



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Special Issue 3, April 2017

# Segmentation and Classification for Improved Karyotyping of Multispectral Chromosome Images

<sup>1</sup>K Hemavathi, <sup>2</sup>D Menaka,

PG Scholar, Dept. of Electronics and Communication Engineering, Sri Venkateswara College of Engineering,  
Sriperumbatur, Chennai, India

Assistant Professor, Dept. of Electronics and Communication Engineering, Sri Venkateswara College of Engineering,  
Sriperumbatur, Chennai, India

**ABSTRACT** : Human chromosomes has the origin of some diseases. To identify the abnormalities, image processing techniques are used. The input of chromosome image is taken from Multicolor Fluorescence In-situ Hybridization (MFISH) Database. In this paper, an improved algorithm is introduced to overcome the limitations of noise. Fuzzy C means clustering algorithm is used to regulate the centers of each intensity cluster for a grayscale image. Watershed transformation is another segmentation method, that segment the image using morphological operations. Watershed transformation is compared with Fuzzy C-means clustering algorithm using evaluation methods and it is shown that Fuzzy C-means clustering has better segmentation accuracy. For the classification stage, pixel classification was applied using k-nearest neighbour algorithm. The proposed system has higher segmentation and classification accuracy than the previous approaches.

**KEYWORDS** : Fuzzy C-Means Clustering; Watershed Transformation; Image Segmentation; Multicolor Fluorescence In-Situ Hybridization (M-FISH); Image Classification.

### I. INTRODUCTION

Human body is organized from cells. Each cell contains a nucleus. The nucleus has very large molecule called DNA (Deoxy Ribonucleic Acid). At cell division this molecule is compressed to a shape known as chromosome. Naturally human have 23 pairs of chromosomes. Each chromosome has 4 arms. The crossing point of arms is called as centromere. Multicolor fluorescence *in situ* hybridization (M-FISH) is a combinatorial labeling technique which is developed for the analysis of human chromosomes[4]. This technique has been used for the characterization of chromosomal translocations, and to study mutagenesis, tumors, and radiobiology. 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. 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"). Here, five spectra are sufficient to distinguish the 24 classes of chromosomes in human genome. In addition to that, DAPI channel is used to stain each chromosome such that all of the chromosomes are visible in a DAPI channel[1]-[4].

By simultaneously viewing six different channel images, pixel-wise classification of human chromosome is possible. M-FISH images of a male cell, where 22 autosomes and 2 sex chromosomes are classified from a 5-channel spectral image data and are displayed using 24 pseudocolors. For a normal cell, each chromosome should be painted with the same color. Otherwise, it indicates the presence of chromosomal abnormalities, which are often associated with certain genetic diseases or cancers. The successfully detection of chromosomal abnormalities depends on accurate pixel-wise classification techniques. Many attempts have been made to automate image analysis procedure, the reliability of the technique has not yet reached the level for clinical application[5]-[8]. The sizes of the misclassified regions are often larger than the actual chromosomal rearrangements and chromosomal gain or lost, which may leads to



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Special Issue 3, April 2017

incorrect interpretation by cytogeneticists. To improve the detection of chromosomal abnormalities for clinical diagnosis, accurate segmentation and classification algorithms have to be developed [1].

In this paper, fuzzy c-means clustering algorithm and watershed transformation algorithm is used for segmentation and applied to the classification of M-FISH images by considering intensity homogeneities, which often exist in the images. For classification, k-nearest neighbor algorithm was developed and classified into 24 classes which are represented in pseudocolors.

## A. Image database:

Database consist of 200 Multispectral images of size 517 X 645 pixels[16]. 17 images are marked as extreme (EX), that are “difficult to karyotype”. ASI, PSI, Vysis are the probes used. Each MFISH image set consist of 5 multispectral images recorded at different wavelengths, DAPI and its “ground truth” image according to ISCN (International System for Human Cytogenetic Nomenclature) for each MFISH image except for EX images. The MFISH image description is shown in table 1.

Table 1: MFISH image Description

PROBE OF M-FISH IMAGE	PLANES OF M-FISH IMAGE
3	532
5	Cy 5.5
6	568
A	S. Aqua
C	Cy 5
D	DAPI
E	DEAC
F	Far Red
G	S. Green
I	FITC
O	S. Orange
R	S. Red
T	Texas Red
Y	S. Gold

Ground truth image is labeled so that the gray level of each pixel represents its class number (chromosome type) where background pixel values are zero and pixels in the overlapped regions values are 255. It is used to determine the accuracy of M-FISH images classification. But translocations are marked such that the full chromosome is labeled with the class which makes up the most of the chromosome. Images used for training and testing are taken from this dataset.

## B. Karyotype:

Karyotype is the tabular representation of human chromosomes in a cell. Karyotyping of human chromosomes are shown in figure 1.1. In this representation, the chromosomes are ordered by length from largest (chromosome 1) to smallest (chromosome 22 in humans), followed by sex chromosomes. Karyotypes are very useful for accurately diagnosing the genetic factors behind various diseases[10]. The karyotyping of human chromosome images is shown in Fig 1.

Manual karyotyping is time-consuming, expensive and need well trained personnel. During the early period of chromosome analysis, researchers used grayscale images and features such as size, shape, centromere position and

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Special Issue 3, April 2017

banding pattern for classification. Since 1996, a staining method called M-FISH was introduced to produce color images. This simplifies the karyotyping and detection of subtle chromosome aberrations. Combinatorial labelling of 5 fluorophores is used to assign a specific fluor combination to each of the chromosomes, so that each chromosome type can be visualized in a unique color. A sixth fluorophore, DAPI (4 in, 6-diamidino-2-phenylindole), is counterstained to all chromosome.

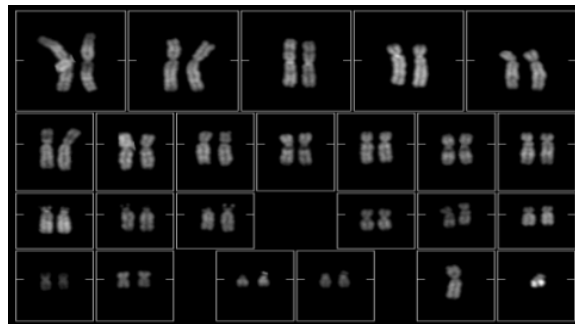


Fig 1 : karyotype

## II. METHODS

### A. Preprocessing:

Pre-processing steps are required mainly for noise removal. Some M-FISH images contains undivided cells as part of the image. So cell removal is the most important pre-processing stage before segmentation. To reduce shot noise and white noise, weighted median filter and low pass filter respectively are employed using 3x3 window neighborhoods[ 4]. The preprocessing stage achieved significant improvements in the classification results.

### B. Image segmentation:

Chromosome segmentation, which is the process of separating chromosomes from back ground of the M-FISH chromosome image, is the most difficult task in chromosome analysis. This is because there is no universally accepted segmentation technique suitable for non-rigid objects like chromosomes, and it does not guarantee hundred percent accuracy in segmentation. Moreover, overlapping of chromosomes makes segmentation and classification a challenging one. We have 6 gray scale images with intensity of corresponding fluorophore as pixel values. Six fluorophores are used for combinatorial labelling of all 24 classes of chromosomes based on whether the dye is attached with the particular pixels or not. In this paper, two algorithms were applied to different images in the dataset and the classification accuracy was calculated.

#### 1. FUZZY C-MEANS CLUSTERING ALGORITHM (FCM) :

Image segmentation is one of the most widespread means to classify correctly the pixels of an image in a decision oriented applications. Image segmentation is a technique that partitions an image into uniform and non-overlapping regions based on some likeness measure[13]. Clustering, the unsupervised classification of patterns into groups is one of the most important tasks in exploratory data analysis[1]. Primary goals of clustering include gaining insight into data, classifying data, and compressing data.

Clustering has a long and rich history in a variety of scientific disciplines including anthropology, biology, medicine, psychology, statistics, mathematics, Engineering and computer science. Fuzzy C-means (FCM) algorithm is one of the most popular fuzzy clustering methods widely used in various tasks of pattern recognition, data mining, image processing, gene expression data recognition *etc.* Modification and generalization of FCM algorithm is a prevailing research stream in fuzzy clustering in recent decades.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Special Issue 3, April 2017

General description:

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function shown in eqn (1) as follows:

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2 \quad 1 < m < \infty \quad \text{eqn. (1)}$$

where  $m$  is any real number greater than 1,  $u_{ij}$  is the degree of membership of  $x_i$  in the cluster  $j$ ,  $x_i$  is the  $i$ th of  $d$ -dimensional measured data,  $c_j$  is the  $d$ -dimension center of the cluster, and  $\|*\|$  is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function, with the update of membership  $u_{ij}$  and the cluster centers  $c_j$  are shown in eqn (2) and (3) by:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad \text{eqn. (2)}$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad \text{eqn. (3)}$$

This iteration will stop when  $\max_{ij} \left\{ |u_{ij}^{(k+1)} - u_{ij}^{(k)}| \right\} < \epsilon$ , where  $\epsilon$  is a termination criterion between 0 and 1, whereas  $k$  is the iteration step. This procedure converges to a local minimum or a saddle point of  $J_m$ .

FCM Algorithm:

The algorithm steps for Fuzzy C-Means clustering are,

1. Initialize  $U=[u_{ij}]$  matrix,  $U^{(0)}$
2. At  $k$ -step: calculate the center vectors  $C^{(k)}=[c_j]$  with  $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$$

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Special Issue 3, April 2017

3. Update  $U^{(k)}$ ,  $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. If  $\|U^{(k+1)} - U^{(k)}\| < \epsilon$  then STOP ; otherwise return to step 2.

*FCM computation:*

FCM has the centroid of a cluster is computed as being mean of all points, weighted by their degree of belonging to cluster. The degree of being in a certain cluster is related to the inverse of the distance to the cluster. By iteratively updating the cluster center and membership grades for each data point, FCM iteratively moves the center to 'right' location within a data set[5].

The performance depends on initial centroids using an algorithm to determine all the centroids (eg..arithmetic means of all data points). Run FCM several times each starting with different initial centroids. It performs unsupervised clustering methods. FCM employs fuzzy partitioning such that a data point can belong to all the groups with different membership grades between 0 and 1.The results of fuzzy c-means clustering segmentation is given below in Fig 2.

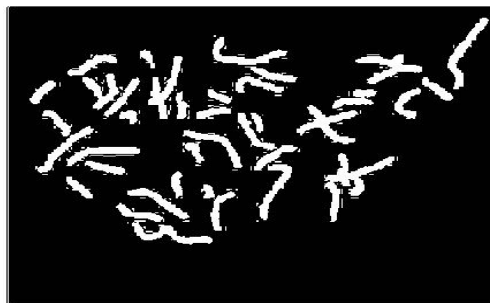


Fig 2 : FCM segmentation result

*Advantages of FCM algorithm:*

The advantages of Fuzzy C-Means clustering algorithm are,

- FCM is an iterative algorithm
- The aim of FCM is to find cluster center (centroid) that minimize a dissimilarity function

## 2. WATERSHED TRANSFORMATION

The watershed transformation technique is well known to be a very powerful segmentation tool. Gray level images are considered as topographic reliefs, each relief is flooded from its minima and when two lakes merge, a dam is built the set of all dams define the so called watershed[10]. Such representation of the watershed simulates the flooding process. The results of watershed transformation is shown in Fig 3.

*Advantages of watershed transformation:*

The advantages of watershed transformation are given below,

- The watershed transformation always provides closed contours, which is very useful in image segmentation.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Special Issue 3, April 2017

- It requires low computation times in comparison with other segmentation methods.

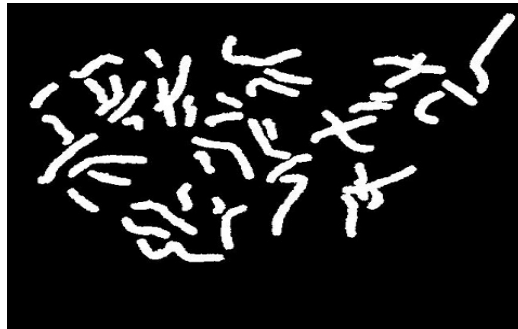


Fig 3 : watershed transformation result

## C. Image Classification:

The problem of object classification can be specified as a problem to identify the category or class that the new observations belong to based on a training dataset containing observations whose category or class is known. k-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. k-nearest neighbor algorithm is among the simplest of all machine learning algorithms[6]. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabelled query point is simply assigned to the label of its k nearest neighbors. Typically the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be a odd integer. However, there can still be ties when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance:

$$d(x, y) = \|x - y\| = \sqrt{(x - y) \cdot (x - y)}$$
$$= (\sum ((x_i - y_i)^2))^{1/2}$$

A main advantage of the KNN algorithm is that it performs well with multi-modal classes because the basis of its decision is based on a small neighborhood of similar objects. Therefore, even if the target class is multi-modal, the algorithm can still lead to good accuracy. However a major disadvantage of the KNN algorithm is that it uses all the features equally in computing for similarities. This can lead to classification errors, especially when there is only a small subset of features that are useful for classification. The result of multispectral image classification is shown in Fig 4.

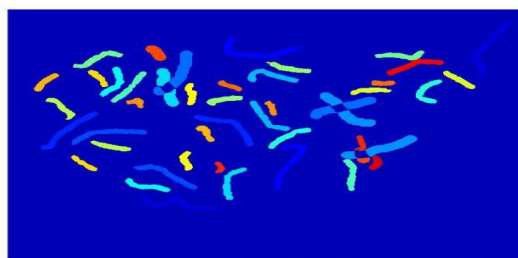


Fig 4 : MFISH image classification



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Special Issue 3, April 2017

## III. RESULTS AND CONCLUSION

Many segmentation and classification algorithms are applied for MFISH chromosome images. In this paper, FCM clustering algorithm and watershed transformation was applied to segmentation and pixel based classification was done using k-nearest neighbor algorithm. The accuracy for segmentation and classification was calculated using the below formula .

### Image segmentation results:

The segmentation accuracy of chromosome images are calculated by using this formula:

$$\text{Segmentation accuracy} = \frac{\text{chromosome pixels correctly segmented}}{\text{total no. of chromosome pixels}}$$

Comparing the results of both the segmentation process using evaluation methods is shown in table 2. Fuzzy C-Means clustering has higher accuracy compared to watershed transformation.

Table 2 : segmentation accuracy

No. of images	FCM segmentation	Watershed segmentation
1	89.09	87.62
2	89.11	86.39
3	90.88	87.07
4	90.54	88.12

### Image classification results :

The classification accuracy of chromosome images are calculated by using this formula:

$$\text{Classification accuracy} = \frac{\text{chromosome pixels correctly classified}}{\text{total no. of chromosome pixels}}$$

For image classification using k-nearest neighbour algorithm, it achieves higher accuracy of 91.68%. This algorithm is tested with 4 datasets. The future work will be applying this technique for more number of datasets which can be further increases in classification accuracy.

## REFERENCES

1. H. Cao; H.W Deng; M. Li; Y. P. Wang; "Classification of Multicolor Fluorescence In Situ Hybridization (M-FISH) Images With Sparse Representation", IEEE Transactions on nanobioscience, vol. 11, no. 2, June 2012.
2. Majid .E; Hadi .S; Masoumeh .V; "A novel classification method: A hybrid approach based on extension of the UTADIS with polynomial and PSO-GA algorithm", ELSEVIER Applied Soft Computing 49 (2016) 56–70 , 2016.
3. Ashu Makkar; V. K Govindan; Lijiya. A; "A Re-Segmentation Algorithm for Improved M-FISH Image Segmentation and Classification", (IJAIST) ISSN: 2319:2682 Vol.32, No.32, Dec 2014.
4. Y. Rahimi, Rassoul Amirfattahi, R. Ghaderi, "Design Of A Neural Network Classifier For Separation Of Images With One Chromosome From Images With Several Chromosomes", 978-0-7695-3453-4/08, IEEE DOI 10.1109/BROADCOM, 2008.
5. Karvelis P.S., D. I. Fotiadis, M. Syrrou, and I. Georgiou, "A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification," In Proc. Of 28th IEEE EMBS Ann. Intern. Conf., 2006, PP. 3009-3012.
6. Yu. Ping Wang, and A.K. Dandpat, "Classification of multispectral fluorescence in situ hybridization images with fuzzy clustering and multiscale feature selection", IEEE International Workshop on Genomic Signal Processing and Statistics, 28- 30 May 2006, PP.95 – 96.
7. Sampat M.P., K. R. Castleman, and A. C. Bovik, "Pixel-by-Pixel Classification of M-FISH Images", Proceedings of the Second Joint Conference of the IEEE EMBS and BMES Conference, Vol.2, pp. 999-1000, Oct 2002.



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Special Issue 3, April 2017

8. Petros S. Karvelis, Dimitrios I. Fotiadis and Alexandros Tzallas, "Region Based Segmentation and Classification of Multispectral Chromosome Images," Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007.
9. D. L. Pham, and Jerry L. Prince, "Adaptive Fuzzy Segmentation of Magnetic Resonance Images", IEEE Trans. Med. Imag., vol. 18, no. 9, pp. 737-752, Sep. 1999.
10. D. L. Pham and J. L. Prince, "An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities", Pattern Recognit. Lett., vol. 20, pp. 57-68, 1998.
11. Sreejini K S, Lijiya A and V K Govindan, "M-Fish Karyotyping - A New Approach Based On Watershed Transform", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, No.2, April 2012.
12. Jingyao Li, Dongdong Lin and Yu-Ping Wang, "Classification Of Multicolor Fluorescence In-Situ Hybridization (MFISH) Image Using Structure Based Sparse Representation Model With Different Constrains", IEEE 10th International Symposium on Biomedical Imaging, April 2013.
13. Hyohoon Choi, Alan C. Bovik, and Kenneth R. Castleman, "Maximum- Likelihood Decomposition of Overlapping and Touching M-FISH Chromosomes Using Geometry, Size and Color Information", 2006.
14. J. Li; Dong Lin; Y. P. Wang; "Segmentation of Multicolor Fluorescence In- Situ Hybridization (M-FISH) Image Using an Improved Fuzzy C-Means Clustering Algorithm While Incorporating Both Spatial and Spectral Information", IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015.
15. P. Karvelis; Aristidis .L.; D. I.Fotiadis; "Semi Unsupervised M-FISH chromosome image classification", IEEE 978-1-4244-6561-3/10/, 2010
16. Petros S. Karvelis, Dimitrios I. Fotiadis, Ioannis Georgiou and Marika Syrrou, "A Watershed Based Segmentation Method for Multispectral Chromosome Images Classification", 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.
17. <https://sites.google.com/site/xiaobaocao006/database-for-download>