Middle-East Journal of Scientific Research 23 (8): 1772-1779, 2015
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2015.23.08.22255

Survey on Techniques Used for M-FISH Image Segmentation for Classification of Chromosomes

V.A. Pravina

ECE Hindusthan College of Engineering and Technology Coimbatore, India

Abstract: Multicolor Fluorescence Insitu Hybridization is a newly developed chromosome imaging method where eachchromosome class appears to have a different color. This methodnot only simplifies the detection of subtle chromosomal abnormalitiesbut also makes the analysis of chromosome images easier; bothfor human inspection and computerized examination. The segmentation and classification of Multicolor Fluorescence Insitu Hybridization (M-FISH) images can be used to detect chromosomal aberrations that can be used for cancer and genetic disease analysis. In this paper various methods to the problem of image segmentation and classification are explored. Here the possibility of using these algorithms to segment general images is presented and the issues involved in such algorithms are discussed.

Key words: Chromosomes · M-FISH · Pixel wise classification · Clustering Chromosome · Image Classificationet

INTRODUCTION

Chromosomes are the structures that contain genes, which store in strings of DNA all data necessary for an organism's development and maintenance which is an intricate schematic for cells and organisms. Images of chromosomes contain vital information about the health of a human being. A chromosome anomaly, abnormality or aberration reflects an atypical number of chromosomes or a structural abnormality in one or more chromosomes. A karyotype refers to a set of chromosomes from an individual which can be compared to a "normal" karyotype for the species through genetic testing. A chromosome anomaly may be identified or confirmed in this manner. Chromosome anomalies usually occur when there is an error in cell division following meiosis or mitosis. Chromosomal abnormalities are disturbances in the normal chromosomal content of a cell and are a major cause of genetic conditions in humans, such as down syndromeand turner syndrome although most aberrations have little to no effect. Some chromosome abnormalities do not cause disease in carriers, such as deletions, or chromosomal insertions, although they may lead to a higher chance of bearing a child with a chromosome illness. Abnormal numbers of chromosomes or chromosome sets, called aneuploidy, may be fatal or may give rise to genetic disorders. Genetic counseling is offered for families that may carry a chromosome rearrangement. There are many types of chromosome anomalies. They can be organized into two basic groups, numerical and structural anomalies.Numerical anomalies is called aneuploidy which means abnormal number of chromosomes and occurs when an individual is missing either a chromosome from a pair (monosomy) or has more than two chromosomes of a pair (trisomy, tetrasomy, etc.). In humans an example of a condition caused by a numerical anomaly is Down syndrome, also known as Trisomy 21 (an individual with Down syndrome has three copies of chromosome 21, rather than two). Trisomy has been determined to be a function of maternal age. An example of monosomy is Turner Syndrome, where the individual is born with only one sex chromosome, an X. Structural abnormalities occur whenthe chromosome's structure is changed. They can take several forms as listed below [1]:

- Deletions: A portion of the chromosome is missing or deleted. Known disorders in humans include Wolf-Hirschhorn syndrome, which is caused by partial deletion of the short arm of chromosome 4; and Jacobsen syndrome, also called the terminal 11q deletion disorder.

Corresponding Author: V.A. Pravina, ECE Hindusthan College of Engineering & Technology Coimbatore, India.

- Duplications: A portion of the chromosome is duplicated, resulting in extra genetic material. Known human disorders include Charcot-Marie-Tooth disease type 1A which may be caused by duplication of the gene encoding peripheral myelin protein 22 (PMP22) on chromosome 17.
- Translocations: A portion of one chromosome is transferred to another chromosome.
- Inversions: A portion of the chromosome has broken off, turned upside down and reattached, therefore the genetic material is inverted.
- Insertions: A portion of one chromosome has been deleted from its normal place and inserted into another chromosome.
- Rings: A portion of a chromosome has broken off and formed a circle or ring. This can happen with or without loss of genetic material.
- Isochromosome: Formed by the mirror image copy of a chromosome segment including the centromere[2].

Chromosome uncertainty syndromes are a group of disorders characterized by chromosomal instability and breakage. Thesefrequently lead to an increased tendency to develop certain types of malignancies.

Images of chromosomes taken during cell division contain valuable information about the well-being of anindividual.Images of chromosomes are useful for diagnosing genetic disorders and for studying cancer. Fig. 1 shows an example of M-FISH images of a male cell, where autosomes and sex chromosomes are classified from a five-channel spectral image data.

For a normal cell, each chromosome should be dyed with the same color. Else, it indicates that chromosomal anomalies might exist, which are associated with certain genetic diseases and cancers. The detection of chromosomal anomalies depends on accurate pixel-wise classification techniques[3]. Though many efforts have been made to automate image analysis procedure, 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 among image sets, and emission spectral overlays between fluorophores. The sizes of the misclassified regions are often larger than the actual chromosomal rearrangements or lost, which often lead to improper interpretation by cytogeneticists. To improve the detection of chromosomal abnormalities for clinical diagnosis, exact segmentation and classification algorithms have to be developed.

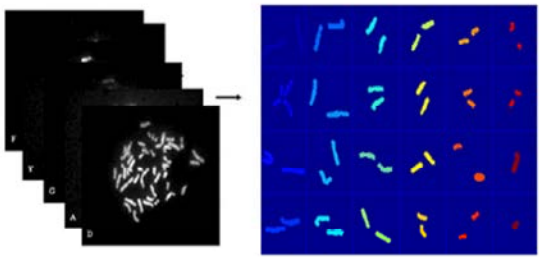


Fig. 1: M-FISH image and its karyotype

Methods for Analyzing Chromosomes: Chromosomes are structures that contain the genetic information of cells. Images of chromosomes taken during cell division contain valuable information about the wellbeing of an individual. Chromosome images are useful for diagnosing genetic disorders and for studying cancer. Thus the analysis of chromosomes is an important procedure in cytogenetic studies.

Grayscale Chromosome Images: Segmentation is the process of dividing the image into segments, each of which has some significance to the observer. It is desired to segment the image into background and chromosome pixels, and to divide further the chromosome pixels into individual chromosome type pixels. Segmenting a chromosome image into background and chromosome is a fairly straight forward task accomplished by thresholding [4]. However, dividing the chromosome pixels into individual chromosomes is quite difficult since chromosomes often touch and overlap. At the point of overlap, pixels may belong to multiple chromosomes. Various methods have been proposed but they fail when chromosomes are bent and do not handle overlaps. Classification usually follows segmentation in chromosome image analysis. After segmentation, chromosomes have a number of features, including length, centromere index and banding pattern that can be used to classify them. However, these parameters are difficult to extract automatically.

Karyotyping: Once segmented and classified, it is simple to arrange the chromosomes into a karyotype as shown in Fig. 2 for examination. There are 46 chromosomes which consist of 22 pairs of similar, homologous chromosomes and two sex-determinative chromosomes. Thus there are 24 types, or classes, of chromosomes. The process of assigning the chromosomes to the different classes is known as Karyotyping. These detect deviation from a normal cell structure.

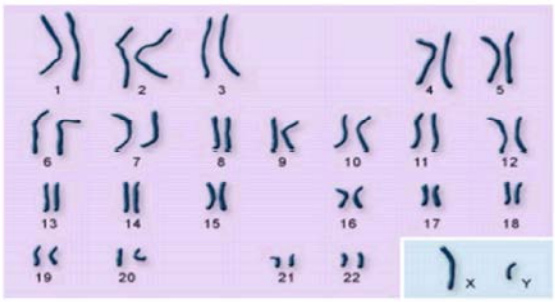


Fig. 2: Karyotype

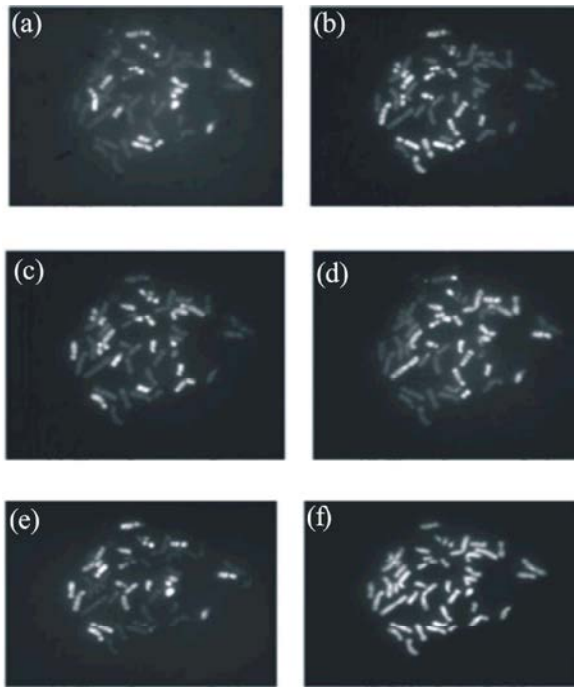


Fig. 3: M-FISH image set

Images of chromosomes are analyzed by cytogeneticists to obtain important information about the health of an individual. However, manual examination of these images is a laborious and time-consuming process and requires skilled lab technicians. Many successful attempts have been made to automate parts of the chromosome image analysis procedure. However it has been found that it is difficult to automate [5].

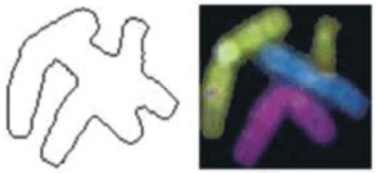
M-FISH Imaging: Multicolor Fluorescence In Situ Hybridization (M-FISH) is a combinatorial labeling technique that is developed for the analysis of human chromosomes. The technique has been used for the characterization of chromosomal translocations, to search for cryptic rearrangements, and to study mutagenesis, tumors and radiobiology. In this technology, chromosomes are labeled with five dyes and a DNA stain known as DAPI 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. An M-FISH data set consists of six images where each image is the response of the chromosome to a particular fluor. A typical M-FISH dataset is shown in Fig. 3. Fig. 3(a) to 3(e) are the images of the responses of the five fluors which are Spectrum Aqua, Far Red, Spectrum Green, Spectrum Red, Spectrum Gold, respectively. Fig. 3(f) shows the response of the DNA stain DAPI. DAPI attaches to DNA and thus chromosomes are seen in the image [6].

Table 1: M-FISH flour labeling

chromosome	Aqua	Green	Gold	Red	Far Red
1	0	0	1	0	0
2	0	0	0	1	0
3	1	0	0	0	0
4	0	1	0	1	1
5	0	0	1	0	1
6	0	1	0	0	0
7	0	0	0	1	1
8	0	0	0	1	1
9	0	0	1	1	0
10	1	0	1	0	0
11	1	0	0	1	0
12	0	1	1	0	0
13	1	1	0	0	0
14	0	1	1	1	0
15	1	0	1	1	0
16	0	1	0	0	1
17	0	1	0	1	0
18	0	0	1	1	1
19	0	1	1	0	1
20	1	0	0	1	1
21	1	1	1	0	0
22	1	1	0	1	0
X	1	0	0	0	1
Y	1	0	1	0	1

Then each chromosome is labeled by a unique combination of the five fluors.

Several such sets of fluors have been developed for M-FISH imaging. One such set of five fluors and the fluor labeling table is shown in Table 1. Here the first column represents the chromosome number. Names of the five different fluors are shown in the first row. A 1 indicates that a particular chromosome is labeled by the fluor and a 0 indicates that the chromosome is not labeled by the fluor [7, 8].



(a) Boundary of cluster (b) Multi spectral information incluster

Fig. 4: Comparison of two types of cluster information

Table 2: Chromosome misclassification rates

	Chromosome Misclassification Rates	
Singles	Joint Segmentation-Classification	Only Pixel Classification
	8.1%	15%

Thus each chromosome is labeled by a specific combination of dyes.This M-FISH imaging technique has the following advantages. They are listed as follows.

- The task of chromosome classification is greatly simplified. Instead of estimating the features such as centromere positions and banding patterns, which may be difficult to measure, one has to look at the spectral information within that chromosome.
- It resolves touching and overlapping chromosomes without losing the ability to detect translocations and rearrangements.
- No longer are centromere location, banding pattern, and other complicated, difficult to measure, features necessary to determine a chromosome's class since color alone is theoretically sufficient to determine the class.
- It is possible to detect smaller translocations and rearrangements since they are easily noticed as a single chromosome with two different colors in it as shown in Fig. 4.

If one observes Fig.4 (a), it is not clear about the segmentation. Even to human observers it is not apparent whether there is an overlap involved or even how many chromosomes are included in this cluster. However, by looking at the M-FISH multispectral information, a human being would very easily be able to determine what proper segmentation should be since each chromosome has its own color. All these images provide significantly more information than grayscale chromosome images and promise significant improvements in the accuracy of chromosome identification, classification and anomaly detection. Grayscale methods were often forced to perform segmentation followed by classification, since the grayscale classification features could only be measured on a segmented chromosome. But in M-FISH classification can be performed independently of segmentation [9, 10].

Image Segmentation and Classification Methods: In this section the algorithms that were developed for the segmentation and classification of chromosomes are discussed.

Pixel Wise Classification: In [8] a new method for automatic chromosome identification by exploiting the multispectral information in M-FISH chromosomes images is presented. In this paper, chromosome segmentationand classification is jointly done. A likelihood function proposed here was used as an indicator of errors in segmentation and classification. It was also used to indicate chromosomal anomalies, which can be used for diagnosing cancer and wide variety of inherited diseases**.**

However even with preprocessing and post processing, classification accuracy was not high enough and it led to higher misclassification rate as shown in Table 2.

In [3] a novel method for segmentation and classification of M-FISH chromosome images is presented. The segmentation was based on the multichannel watershed transform, in order to define regions of similar spatial and spectral characteristics. Then, a Bayes classifier, task-specific on region classification, was applied. By introducing the classification of each watershed region, the proposed method achieved substantially better results compared to other methods at a lower computational cost. Here the method divides M-FISH image into regions, i.e., groups of pixels which are assumed to be members of the same chromosome class.

The method compares a set of pixels with the training class distributions instead of comparing a data vector with the distributions of the trained classes in a 5-D space. Two indicative cases where the proposed method is superior compared to the pixel-by-pixel classification method are presented in Fig. 5. In these two cases, pixel-by-pixel classification produces noisy results making the decision of the expert difficult since these artifacts can be misinterpreted as chromosome abnormalities. This is obvious by the misclassifications errors produced by the pixel-by-pixel algorithm as shown in Fig. 5. If a classification is performed on a pixel-by-pixel basis, the classification will be dominated by noisy painting inhomogeneities. Table 3 shows a comparison of several different segmentation and classification algorithms.

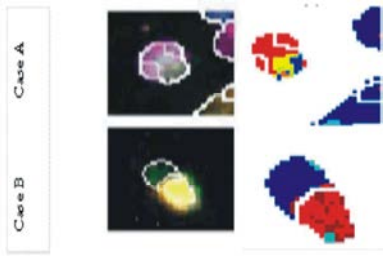


Fig. 5: Image dominated by inhomogeneities

Table 3: Pixelwise classification output

Method	Year	Images	Acpc[1] (%)
Sampat *et al.* [10]	2002	5	91.4
Choi *et al.*[11]	2004	10	97.1
Sampat *et al.*[7]	2005	5	90.5
Wang *et al.*[12]	2005	5	87.5
Schwarkopf *et al.*[8]	2005	183	68.0
karvelis *et al.*[4]	2006	15	89.9

Table 4: Classification accuracy result

	MD Classifier			ML Classifier		
Methods	NP	BC	EM	NP	BC	EM
Accuracy	47.1	60.1	68.7	47.8	62.4	72.7

Table 5: Classification result

	Pixel by pixel classification		Watershed Area Classification	
Images	ACC (%)	AUC_{total}	ACC (%)	AUC_{total}
1	93.84	0.990	96.11	0.996
2	86.40	0.987	97.11	0.995
3	82.20	0.978	95.81	0.994
4	72.70	0.933	85.30	0.953
5	80.60	0.965	93.78	0.989
6	74.50	0.952	91.38	0.990
7	61.50	0.894	72.50	0.931
8	66.10	0.916	74.30	0.953
9	82.00	0.964	91.44	0.987
10	83.30	0.969	93.77	0.995
11	77.80	0.956	91.57	0.985
12	64.10	0.923	82.00	0.965
13	65.50	0.938	86.90	0.960
14	85.30	0.985	96.21	0.996
15	82.30	0.965	94.75	0.991
overall	77.21±9.5	0.95±0.03	89.53±7.9	0.98±0.02

The table shows that the average pixel-by-pixel classification accuracy for the whole set was only 68% with standard deviation 17.5% [11].

In [1], a new feature normalization method for M-FISH images that reduces the difference in the feature distributions among different images using the expectation maximization (EM) algorithm was introduced. Also a new unsupervised, nonparametric classification method for M-FISH images was adopted. The performance of the classifier was as accurate as the maximum-likelihood classifier, whose accuracy also significantly improved after the EM normalization. A significant improvement was achieved on the pixel classification accuracy after the new feature normalization. Indeed, the overall pixel classification accuracy improved by 20% after EM normalization. The classification accuracy obtained was as tabulated in Table 4.

As the Table 4 shows, the overall classification accuracy without any normalization was about 50%, which increased significantly after background correction to about 60%, and further improved with EM normalization to about 70% for both classification methods. EM normalization increased the classification accuracy from 50% to70%, which is a 40% increase in accuracy. However this is not sufficient for clinical use.

Region Based Segmentation And Classification: In [4] an automated method for the classification of multispectral chromosome images based on the watershed transform has been presented. Initially, the chromosome image is decomposed into a set of primitive homogeneous regions [5].

Each segmented region is then classified using a Bayes classifier. This methodology has been evaluated using the commercially available M-FISH database and an overall accuracy of 89% was reported. Classification results of both pixel wise method and region based method are tabulated as shown in Table 5. However the classification accuracy result obtained by region based method was not sufficient for clinical use.

In [10] An automated method for the segmentation and classification of multispectral chromosome images has been presented. The chromosome image is first decomposed into a set of homogeneous regions. Each region is then classified using a regionBayes classifier. The methodology has been evaluated using available M-FISH database and an overall accuracy 83.59% and 89.88% was reported for the segmentation and classification respectively.

The proposed classification method (RBC) was compared with a pixel-by-pixel classification technique [3-6] and shows better classification accuracy (89.88%). The result obtained by region based method was 89% however it was not sufficient for clinical use.

Fuzzy C Means Clustering: In [2] it partitions a collection of n vector xi, i=1...n into c fuzzy groups and finds a cluster center in each group such that a cost function of

Table 6: Comparison of classification results of pixel wise methods with region based method

	MIS		RBC	
Images	Without WB (%)	With WB (%)	Pixel by Pixel [3] (%)	Proposed Method (%)
1	49.67	89.94	86.60	97.48
2	64.01	89.98	86.70	97.62
3	45.86	89.09	85.00	95.04
4	69.59	72.14	85.20	96.17
5	65.44	76.44	83.90	96.18
6	57.61	90.76	56.30	66.30
7	71.71	77.20	82.00	95.66
8	70.89	85.30	86.40	93.90
9	75.14	62.10	81.70	80.50
10	51.89	90.13	89.30	96.59
11	57.81	92.02	86.50	05.62
12	63.97	95.41	61.80	72.70
13	64.98	76.00	70.70	94.32
14	50.31	93.84	75.50	85.70
15	78.29	73.46	82.40	84.40
overall	62.48±9.96	83.89±9.89	80.01±9.79	89.8±9.85

dissimilarity measure is minimized. FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. However, FCM still uses a cost function that is to be minimized while trying to partition the data set. The membership matrix U is allowed to have elements with values between 0 and 1. However, the summation of degreesof belongingness of a data point to all clusters is always equal to unity. The objective function can be formulated as

$$J_{fcm} = \sum_{i \in D} \sum_{k=1}^{NC} u_{ik}^{m} \, \| y_i - c_k \|^2 \quad (1)$$

where uik is the membership function with values between 0 and 1, ck is the cluster center; m is the weighing exponent on each fuzzy membership which determines the amount of fuzziness, D is the area of the image, NC is the number of clusters.

The FCM method when applied to human chromosomes produced the following results [9].

As shown in Fig. 6(a), although the intensity of the chromo-some at location C was low, it should be clearly identified by a trained cytogeneticist. Fig .6(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.

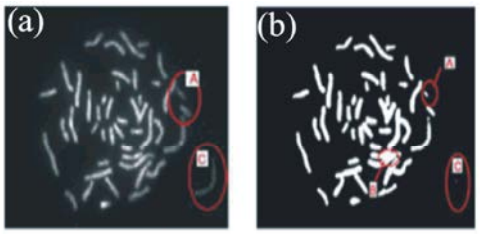


Fig. 6: FCM segmentation result

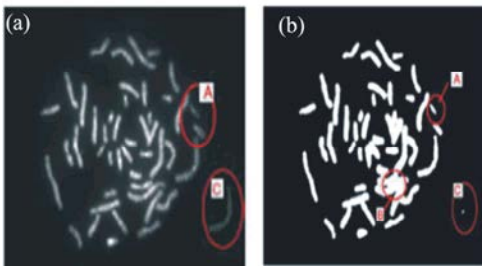


Fig. 7: AFCM segmentation result

Adaptive Fuzzy C Means Clustering: The fuzzy C means algorithm (FCM), in particular, can be used to obtain segmentation via fuzzy pixel classification. Unlike hard classification methods which force pixels to belong exclusively to one class, FCM allows pixels to belong to multiple classes with varying degrees of membership. This approach allows additional flexibility in many applications. The Adaptive Fuzzy C Means algorithm (AFCM), produces a fuzzy segmentation while compensating for intensity inhomogeneities. Here [6] a new objective function for obtaining fuzzy segmentations for images with intensity inhomogeneities and an iterative algorithm for minimizing this objective function is proposed. The objective function proposed could be expressed by,

$$J_{afcm} = \sum_{i \in D} \sum_{k=1}^{NC} u_{ik}^{m} \, \| g_i y_i - c_k \|^2 + \lambda_1 \sum_{i \in D} (G_i^{'})^2 + \lambda_2 \sum_{i \in D} (G_i^{''})^2 \quad (2)$$

The objective function contains a multiplier field term that models the brightness variation caused by the inhomogeneities. AFCM used a gain field to modify the centers of each cluster and to compensate the slowly changing inhomogeneities effects. Here the energies of first and second order derivatives of the gain field are also employed [12].

In [9] AFCM method was applied to chromosomes. The segmentation result obtained is also as shown.

The AFCM based segmentation result as in Fig. 7(b) is relatively better than FCM-based method. It covers the chromosome at area A by taking spatial contextual

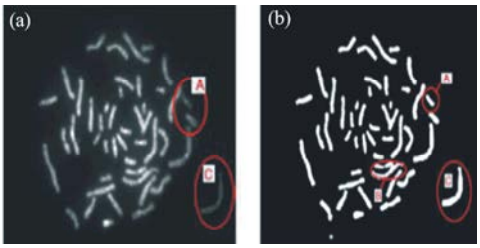


Fig. 8: IAFCM segmentation result

information into consideration. However, it failed to find the whole chromosome at area C as in Fig. 7(a). Also it resulted in messing up of two chromosomes indicated by an area red circled as B as shown in Fig. 7(b).

Improved Adaptive Fuzzy C Means Clustering: In order to improve the segmentation and classification of chromosomes, a new algorithm known as Improved Adaptive Fuzzy C means Clustering algorithm was employed. This IAFCM algorithm used a new objective function with a different regularization term, which appears to be more effective in improving the results. The objective function of the IAFCM can be calculated by,

$$J_{iafcm} = \sum_{i\in D} \sum_{k=1}^{NC} u_{ik}^{m} \left\| y_i - g_i c_k \right\|^2 + \lambda \sum_{i\in D} (g_i - (H * g)_i)^2 \quad (3)$$

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; m is the weighting exponent on each fuzzy membership, which determines the amount of fuzziness; D is the whole area of the image; NC is the number of clusters; $\{g_i \mid i \in D\}$ is the gain field to be found; and H is a (2r+1)*(2r+1) average convolution kernel. IAFCM method when applied to chromosomes provided the result as shown in Fig. 8(b).

The IAFCM segmentation result not only found the chromosomes at area C and area A but "cleaned up the mess" that existed in between two chromosomes. The work has shown that IAFCM segmentation yields better segmentation and classification than that of previous methods.

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. The algorithms developed earlier mainly focus on the correction of inhomogeneous background that smoothly and slowly vary through the image space. In this paper segmentation and classification of images using different soft computing techniques are discussed. From the survey it is clear that, the use of IAFCM algorithm yielded substantially better segmentation and classification results contributing to improved diagnosis of genetic diseases and cancers.

REFERENCES

1. Choi H, A. C. Bovik and K. R. Castleman, 2008. Feature normalization via expectation maximization and unsupervised nonparametric classification for M-FISH chromosome images, IEEETrans.Med.Imag., 27(8): 1107–1119.
2. Jiang, L. and W. Yang, 2003. A modified fuzzy c-means algorithm for segmentation of magnetic resonance images, in Proc. 7th Int. Conf. Digital Image Comput.: Tech. Appl, pp: 225-232.
3. Karvelis, P.S., A.T. Tzallas, D.I. Fotiadis and I. Georgiou, 2008. A multichannel watershed-based segmentation method for multispectral chromosome classification, IEEE Trans. Med. Imag., 27(5): 697-708.
4. Karvelis, P.S., D.I. Fotiadis, I. Georgiou and M. Syrrou, 2006. A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification, in Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., New York, pp: 3009-3012.
5. Karvelis, P.S., D.I. Fotiadis and A. Tzallas, XXXX. Region based segmentation and classification of multispectral chromosome images, in Proc. 20th IEEE Int. Symp. Comput.-Based Med. Syst., pp: 251-256.
6. Pham, D.L. and J.L. Prince, 1988. An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, Pattern Recog. Lett., 20: 57-68.
7. Sampa,t M.P., A.C. Bovik, J.K. Aggarwal and K.R. Castleman, 2005. Supervised parametric and non-parametric classification of chromosome images,Pattern Recog., 38: 1209-1223.
8. Schwarzkopf, W.C., A.C. Bovik and B.L. Evans, 2005. Maximum likelihood techniques for joint segmentation-classification of multispectral chromosome images, IEEE Trans. Med. Imag., 24(12): 1593-1610.

9. Hongbao Cao, Hong-Wen Deng and Yu-Ping Wang, 2012. Segmentation of M-FISH Images for Improved Classification of Chromosomes With an Adaptive Fuzzy C-means Clustering Algorithm, IEEE Transactions On Fuzzy Systems, 20(1).
10. Sampat, M.P., A.C. Bovik, J.K. Aggarwal and K.R. Castleman, 2002. Pixel by pixel classification of MFISH images, in Proc. 24th IEEE Ann. Intern. Conf. (EMBS), Houston, TX, pp: 999-1000.
11. Choi, H., K.R. Castleman and A. Bovik, 2004. Joint segmentation and classification of M-FISH chromosome images, in Proc. 26th IEEE Ann. Intern. Conf. (EMBS), San Francisco, CA, pp: 1636-1639.
12. Wang, Y. and K.R. Castleman, 2005. Normalization of multicolor fluorescence in situ hybridization (M-FISH) images for improving color karyotyping, Cytometry, 64: 101-109.