



Occlusion Detection in M-FISH Human Chromosome Images

Lijiya A, Joshy Joseph, V.K.Govindan

Asst. Professor, Department of CSE, National Institute of Technology Calicut ,India

lijiya@nitc.ac.in

M.Tech Student, Department of CSE, National Institute of Technology Calicut ,India

joshyjoseph.act@gmail.com

Professor, Department of CSE, National Institute of Technology Calicut ,India

vkg@nitc.ac.in

ABSTRACT

Automation of Chromosome Analysis has long been considered a tedious task due to the partial occlusion of chromosomes. This calls for a non-trivial, dedicated procedure to segment chromosomes. In this paper, a new method is proposed which detects and separates occluded chromosomes, by separating out the chromosome cluster from the M-FISH image, followed by detecting the cut-points along which these clusters can be split into multiple regions. These regions are then combined into separate partial chromosomes based on difference matrix. After this stage the invisible regions due to occlusion is reconstructed based on the visibility in the five channels. The performance of the new proposal was compared with the existing work and observed better performance in resolving occlusions. With 15 occluded chromosome images tested, 90% accuracy was obtained

Indexing terms/Keywords

Occlusion Detection, M-FISH images, Chromosome Detection.

Academic Discipline And Sub-Disciplines

Image Processing, Pattern Recognition.

SUBJECT CLASSIFICATION

Computer Science.

TYPE

Research Work.

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS AND TECHNOLOGY

Vol. 13, No. 4

editorijctonline@gmail.com

www.cirworld.org/journals

INTRODUCTION

Chromosomes are structures found in the nucleus of cells, which are part of the DNA and contain all the genes of the host organism. Normally Human Beings have 46 chromosomes which are grouped into 23 pairs. The first 22 pairs are called autosomes, and the last pair is "sex chromosome" or gonosome, which could be either XX or XY. Chromosome analysis can be useful in diagnosing many diseases. Karyotyping is a test to identify and evaluate the size, shape and number of chromosomes in a sample of body cells. Extra or missing or abnormal positions of chromosome pieces can be identified by Karyotyping. All these defects can cause health issues for the host. A Karyotype is a photograph in which the chromosomes are arranged in pairs in the ascending order of the chromosome number.

Traditionally, the procedure of assigning each chromosome to a class (Karyotyping) was being carried out based on the visual scanning of chromosome images by experts in the domains of biology [1]. This is a time consuming and expensive process, and requires scientists with extra-ordinary skills. Hence, automated image chromosome analysis could prove a boon in this scenario. In Multicolor fluorescence in-situ hybridization (M-FISH) technology, chromosomes are labeled with five dyes and a DNA stain known as DAPI, which attaches to DNA and labels all chromosomes. Each dye is visible only in a particular wavelength and can be captured using a specific filter. Therefore, M-FISH signals can be obtained as multi-spectral or multi-channel images, in which each chromosome can be stained to be made visible or not visible. Hence, five spectrums are sufficient to distinguish the 24 classes of chromosomes in human genome.

The major challenge faced by automatic karyotyping methods is the presence of occlusion. Occlusion refers to the partial or complete concealment of a chromosome by other chromosomes. The presence of occlusion reduces the accuracy of extracted feature for chromosome classification. Hence, manual intervention is required to finish the process of Karyotyping which could consume more time. The presence of occlusion in an image of metaphase Chromosomes is shown in Figure 1.

A new method is proposed to detect occlusion and separate the occluded chromosomes from the M-FISH image. The proposed method involves determining the connected components and detecting the cluster, detecting cut-points within the cluster and splitting the cluster into multiple regions, merging the regions together and separating the chromosomes from the chromosome clusters based on visibility in the five channels.

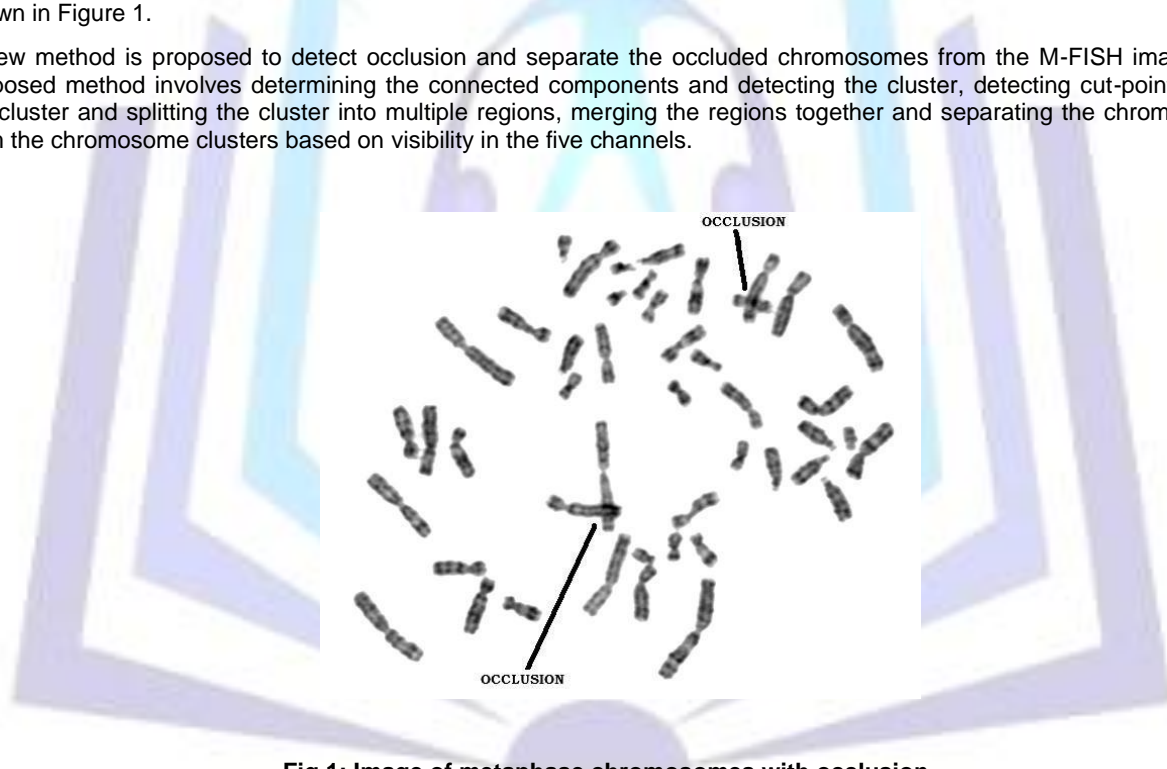


Fig 1: Image of metaphase chromosomes with occlusion

LITERATURE SURVEY

There are already some attempts by researchers on detecting and separating occluded chromosomes. Notable approaches on the gray scale chromosome images includes various segmentation methods, such as thresholding [2, 3] and region-growing [4,5] heuristic search based edge-linking methods [6,7], and shape decomposition method that makes use of fuzzy subset theory[8].

Segmentation methods try to separate the touching chromosomes by classifying the chromosome pixels and the background pixels into two different segments by thresholding chromosome images. Since these methods do not depend on the shape of objects they fail in many cases. Heuristic search based edge-linking methods try to separate touching chromosomes by searching for a minimal-cost connected path that separates the chromosomes. Since the objective of such methods is to find the separating path between the objects and not the objects themselves, their link to the specific problem of chromosome separation is done by a rough determination of the path characteristics which leads to inferior results.



Wade Schwarzkopf et al.[9] proposed a new method for classifying M-FISH chromosome in the presence of occlusion. The approach first determines all the cut points. Once the proper cut points are found out, separation of the chromosome from the cluster is done by searching for a cut region having minimum Shannon's entropy calculated using the 6-feature intensity values extracted from 6 channels of the image. In another paper [10], they proposed an extension to their previous work. This work employs nearest neighbor distances to estimate entropy from raw image data to accomplish minimum entropy segmentation without requiring pixel classification. A problem found with this method is that it fails to find the occluded regions if two chromosomes of same type overlapped each other.

Choi et al. [11] presents a novel decomposition method, for overlapping and touching chromosomes, that utilizes the geometry of a cluster also into consideration. They proposed a set of hypothesis for each type of chromosome cluster and used a maximum-likelihood classification algorithm on these hypotheses for classification. Authors claim about 90% of accuracy was obtained for two or three chromosome clusters, which consist about 95% of all clusters with two or more chromosomes.

Petros et al. [12] proposes a technique that consists of three main steps: the recursive watershed transform computation, the computation of gradient path and the region merging process. Watershed segmentation is applied on binary image to split the image into many small regions. Next, all gradient paths are computed and the binary chromosome area is split along the gradient path. Regions are classified independently using Bayes classifier. Then all neighboring regions that belong to the same class are merged together. The authors claim 90.6% success rate for touching chromosomes and 80.4% for overlapping groups of chromosomes.

Mousami et al.[14] proposed a technique for separating the touching chromosomes using the modified snake algorithm to disentangle the cluster of touching chromosomes from the metaphase image and then a greedy approach based on combinatorial computational geometry of the pixels on the boundary of the cluster is used to identify and resolve the set of touching chromosomes. Performance of this work lies in the ability of the algorithm to successfully separate the clusters of any number of touching chromosomes.

PROPOSED METHOD

1. Finding connected components:

DAPI image is segmented using global thresholding for separating the chromosome from the background. Then, all the connected components are detected. A set S of pixels is a connected component if there exists at least one path in S that joins every pair of pixels $\{p, q\} \in S$ and the path must contain only pixels in S . In other words, a connected component is a set of foreground pixels such that there exists an 8-connected path between every pair of pixels. An issue observed after segmentation using global thresholding was that a set of chromosomes which are adjacent may appear as a single cluster of chromosomes after thresholding. This happens due to the intensity in-homogeneity in MFISH images. To overcome this problem we propose the following method.

Algorithm 1

- for each connected component in the thresholded image,
 1. Identify the region corresponding to the connected component from the DAPI image.
 2. Apply k-means clustering with $k=3$ and initial centroid values as low, middle, and high intensity values of image pixels on the identified portion of the image.
 3. Set pixels within the low value cluster as background and set the middle and high value clusters as foreground.
 4. Detect the connected-components from the foreground pixels and eliminate the connected-components with size less than a pre-defined threshold.
 5. If there are more than two connected-components, then replace the previous set of the connected-component(thresholded image) with the new result
-

2. Detecting cut-points and splitting the cluster into multiple regions:

Detecting cut-points is the most important part of separating chromosomes and identifying occlusion. To detect cut-points, first of all, separate out boundary of the cluster. Then split the boundary of the cluster into various line segments. Now the boundary has become a sequence of line segments. An example is shown in Figure 2. For finding the cut-points, the following method is applied.

Algorithm 2 :

- for each line segment, say $line_i$
 1. Take the next line segment adjacent to $line_i$, say $line_{i+1}$.
 2. Extend $line_i$ and $line_{i+1}$ by k , $\{5 \leq k \leq 10\}$ pixels from their coinciding point.
 3. If at least $k/2$ number of extended pixels, of each line segment, are inside the boundary of the cluster then their coinciding point is a cut-point.
-

Figure 3 shows the cut points found in cluster. After cut-points are found out, split the cluster into multiple regions by drawing a line between the cut-points that are adjacent to each other. Two cut-points are said to be adjacent, if there exists a path of line segments in the cluster from first cut-point to the second such that no other cut-points come in the path. Now, the cluster of chromosomes is split into multiple regions where each region is a part of a single chromosome unless it is the region of occlusion. Figure 4 illustrates this.

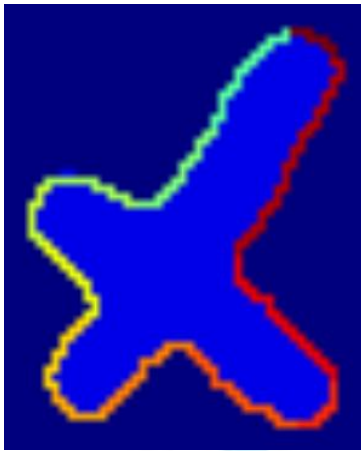


Fig 2: boundary split into line segments.

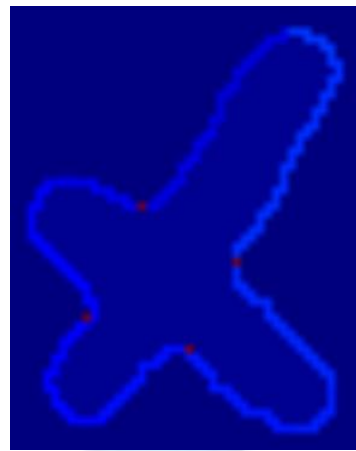


Fig 3: cut-points multiple regions

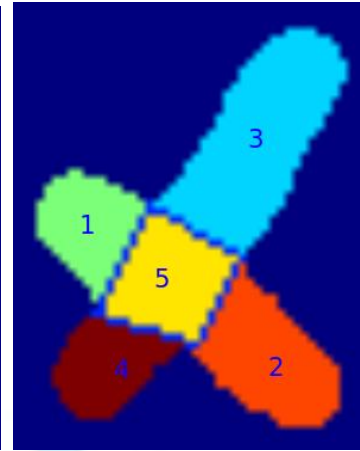


Fig 4: cluster splits into multiple regions

3. Merging the regions together based on difference matrix:

For finding the regions which are part of a single chromosome, a difference matrix is constructed first. This matrix shows the closeness of each region with other regions. Method to construct the difference matrix is described in Algorithm 3 given below:

Algorithm 3:

1. Find the regions corresponding to the cluster from each channel.
2. Apply k-means with $k=2$, initial centroids set to low and high intensity pixel values respectively, to each of these regions so that we can separate high intensity valued pixels and the low intensity valued pixels in two groups.
3. Find Perc (i, j) the percentage of high intensity valued pixels in region _{i} in channel _{j} to the total number of pixels in each region _{i} of all the channels, i.e.

$$\text{Perc}(i, j) = \frac{N_{ij} \times 100}{N_i}$$
 where, N_{ij} = Number of high intensity valued pixels in the region of channel _{j}
4. Now Compute $P_k(m, n)$, i.e. the difference between a region with another region given a channel k

$$P_k(m, n) = \text{abs}(\text{Perc}(m, k) - \text{Perc}(n, k))$$
 Where, $0 \leq m, n < \text{no. of regions}$
5. Construct the difference matrix $D(m, n)$ of size $m \times n$ by doing a matrix addition of all the five $P_k(m, n)$

$$D(m, n) = \sum_{k=0}^5 (P_k(m, n))$$

The difference matrix represents the similarity between every pair of regions i, j ; i.e. smaller the value corresponding to the regions i, j in difference matrix, higher the probability that the regions are the part of a single chromosome. Now, we have to combine the regions which are part of the same chromosome. It can be done as in Algorithm 4 given below:

Algorithm 4:

1. Find the region with smallest value in the difference matrix $D(m, n)$, and let the index of this element be i and j
2. These (i, j) corresponds to the regions which are most similar to each other.
3. Combine these regions $\text{reg}_i, \text{reg}_j$, as below :
 - a. If the regions share part of the boundary, merge the pixels in the first region with the second region. Mark regions as merged chromosomes
 - b. Otherwise, combine the regions through the pixels connecting these regions such that resulting region contains all the pixels. Mark regions as cross-cut chromosomes.
4. Add the combined region into the partial-chromosome set.
5. Set the values of i th and j th row of $D(m, n)$ to infinity.
6. Repeat steps 1 to 5 until either the entire elements are infinite or smallest element in the matrix is greater than a threshold.
7. If there are any non-infinity elements in $D(m, n)$, check whether this region is already a part of a cross-cut chromosome. If not, add this region to partial-chromosome set.

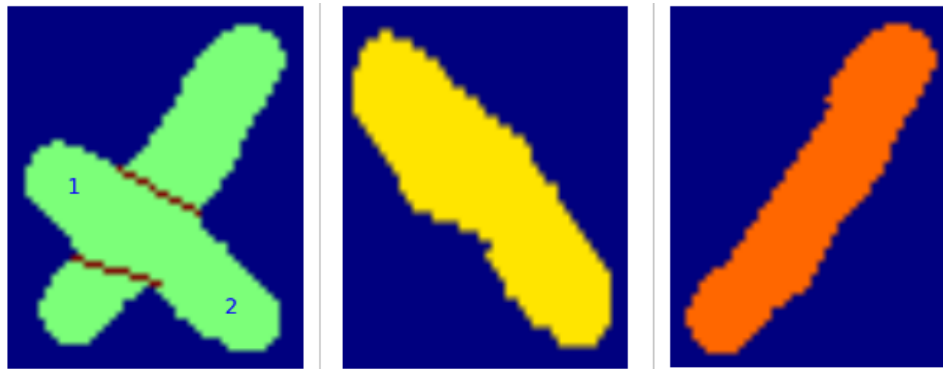


Fig 5: Separated partial chromosomes

4. Separating the chromosomes based on the visibility in the five channels:

We have now got partial-chromosome information. Because of occlusion, some part of the chromosome is not included in the extracted region. Getting this information is a challenging one as the occluded portion of a chromosome is visible only in a subset of channels of the M-FISH image. So, it is necessary to grow the partial-chromosome already identified to the occluded parts as well. This requires the selection of a channel which clearly makes this chromosome visible among the five regions. Channels where most of the pixels in the cluster are either low or high are avoided. The following algorithm performs the task

Algorithm 5 :

1. for each partial chromosome marked as merged-chromosome
 2. for each channel,
 - a. Compute L_i , the percentage of low value pixels within the partial-chromosome to the total number of pixels within the cluster.
 - b. Compute H_i , the percentage of high value pixels within the partial-chromosome to the total number of pixels within the cluster
 - c. Find the channel with the largest percentage value of L_i and H_i among all the channels and select that channel for growing the image.
 - d. If the highest value is obtained from L_i , grow the region using low pixel values; otherwise grow the region using high pixel values.
 - e. Add each grown partial-chromosome to the chromosome set.
 3. Add partial chromosome marked as merged-chromosome to the chromosome set
-

5. Occlusion Detection:

Now, the occluded portion is detected by taking the intersection between these identified chromosome pixels. Occluded pixels are those pixels with in these identified chromosomes which are shared by more than one chromosome.

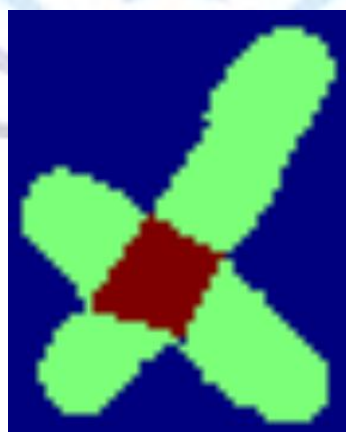


Fig 6: Middle portion show occluded region



RESULT

The algorithm was tested on 15 images from the database [13]. The results were compared with an already existing method [11]. Some of the images of occluded chromosomes processed are shown in Figures 5 to 10 .

Table 1: Accuracy of the proposed approach with and without occlusions the energy efficient route on the basis of minimum total transmission energy of the route.

Number of chromosome in cluster	Accuracy	
	Proposed method	Choi[11]
1	100%	100%
2	92 %	89%
3	90 %	89%

The proposed work was compared with the existing works as given in Table 1, and the results were analyzed. An overall accuracy of 90% was observed when the number of chromosomes in the occluded cluster is 3. It was also noted that for some cases, the proposed work gave better results while giving the same result in other cases. An improvement of this method over the existing methods is that in some of the cases this method can also detect the occlusion in cluster even if the same chromosome pair occluded each other.

CONCLUSION AND FUTURE WORK

A new method is proposed to detect and separate occluded chromosomes by separating out the chromosome cluster from the M-FISH image. Overlapping chromosomes were successfully detected and separated in most of the cases. When tested with 15 images with occluded chromosomes, an accuracy of 90% was observed. The overall accuracy of the method greatly depends upon the accuracy of Cut-Point detection, which in turn depends upon the accuracy of detecting Line Segments. Use of improved techniques for line detection may improve the overall accuracy of Karyotyping with occlusions. Another issue observed in the proposed approach is that some region which is actually not a part of the occluded chromosome is also added to the chromosome in some cases. This issue needs to be addressed for further improving the classification accuracy.

REFERENCES

- [1] M Thompson, R. McInnes, and H Willard. "Genetics in Medicine", Saunders. 1991.
- [2] J. Piper, E. Granum, D. Rutovitz, and H. Rutledge. "Automation of chromosome analysis", IEEE Trans. Signal Processing, 1980.
- [3] P. Nickolls, J. Piper, D. Rutovitz, A. Chisholm, I. Johnstone, and M. Robertson. "Pre-processing of images in an automated chromosome analysis system", IEEE Trans. Pattern Recognition, 1981.
- [4] J.M. Chassery and C. Gaybay. "An iterative segmentation method based on a contextual color and shape criterion", IEEE Trans. Pattern Analysis and Machine Intelligence, pages 794-800, 1984.
- [5] C. Gaybay. "Image structure representation and processing: A discussion of some segmentation methods in cytology", IEEE Trans. Pattern Analysis and Machine Intelligence, pages 140-146, 1986.
- [6] J. Liang. "Intelligent splitting in the chromosome domain", IEEE Trans, Pattern Recognition, pages 519-532, 1989.
- [7] D.H. Cooper, N. Bryson, and C.J. Taylor. "An object location strategy using shape and greylevel models", Image and Vision Computing IEEE, pages 50-56, 1989.
- [8] L. Vanderheydt, F. Dom, A. Oosterlinck, and H. Van Den Berghe. "Two-dimensional shape decomposition using fuzzy subset theory applied to automated chromosome analysis", IEEE Trans, Pattern Recognition, pages 147-157, 1981.
- [9] Wade C. Schwartzkopf, Alan C. Bovik, Brian L. Evans, "Minimum entropy segmentation applied to multi-spectral chromosome images", IEEE Trans, 2001
- [10] Wade C. Schwartzkopf, Alan C. Bovik, Brian L. Evans. "Entropy estimation for segmentation of multi-spectral chromosome images", IEEE Trans, 2002.



- [11] Hyohoon Choi, K.R. Castleman and A.C. Bovik. "Joint segmentation and classification of m-fish chromosome images", In Engineering in Medicine and Biology Society, 2004. IEMBS04. 26th Annual International Conference of the IEEE, (25):1536-1639, September 2004.
- [12] Petros Karvelis, Aristidis Likas a, and Dimitrios. I. Fotiadis. "Identifying touching and overlapping chromosomes using the watershed transform and gradient paths", IEEE Trans. Med.Imaging, 2010.
- [13] The ADIR M-FISH Image Database. <http://www.adires.com/05/Project/MFISHDB/MFISHDB.shtml>, Accessed: 21-Jul-2013
- [14] Mousami V Munot, Madhuri A. Joshi and Nikhil Sharma, "Automated Karyotyping of Metaphase Cells with Touching Chromosomes", International Journal of Computer Applications (0975 – 8887) Volume 29– No.12, September 2011.
- [15] L. Ji, "Fully automatic chromosome segmentation," Cytometry, vol. 17, pp. 196–208, 1994The ADIR M-FISH Image Database. <http://www.adires.com/05/Project/MFISHDB/MFISHDB.shtml>, Accessed: 21-Jul-2013
- [16] Vadivel, R and V. Murali Bhaskaran, 'Energy Efficient with Secured Reliable Routing Protocol (EESRRP) for Mobile Ad-Hoc Networks', Procedia Technology 4, pp. 703- 707, 2012.

