineering, Tiangjin University, Tianjin, China, in 2002 and 2005, respectively. He received the Ph.D. degree from the Biomedical Engineering Department, Louisiana Tech University, Ruston.

He was a Postdoctoral Research Associate with the Department of Electronic Engineering and Computer Science, University of Missouri, Kansas City, from November 2009 to August 2010. Since September 2010, he has been a Postdoctoral Research Associate with the Department of Biomedical Engineering, Tulane University. He has about 14 publications, and his research interests involve signal processing, image processing, pattern recognition, and computational modeling.

Hong-Wen Deng received the Bachelor's degree in ecology and environmental biology and the Master's degree in ecology and entomology, both from Peking University, Beijing, China. He received the Master's degree in mathematical statistics and the Ph.D. degree in quantitative genetics, both from the University of Oregon, Eugene.

He was a Postdoctoral Fellow with the Human Genetics Center, University of Texas in Houston, where he conducted Postdoctoral research in molecular and statistical population/quantitative genetics. He also served as a Hughes Fellow with the Institute of Molecular Biology, the University of Oregon. He was a Professor of medicine and biomedical sciences at Creighton University Medical Center, Omaha, NB, Professor of orthopaedic surgery and basic medical science, and the Franklin D. Dickson/Missouri Endowed Chair in Orthopaedic Surgery with the School of Medicine, University of Missouri-Kansas City. He is currently the Chair of the Tulane University Biostatistics Department, New Orleans, LA, and the Director of Center of Bioinformatics and Genomics. He has widely published more than 400 peer-reviewed articles, 10 book chapters, and three books. His research interest includes the genetics of osteoporosis and obesity.

Dr. Deng is the holder of multiple National Institutes of Health RO1 awards and has received multiple honors for his research.

Yu-Ping Wang (SM'06) received the B.S. degree in applied mathematics from Tianjin University, Tianjin, China, in 1990 and the M.S. degree in computational mathematics and the Ph.D. degree in communications and electronic systems from Xi'an Jioatong University, Shaanxi, China, in 1993 and 1996, respectively.

After his graduation, he had visiting positions at the Center for Wavelets, Approximation, and Information Processing, the National University of Singapore and Washington University Medical School, St. Louis, MO. From 2000 to 2003, he was a Senior Research Engineer with Perceptive Scientific Instruments, Inc. and then with Advanced Digital Imaging Research, LLC, Houston, TX. In the Fall of 2003, he returned to academia as an Assistant Professor of computer science and electrical engineering with the University of Missouri-Kansas City. He is currently an Associate Professor of biomedical engineering and biostatistics at Tulane University, New Orleans, LA, and a member of Tulane Center of Bioinformatics and Genomics and Tulane Cancer Center. He is also a Visiting Professor with Shanghai University for Science and Technology, Shanghai, China. His research interests include the interdisciplinary biomedical imaging and bioinformatics areas, where he has about 100 publications. He has served on numerous program committees and National Science Foundation/National Institutes of Health review panels.

Dr. Wang was a Guest Editor for the *Journal of VLSI Signal Processing Systems* for a special issue on genomic signal processing and is a member of the Machine Learning for Signal Processing technical committee of the IEEE Signal Processing Society.