ISSN 2278-5612

Karyotyping of Human Chromosomes Using M_FISH Images

Mumthas T K[1] ,Lijiya A[2] and V K Govindan[3]

Department of Computer Science and Engineering, National Institute of Technology, Calicut, India

[1]mumthastk@gmail.com, [2]lijiya@nitc.ac.in, [3]vkg@nitc.ac.in

ABSTRACT

Karyotyping is a technique used to align the chromosomes in the decreasing order of size so that the structural and numerical changes can be easily identified. Traditional chromosome analysis is simplified by the introduction of M_FISH (multiplex fluorescence in-situ hy-bridization), a combinatorial labeling technique in which each chromosomes appears in distinct color. In this paper, Bayes classification of M-FISH chromosome images using the features intensity in five channel as well as size of the chromosome is presented. Watershed transform is employed for segmentation . Also reclassification of small misclassified segments to the neighbor region is performed as post-processing to improve the performance. The classifier was trained and tested on a set of images from the dataset [2] and an overall accuracy of 91.09% was obtained.

Keywords

Chromosome analysis; M-FISH; Watershed Segmentation; Bayes Classifier

Council for Innovative Research

Peer Review Research Publishing System

Journal: International Journal of Management & Information Technology

Vol.4, No.2

editor@cirworld.com

www.cirworld.com, member.cirworld.com

INTRODUCTION

Chromosomes are microscopic structures found in the cell nuclei that determine the characteristics of an individual. Human cells have a total of 46 chromosomes which are arranged into 22 pairs of autosomes and sex chromosomes XX or XY. Chromosomes contain thousands of genes. Every chromosome is characterized by a neck like centromere and two arm regions, the p and q arms corresponds to the short and long arm regions. Normally chromosomes become visible when they replicate in a process known as mitosis. They can be stained and imaged by a microscope during this stage.

Chromosome analysis deals with counting the number and analyzing the structural changes in the chromosomes. An abnormality in DNA causes genetic disorder. Abnormalities in DNA can be caused due to unusual number of chromosome, deletion of part of a chromosome, duplication of genetic material within a chromosome, translocations, etc. Many of the genetic problems that cause disorder or disease can be identified by chromosome analysis. Some of the genetic diseases caused by chromosomal abnormalities include autism, down syndrome, mental retardation, etc. Also many of the cancers are caused by mutations in the chromosomes. Hence chromosome analysis has a vital role in diagnosing cancers and various genetic diseases.

M-FISH (multiplex fluorescence in-situ hy-bridization) is a cytogenic technique developed for the analysis of human chromosomes. It uses five fluorophores to assign a specific fluor combination to each of the chromosomes so that each chromosome type can be identified in unique color. An M_FISH image is captured using a fluorescent microscope with optical filters. 2^{N-1} objects can be labeled with N fluorophores; hence five fluorophores are used to label 24 chromosomes.Specific flour combination is assigned to each chromosomes using combinatorial labeling [3] of 5 flourophores. A sixth fluorophore, DAPI (4'-6-Diamidino-2-phenylindole) which labels all chromosomes, is used to create gray scale image. So an M_FISH multispectral image consists of 6 images. A pixel is represented as five-dimensional vector; each vector corresponds to the magnitude of each flour. An M-FISH image for the VYSIS probe is shown in "Figure 1.1". The introduction of this multispectral imaging technique made the traditional chromosome analysis simpler. i.e. centromere detection, length comparison and banding pattern analysis are not required; smaller translocations and rearrangements can also be identified.

The accuracy of pixel classification determines the success of this technique. Supervised and Unsupervised methods are described in the literature. Supervised methods require a set of images and the corresponding ground truth images from the dataset for training the classifier and the performance is evaluated to the dataset. Unsupervised methods directly classify the M-FISH image without training the classifier.

In this paper, Bayes classification of M-FISH chromosome images using the features intensity in five channel as well as size of the chromosome based on watershed segmentation is presented. Some of the related works in the areas of Segmentation and Classification methods of M_FISH chromosomes are included in Section 2. Segmenting the chromosome pixels from the background and Classification of the segmented results (chromosomes) are presented in section 3 and 4 respectively. The section 5 deals with the experimental and performance study of the results obtained from different approaches and finally the paper is concluded in Section 6.

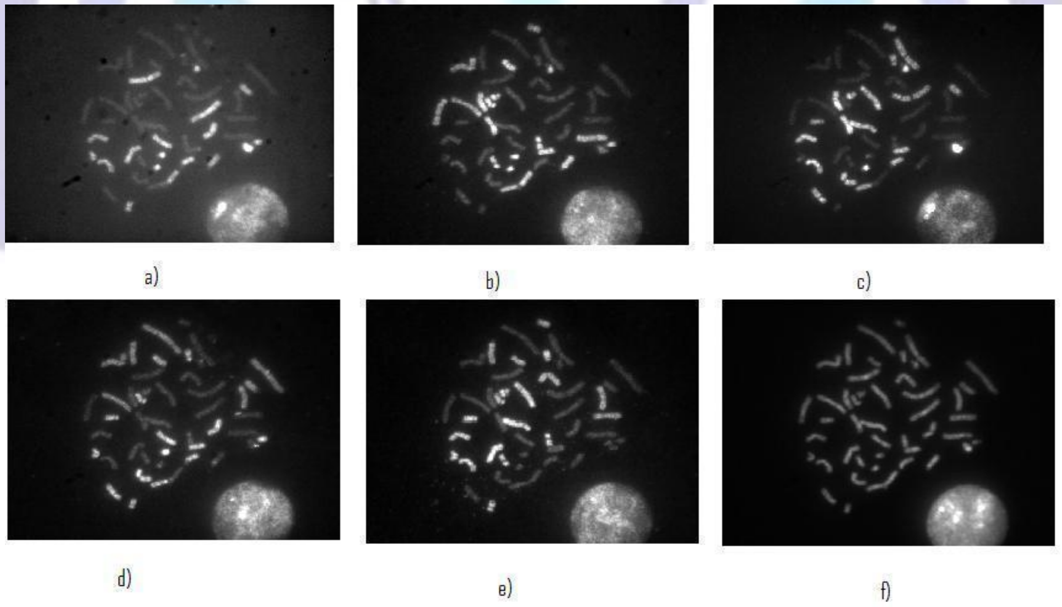


Fig 1.1: Six channel M-FISH image data a) Aqua flour b) Green flour c) Gold flour d) Red flour e) Far Red flour f) DAPI stain

RELATED WORKS

Automatic color karyotyping on M-FISH images that requires analysis of numerical and structural abnormalities is an active topic of current research. Many attempts have been made to automate image analysis procedure. Some of the existing publications in this topic are reviewed in the following.

Speicher M.R. et al. [1] introduced **M-FISH** (multiplex fluorescence in-situ hy- bridization), which made the analysis of chromosomes much easier. It is an unsupervised classification method and does not require training data. It provides simultaneous visualization of chromosomes. They have used epifluorescence filter sets and computer software for this purpose. All the 24 chromosomes types are identified using Combinatorial labeling [3]. An accurate alignment of source images and correction of chromatic cross talk are required in this approach. This method requires some manual image manipulation and that is a complicated task.

An algorithm for automatic pixel by pixel classification of M-FISH images was proposed by Mehul P. Sampat, et al. [4]. Pixel by pixel classification of M-FISH image into 24 chromosomes and the background as the 25 classes was done by a 6-feature 25 class Bayes Classifier. The classification accuracy obtained in this method is 91.4%. The non-overlapping data sets were only taken for training and testing the classifier.

Wade Schwartzkopf, et al. [5] proposed a segmentation algorithm for M-FISH images to decompose clusters of touching and overlapping chromosomes. The methods used include chromosome segmentation by thresholding, followed by labeling of connected components. Here the pixels in the connected components are classified using multi spectral information; then the clusters are divided by choosing cut points on the boundary of the cluster. He also presented an automatic identification method for chromosome by performing segmentation and classification jointly [6]. A likelihood function is defined in terms of three components: Lmulti (.) that uses multispectral information, Lsize (.) a function of a possible chromosome segment size and W (.) a weight function that accounts for overlaps. Connected component analysis is used to parse the image into individual chromosomes and clusters of touching and overlapping chromosomes. Touching or overlapping of chromosomes of the same class cannot be segmented in this approach.

A Fuzzy C-means Clustering (FCM) approach for multi-spectral FISH image classification is proposed by Yu-Ping Wang, et al. [7]. The FCM employs a fuzzy partitioning such that a data point can belong to all classes with different membership grades between 0 and 1. The data sets were normalized by Color compensation, dimension reduction and registration. The FCM algorithm together with the data normalization leads to improved classification rate.

P. S. Karvelis, et al. [8] presented a region based method for chromosome classification in M-FISH image. Here the chromosome image is decomposed into regions using watershed transform and then classified using Bayes classifier. The average intensity value of each channel is used as the feature. This classification resulted in an overall accuracy of 89% for 15 images from the commercially available M-FISH database. This approach reduces the computation time since only watershed regions have to be classified instead of individual pixels. An extension to this work is also presented in “A multichannel watershed-based segmentation method for multispectral chromosome classification” [9]. Here the contrast information from the different spectral channels is used as part of the gradient magnitude computation. Unwanted regions are further reduced by superimposing a binary mask of the DAPI channel. After classification, adjacent regions of the same class are merged to one single region. The authors claimed an overall accuracy of 82% with standard deviation 12% for the entire images of M-FISH datasets excluding the 17 images which are reported as “difficult to karyotype ”. Sreejini K.S, et al [10] presented a modified approach to [8]. The watershed regions were classified using the features, mean and standard deviation of intensity. It also includes reclassifying small segments to the neighboring larger segments so that the chances of misclassification can be reduced.

Supervised parametric and non-parametric classification of M-FISH images was done by .M. P. Sampat, et al. [11]. The methods adopted include Maximum likelihood estimation for supervised classifications, nearest neighbor and k-nearest neighbor for unsupervised classifications. The metric used is the Euclidean distance in both nearest neighbor and k-nearest neighbor. The values of k used are 5,7 and 9 neighbors. Pixel selection was done by means of Laplacian of Gaussian edge detector and then a morphological operator was applied for edge dilation so that segmentation will be more accurate. A 5-by-5 majority filter was applied to the classification output to remove the errors observed after classification. The authors claimed that Non-Parametric methods performed better than the parametric method and the highest classification accuracy was obtained by the k-nearest neighbor method. The optimal value for k is 7 for this classification. This method was tested only for 5 images from the data sets available for chromosome images.

A karyotyping technique presented by Hyohoon Choi et al. [12] provides analysis of numerical and structural abnormalities of human chromosomes. The M-FISH image contains crosstalk between channels due to spectral overlap and broad sensitivity of sensors, resulting in color spreading. The author modeled this color spreading as a linear transformation and obtained a spread matrix. The inverse of this matrix, the color compensation matrix, is used to correct the spreading and to recover the true intensities. This method resulted in an improvement in image quality. Some of the noises that might occur in the image due to non uniform hybridization inside chromosomes and among different chromosomes are not considered. Also images that contains misalignment is also not considered here.

An enhanced classification of multichannel Chromosome images using Support Vector Machines (SVM) was presented by P.S. Karvelis, et al. [13]. Segmentation and classification are done by means of Watershed Transform and Support Vector Machines respectively. Gradient magnitude computation and grayscale reconstruction was done as part of watershed transform. The classifier resulted in an overall accuracy of 80.20% which outperformed the Bayes classifier, when it is tested for 20 images from the M-FISH data set.

Petros Karvelis et al. [14] also presented a Semi Unsupervised method for the classification of M-FISH chromosome images. In order to extract the pixels which belong to chromosomes, an automated thresholding is adopted. A normalization procedure that reduces the difference in the feature distributions among different images is then applied. Since feature normalization is applied after segmentation, classification is further improved. Also prior information, i.e. emission information for each chromosome class is used to initialize cluster centroid.

Hongbao Cao and Yu-Ping Wang [15] adopted an improved adaptive fuzzy c-means clustering algorithm for chromosomes classification of multi-color fluorescence in situ hybridization (M-FISH) images .The algorithm improves the classical fuzzy c- means (FCM) clustering algorithm by introducing a gain field. When compared with other fuzzy c-means clustering based algorithms and region- based segmentation and classification algorithm, this method gave the lowest segmentation and classification error.

Lijiya A, et al. [16] proposed a majority voting scheme for segmenting the chromosomes followed by fuzzy logic classifier. The segmentation was done by means of Laplacian of Gaussian, Global thresholding and 6-feature 2-class K-means. Some pre-processing such as cell removal, noise removal and dilation are performed to improve the performance of karyotyping.

A method that handled occlusions in the chromosomes includes a decomposition method for overlapping and touching chromosomes by Hyohoon Choi, et al. [17] . The fuzzy-logic classifier [11] is adopted for classification. A group of connected pixels is defined as cluster. Three sets of basic elements for clusters are defined as Cross shape cluster, T shape cluster and I shape cluster. The cluster is decomposed into multiple hypotheses and the likelihood of each hypothesis is computed. Among all hypotheses, the one that has the maximum likelihood is chosen as the correct decomposition of the cluster.

Though there are a number of attempts by various researchers in karyotyping, the performances of the automatic karyotyping system is still not acceptable for commercial deployment. So, a great deal of further research is required to improve the performance of karyotyping systems.

SEGMENTATION

It is required to segment the chromosome pixels from the background, since only the chromosomes pixels have to be classified.

Pre-processing

Cell nuclei often present in the FISH image is removed from the DAPI channel based on circularity and size. For circular objects,

$$4\pi*\text{area} / (\text{perimeter})^2 = 1$$

“Figure 3.1” shows the results of pre-processing.

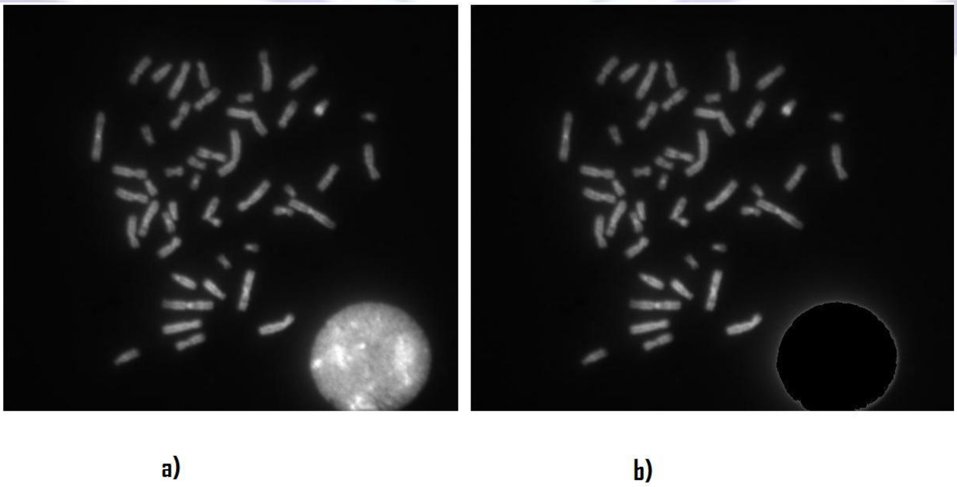
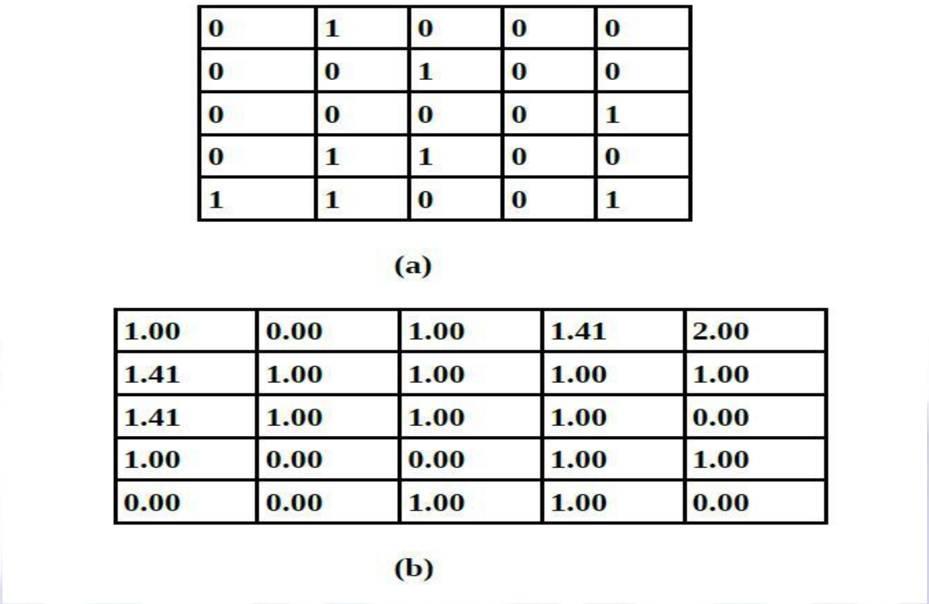


Fig 3.1: a) DAPI channel before cell removal b) After cell removal

Segmentation was done by means of watershed segmentation based on distance transform [18][19]. In order to make the image suitable for watershed segmentation, the distance transform of the image was computed. It finds the distance from every pixel to the nearest non-zero valued pixel. The distance metric used is the Euclidian distance. Let [x, y1] and [x2, y2] be two pixels in a digital image, then the Euclidian distance between them is given by,

$$dEuclidian\ ([x1, y1], [x2, y2]) = \sqrt{(x1 - x2)2 + (y1 - y2)2}$$

“Figure 3.2” shows a binary image and its corresponding distance transform.



0	1	0	0	0
0	0	1	0	0
0	0	0	0	1
0	1	1	0	0
1	1	0	0	1

(a)

1.00	0.00	1.00	1.41	2.00
1.41	1.00	1.00	1.00	1.00
1.41	1.00	1.00	1.00	0.00
1.00	0.00	0.00	1.00	1.00
0.00	0.00	1.00	1.00	0.00

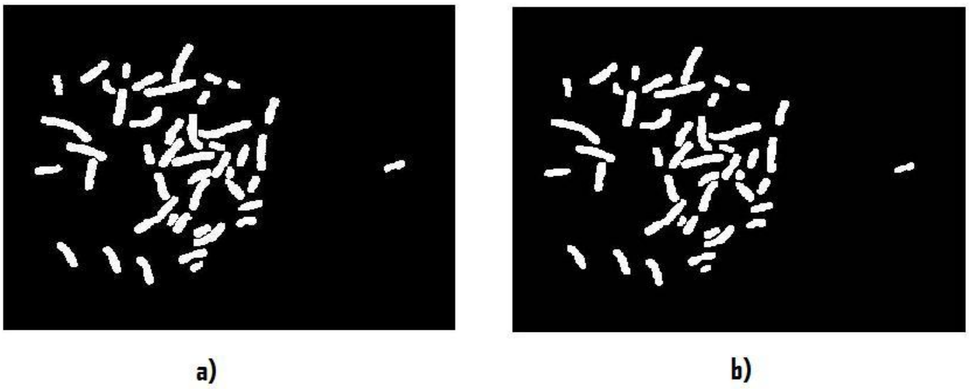
(b)

Fig 3.2: a) A binary image matrix b) Its corresponding distance transform.

Computation of Watershed transform

The image is segmented in to different regions as a result of applying Watershed transform [20]. The watershed transform is a gradient based segmentation technique. The image is considered as a relief map in which the gradient values are related to the height of the surface. If we punch a hole in each local minimum and immerse the whole surface in water, the water level will rise over the regions. A dam is built between the two different bodies of water. The flooding process continues until all the points in the map are immersed. Finally the entire image is segmented by the dams. The dams are called the watersheds and the catchment basins correspond to the segmented regions.

In watershed segmentation, over segmentation is a well known phenomenon. It mainly happens because every regional minimum forms its own catchment basins; even though it is very small and insignificant. Here minima selection is used to overcome over segmentation so that the required features can be extracted from the segmented results. “Figure 3.3” shows the segmentation results for one of the image being tested.



a)

b)

Fig 3.3: a) An M-FISH image truth b) Its segmented image

FEATURE EXTRACTION AND CLASSIFICATION

A feature vector is computed for each segmented area in the M-FISH image data and classified using naive Bayes Classifier.

Feature Extraction

The feature vector is a six dimensional vector which includes intensity information from five channels other than DAPI and the size of the chromosome in which the pixel belongs to. In conventional karyotyping, one of the important features selected was found to be the size of the chromosome. Here also, it is incorporated along with multispectral information so that a better representation of chromosomes can be obtained.

In order to reduce intensity inhomogeneities, a morphological operation, dilation is also performed. It finds the local maxima in the defined neighborhood and replaces the centre pixel with that value. In order to apply dilation, flat structuring element object containing 3 neighbors were used. The intensities which are dilated by the above approach are extracted from all the five channels of an M-FISH image.

Classification

The classification is done by Bayesian classifier [21], they predict class membership probabilities based on the Bayes theorem and the maximum *a posteriori* hypothesis. We classify 46 chromosomes into 22 autosomes and 2 sex chromosomes (w=24)
Let the feature X be a d-component vector with d=6 and let P(X|w_i) be the class conditional probability density function with P(w_i), the prior probability that a feature vector belongs to class $w_{i,}$, i=1 ..24. Then the posterior probability that the feature vector X belongs to class w_i, given the feature vector X can be computed by Bayes formula.

$$P(w_i|X) = \frac{P(X|w_i)P(w_i)}{p(X)}$$

Where $p(X) = \sum_{i=1}^{24} P(X|w_i)P(w_i)$

The terms P(X|w_i), p(X_i) and P(w_i) are computed from training data by fitting a Gaussian Mixture Model, i.e. P(X|w_i) = G(X, μ_i, $\sum_i$)
The general multivariate Gaussian density function in d dimension is given by

$$\mathrm{P}\ (X|w_i) = \frac{1}{(2\pi)^{\frac{d}{2}}|\sum i|^{\frac{1}{2}}} * \exp\left(-\frac{1}{2}(x-\mu_i)^t * \sum_i^{-1}(x-\mu_i)\right)$$

We calculate the posterior probability for each class i, and the class to which the feature belongs is obtained by the Bayes decision rule,

Decide $w_i\ if\ P(w_i|X) > P\left(w_j|X\right) for\ all\ j \neq i.$

Post-processing

It includes reclassifying the small misclassified segments to one of its neighbors. The misclassification shown by the circles in “Figure 4.1” were corrected by post-processing.

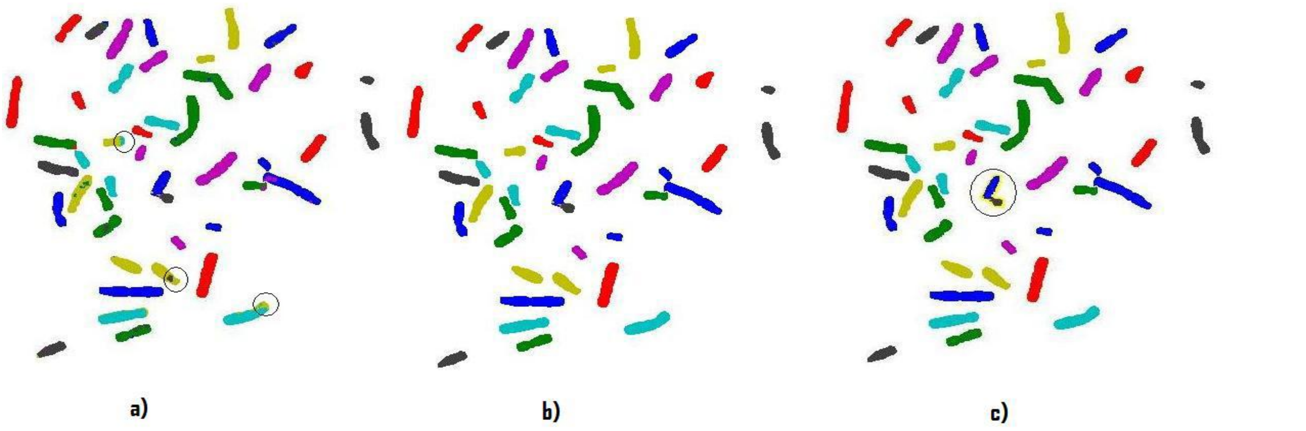


Fig 4.1: a) before post-processing b) after post-processing

The classified pixels are analyzed further to identify the areas of occlusions in the chromosomes using the minimum entropy algorithm [5]. The connected component in the classified image is considered as separate objects and its entropy is calculated. The entropy of a segment will be zero, if all the pixels in that segment are classified in to the same class. The number of different classes with in a segment results into higher values of entropy. The touching chromosome in "Figure 5.1 b)" is identified using the minimum entropy method and is shown in "Figure 5.1 c)"

The proposed approach employs pixel by pixel classification based on intensity and size with pre-processing and post-processing. For the purpose of comparison of the proposed approach, we have also implemented the pixel by pixel classification using intensity only, and pixel by pixel classification using intensity and size. A Bayes classifier was built for the above approaches using training data.

EXPERIMENTAL RESULTS

M-FISH Chromosome Image Database

The database contains 200 multispectral images of size 517x 645. An M-FISH image set contains five spectral images, DAPI channel and ground truth except for "difficult to karyotype" (having many abnormal chromosomes). The dataset contains images for three different probes (ASI, PSI and VYSIS). In the ground truth image, class number is assigned as intensity for chromosome pixels; background pixels and overlapped region values are marked as 0 and 255 respectively.

Segmentation and Classification Accuracy are defined as

$$\textbf{Segmentation Acc.} = \frac{\textit{Chromosome pixels correctly segmented}}{\textit{Total number of chromosome pixels}}$$

$$\textbf{Classification Acc.} = \frac{\textit{Chromosome pixels correctly classified}}{\textit{Total number of chromosome pixels}}$$

"Table 1" shows the classification accuracy for proposed method, pixel by pixel classification with intensity and size method and pixel by pixel classification with intensity only method. The method was tested on 9 images of M-FISH Chromosome Image Database**.** Training was done using a few images that have been hand segmented. Images in the ASI and PSI batch of the dataset are used for training and testing the above approaches. Separate training set was made for ASI and VYSIS probe images. The proposed method obtained an average classification accuracy of 91.09%. It is observed that the conventional criteria along with multispectral information provided better feature selection.

Table 1: Classification Accuracy

#images	Classification Accuracy		
	Pixel by pixel with intensity only	Pixel by pixel with intensity and size	Proposed method
1	86.5	92.3	95.27
2	68.83	85.77	94.83
3	85.17	86	92.88
4	71.60	83.57	91.3
5	85.9	88.7	90.13
6	60.16	86.38	90.28
7	60	77.85	88.9
8	54.23	63.76	88.8
9	57.13	76.38	85.94
Average	70	82.3	91.09

Classification Map

"Figure 5.1" shows the actual ground truth and the class map generated for one of the M_FISH image tested.

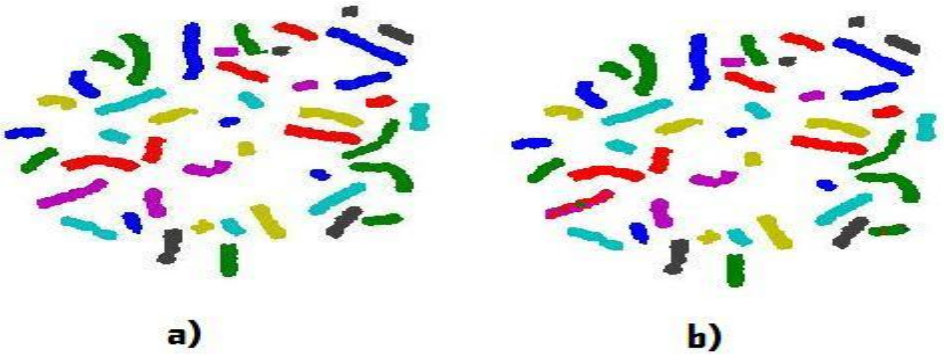


Fig 5.1: a) Actual ground truth b) Generated class map

CONCLUSION

Chromosomes are the structures that contain genetic information of an individual. Automation of human chromosome analysis is a vital task to simplify the process of karyotyping. Though there are a number of attempts by various researchers in karyotyping, the performance of the automatic karyotyping system is still not acceptable for commercial deployment. This paper presents pixel by pixel classification based on intensity and size with pre-processing and post-processing. The approach provided an average classification accuracy of 91.09%. It is observed that introduction of dilation operation and post-processing achieved significant improvement in the classification results. Future work is to extend this approach to include larger set of images and to propose a method to handle occlusions in the chromosomes.

REFERENCES

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

[2] http://www.adires.com/05/Project/ MFISH_DB/ MFI SH_DB.shtml

[3] P. M. Nederlof, S. van der Flier, J.Wiegant, A. K. Raap, H. J. Tanke, J. S. Ploem, and M. van der Ploeg, "Multiplefluorescence in situ hybridization,"*Cytometry*, vol. 11, pp. 126–131, 1990.

[4] Sampat M. P., A. C. Bovik, J. K. Aggarwal, and K. R. Castleman, "Pixel-by-Pixel Classification of MFISH Images," Proc. 24th IEEE Intern. Conf. on Biomedical Engineering Society, 2002, vol. 2, pp. 999-1000.

[5] W. Schwartzkopf, B. L. Evans, and A. C. Bovik, "Minimum Entropy Segmentation Applied to Multi-Spectral Chromosome Images", Proc. IEEE Int. Conf. on Image Processing, 2001, vol. 2, pp. 865-868.

[6] 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.

[7] Wang Y and Ashok Kumar Dandpat, "Classification of M-FISH Images using Fuzzy C-means Clustering Algorithm and Normalization Approaches," 38th Asilomar Conference on Signals, Systems and Computers, vol. 1, Issue 7-10, Nov, pp. 41 – 44, 2004.

[8] P. S. Karvelis, D. I. Fotiadis, M. Syrrou, and I. Georgiou, "A watershed based segmentation method for multispectral chromosome images classification," in Proc. 28th IEEE Ann. Intern. Conf. (EMBS), New York, 2006, pp. 3009–3012.

[9] P Karvelis, A Tzallas, D. Fotiadis, and 1. Georgiou, "A multichannel watershed-based segmentation method for multispectral chromosome classification," IEEE Trans. on Med. Imag., vol. 27, no. 5, pp. 697- 708, 2008.

[10] Sreejini KS, Lijiya, A.and Govindan, VK , "M-FISH Karyotyping- A new approach Based on Watershed Transform, "In International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), pp.105-107, vol.2, no.2, April 2012.

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

[12] Hyohoon Choi, Kenneth R. Castleman and Alan C. Bovik, " Color Compensation of Multicolor FISH Images," IEEE Trans. On Med. Imaging, vol. 28, no. 1, pp. 129-136, January 2009

[13] Georgiou. I., P. Sakaloglou, P. S. Karvelis and D. I. Fotiadis, "Enhancement of the Classification of Multichanne Chromosome Images using Support Vector Machines," 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, September 2-6, pp- 3601-3604, 2009.

[14] Karvelis, Petros and Likas, Aristidis and Fotiadis, Dimitrios I," Semi unsupervised M-FISH chromosome image classification," In 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), pp. 1—4, 2010.

[15] Cao H. and Y. Wang, "Segmentation of M-FISH images for improved classification of chromosomes with an adaptive Fuzzy C-Means Clustering Algorithm," In IEEE international symposium on biomedical imaging: from nano to macro, pp 1442–1445, 2011.

[16] Sangeetha M.K, Govindan, V K and others, "Segmentation and Classification of M-FISH Human Chromosome Images, ," In International Conference on Advances in Computing and Communications (ICACC), pp. 102—105, 2012.

[17] Choi H., A. C. Bovik, and K. R. Castleman, "Maximum-likelihood decompositio overlapping and touching M-FISH chromosomes using geometry, size and color information, " In Proceedings of the 28th IEEE EMBS annual international conference, 2006, pp 3130–3133, 2006.

[18] Qing Chen, Xiaoli Yang and Emil *M.* Petritítle, "Watershed segmentation for binary images with different distance transforms," In Proc. The 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications, 111—116,2004.

[19] Pinaki Pratim Acharjya and Dibyendu Ghoshal, "Watershed Segmentation based on Distance Transform and Edge Detection Techniques," International Journal of Computer Applications vol. 52– No.13, August 2012.

[20] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 2 edition, Prentice- Hall, 2002.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification", SanDiego: Harcourt Brace Jovanovich, Second ed., November 2000Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.