# A Novel Approach for Segmentation of Human Metaphase Chromosome Images Using Region Based Active Contours

Tanvi Arora
Department of Computer Science and Engineering, Dr. B.R Ambedkar National Institute of Technology, India

**Abstract:** *The chromosomes are the genetic information carries. A healthy human being has 46 chromosomes. Any alteration in either the number of chromosomes or the structure of chromosomes in a human being is diagnosed as a genetic defect. To uncover the genetic defects the metaphase chromosomes are imaged and analyzed. The metaphase chromosome images often contain intensity inhomogeneity that makes the image segmentation task difficult. The difficulties caused by intensity inhomogeneity can be resolved by using region based active contours techniques. These techniques uses the local intensity values of the nearby regions of the objects and find the approximate intensity values along both sides of the contour. In the proposed work a segmentation technique has been proposed to segment the objects present in the human metaphase chromosome images using region based active contours. The proposed technique has been quite efficient from prospective of number of objects segmented. The method has been tested on Advanced Digital Imaging Research (ADIR) dataset. The experimental results have shown quite good performance.*

**Keywords:** *Chromosomes, segmentation, active Contours, intensity inhomogeneity.*

*Received December 8, 2015; accepted April 17, 2016*

## 1. Introduction

A human being may have missing fingers, abnormal shape of face, epilepsy, diseases of blood, learning in ability and many more such conditions, the cause of which cannot be detected unless the chromosomes of an individual are analyzed. For this purpose the doctors recommend the effected individuals to go for genetic testing. The genetic testing is done in order to find out the cause of these abnormalities as chromosomes are the genetic information carriers. A healthy human has 46 chromosomes [26], out which 22 chromosomes are paired and the 23rd and 24th chromosome are the sex determining chromosomes, they may be XX or XY in the case of female and male respectively. A person is said to have genetic defect if there is a mismatch either in the number of chromosomes or in the structure of the chromosomes.

The chromosomes are present inside the nucleus of the cell as a thread like structures. In order to capture the chromosomes for the purpose of analysis they are imaged using a light microscope that too during the metaphase stage of cell division, as during that stage the chromosomes are visible as the longest. For the purpose of imaging the cell division phase is inhibited by treating the cells with colchicines or colcemid [4, 6] as shown in Figure 1.

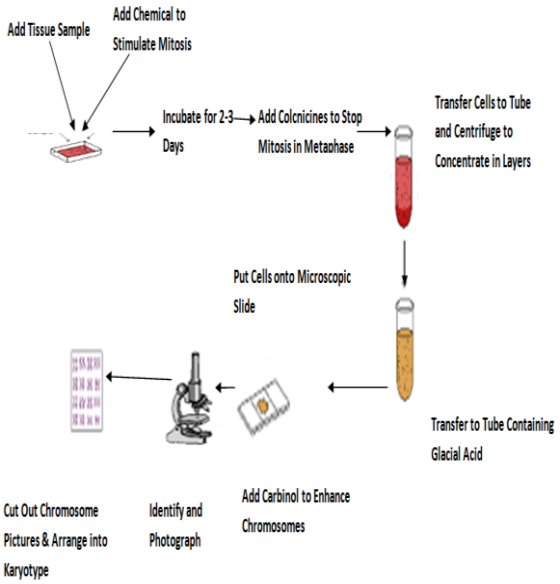


Figure 1. Metaphase chromosome imaging process.

In order to analyze the chromosomes the chromosomes are extracted from the metaspread images and arranged in an order so as to form a karyogram as shown in Figure 2. The steps for the Karyogram generation are illustrated in Figure 3. The segmentation process is very challenging as the captured images have intensity inhomogeneity and secondly due to the non-rigid structure of the chromosomes they are present in different orientations.

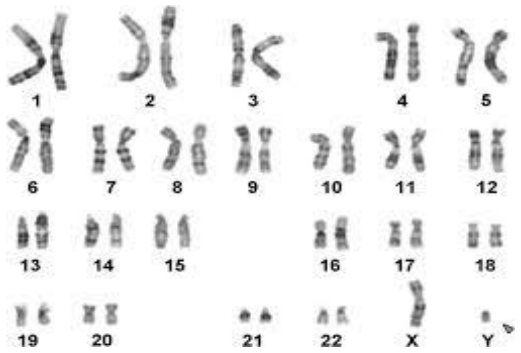


Figure 2. Karyogram of human chromosomes.

In this paper a segmentation approach is being presented to segment the human metaphase chromosome images using region based active contours. This technique uses the local intensity values of the nearby regions of the objects and finds the approximate intensity values along both sides of the contour. The technique is not specific to any particular type of metaspread chromosome image. The touching and overlapping chromosomes will be disentangled by finding the cut points using concave and convex points along the boundary of the clusters of chromosomes. The best cut will be performed using the gradient values of the cut points.

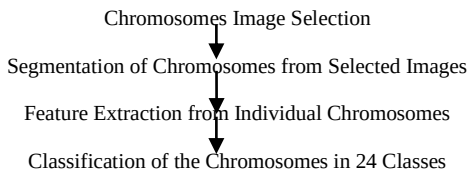


Figure 3. Karyogram generation process.

The remaining paper is organized as follows: section 2 contains related work, section 3 contains materials and methods, section 4 contains results and discussion and section 5 concludes the paper.

## 2. Related Work

In the initial years the karyotyping was done manually where in the technician used to cut the images of chromosomes and arrange them in the order of karyogram. Since a large number of analysis were to be performed, this task was very tiring and time consuming [8]. Then in late 1970's systems were developed for semi-automatic human assisted karyotype generation [1, 8]. A combination of filtering and thresholding techniques have been used for enhancing and segmenting the chromosome images [11, 19] for the creation of automatic karyotyping systems. The proposed techniques were able to decompose the cluster of chromosomes into individual chromosomes but resulted in the over segmentation of the chromosome images. The thresholding based segmentation methods were not able to resolve the problem of touching or overlapping chromosomes [20].

Geometry based segmentation methods were proposed to handle these issues[1, 16, 17, 25].

A computational geometry based approach was proposed for segmentation of touching and overlapping chromosomes, it aimed at finding the cut points and then segmenting the touching and overlapping chromosomes [3, 24].

To overcome the issues associated with touching and overlapping chromosomes in M-FISH images a hypothesis based approach was proposed for pixel classification for the purpose of segmentation [13, 23]. To overcome the partial success of the proposed methods a maximum likelihood framework [7] was proposed, using the pixel classification, geometric information of objects and size. In this multiple hypothesis were framed and then evaluated for the purpose of segmentation.

A recursive watershed based approach was proposed for Multiplex Fluorescence in Situ Hybridization (M-FISH) images, in which the problem of touching and overlapping chromosomes was resolved using multichannel gradient paths. To resolve the ambiguities the region merging techniques were used [12].

A recent advancement has been to segment the chromosomes from the metaspread by measuring the mass of the chromosomes using the optical and X-Ray ptychography [23]. But this method works only for isolated chromosomes. The problem of intensity inhomogeneity in the background of chromosome images and the variability of the fluorescence in chromosomes was addressed by proposing a region based level set approach [2, 9, 14], for segmentation of the Q banded prometaphase images. This segmentation approach was better as far as pixel sensitivity was concerned or number of clusters that were separated. But this approach could not work for images that had many chromosomes in the clusters.

Comparative studies were carried out for segmentation algorithms in which adaptive thresholding and region based level set methods had the best performance [5, 21] as they are able to work quite well in the presence of inhomogeneous conditions.

## 3. Materials and Methods

### 3.1. Image Dataset

To measure the performance of the proposed segmentation technique the images of the metaspread has been taken from Advanced Digital Imaging Research (ADIR) chromosome image database. The ADIR database is having a total of 200 multispectral images and an associated ground truth files giving the details of the translocations, deletions, missing or extra chromosomes.

### 3.2. Performance Evaluation Metrics

The accuracy of the proposed segmentation approach can be measured using accuracy of segmentation. The parameter can be calculated as follows:

- True Positive (TP): Objects termed as single chromosomes and they proved to be single chromosome.
- False Positive (FP): Objects termed as single chromosomes and they proved to be cell artifacts or cluster of chromosomes.
- True Negative (TN): Objects termed as cell artifacts or cluster of chromosomes they proved to be cell artifacts or cluster of chromosomes.
- False Negative (FN): Objects termed as cell artifacts or cluster of chromosomes they proved to be chromosomes.

Accuracy: It is the ability to classify correctly. More the accuracy batter the method.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

### 3.3. Proposed Method

The aims of the proposed segmentation technique are as follows:

1. To segment the chromosomes present in the metaphase images as efficiently as that can be done by an expert human cytogeneticist.
2. To find the cut points and disentangle the touching and overlapping chromosomes.
3. The proposed segmentation technique can segment the chromosomes irrespective of the straining method or the type of image.
4. To compare the results of the segmentation with the standard ground truth information, so as to evaluate the performance of the proposed method for the accuracy of segmentation.

The flow chart of the proposed method is depicted in the below Figure 4. The proposed segmentation method consists of the various steps like object extraction from the metaspread chromosome image, classifying the extracted objects as chromosomes, cluster of chromosomes or artifacts, disentangling the chromosomes from cluster of chromosomes and calculating the accuracy of segmentation. These steps are explained below:

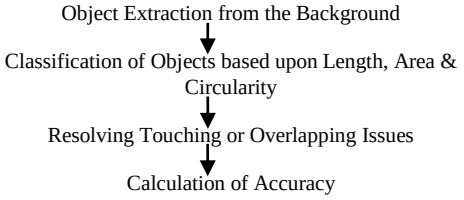


Figure 4. Proposed method for segmentation.

#### 3.3.1. Object Extraction from the Chromosome Metaspread Image

To extract the objects from the chromosome metaspread image the region based active contour model has been used. This model retrieves the intensity of pixels in the nearby regions. It is capable of segmenting the images even in the presence of intensity inhomogeneity and can perform quite well for images of weak boundaries. As illustrated in [16] the following equation can be used to compute the level set evolution:

$$\frac{\partial \varphi}{\partial t} = -\delta_\varepsilon(\varphi)(\lambda_1 e_1 - \lambda_2 e_2) + \nu\delta_\varepsilon(\varphi)\text{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) + \mu\left(\nabla^2\varphi - \text{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right)\right) \quad \delta_\varepsilon \quad (2)$$

is the Dirac delta function that can be computed as follows:

$$\delta_\varepsilon(\text{x}) = \text{H}_\varepsilon(\text{x}) = \frac{1}{\pi}\frac{\varepsilon}{\varepsilon^2 + \text{x}^2} \quad (3)$$

$H_\varepsilon(x)$ is approximated smoothened Heaviside function. $e_1, e_2$ are fitting energy functions they can be computed as:

$$\text{e}_\text{i}(\text{x}) = \int \text{K}_\sigma(\text{y} - \text{x})\left|\text{I}(\text{x}) - \text{f}_\text{i}(\text{y})\right|^2 \text{dy} \quad (4)$$

where i =1,2.

$f_1(x)$ and $f_2(x)$ are the fitting values, they can be computed using:

$$\text{f}_\text{i}(\text{x}) = \frac{\text{K}_\sigma(\text{x}) * \left[\text{M}_\text{i}^\varepsilon(\varphi(\text{x}))\text{I}(\text{x})\right]}{\text{K}_\sigma(\text{x}) * \text{M}_\text{i}^\varepsilon(\varphi(\text{x}))} \quad (5)$$

The Equation (2) has been used to evolve the level set and segment the chromosomes and cluster of chromosomes from the background image. It has three components, the first term is the data fitting term that has been derived from the data fitting energy. This term draws the active contour along the boundaries of the objects. The second term is the arc length term, that is used for smoothening of the contour, the third term is the regularization term, it is used to regularize the level set.

#### 3.3.2. Classifying Objects as Chromosomes

After the objects have been extracted from the background they underwent a test to find out if there is a single chromosome, cluster of chromosomes or cell residues or artifacts. The classification of the objects is done based upon the parameters of area, circularity and length. The area is calculated considering the number of pixels present in the object and it can be calculated as:

$$Area = \sum I(x) \text{ where } \text{I}(\text{x}) \neq 0 \quad (6)$$

The circularity can be calculated by finding the ratio of the Area of the object and Area of the equivalent

circle that has its center at the center of gravity of the object.

$$\text{Circularity} = \frac{\text{Area}_{\text{object}}}{\text{Area}_{\text{circle}}} \quad (7)$$

The length of the object can be calculated by finding the medial axis transform. The length is calculated by counting the pixels along the longest path.

### 3.3.3. Extracting Chromosomes from Chromosome Cluster

The objects that are classified as cluster of chromosomes, may contain either touching chromosomes or overlapping chromosomes. In order to extract the individual chromosomes from the clusters the boundary of the cluster of chromosomes is extracted. From the boundary the curvature of the boundary points is calculated. Using the curvature the convex and concave points of the boundary can be found based upon the angle of the curvature. Let $C_1, C_2, C_3, \ldots$ be the boundary points. The angle of boundary points can be found by using the following:

$$a(i) = \arccos\left(\frac{(C_i - C_{i-k}).(C_{i+k} - C_i)}{\|C_i - C_{i-k}\|.\|C_{i+k} - C_i\|}\right) \text{sgn}\left[\det(C_i - C_{i-k} \ldots C_{i+k} - C_i)\right] \quad (8)$$

As reported in [12] in case of overlapping chromosomes the intensity of the overlapping regions is relatively high and homogeneous. To separate the touching and overlapping chromosomes concept of gradient paths have been used. The cut points are used to find the gradient paths. The cut points are those boundary points that have *a(i)* greater then a particular threshold value. A large number of boundary points qualify as cut points. In order to limit the number of cut points, the points of the neighborhood are grouped and from each group the cut points that have the maximum *a(i)*, are chosen as the cut points for the computation of the gradient paths.

The gradient path is calculated by first choosing any one cut point $c_i$. The initial direction of the gradient path is taken as the bisector of the angle of the cut points adjoining that cut point. The next pixel candidate is taken considering the gradient value of the adjoining pixels. The pixel that has the highest gradient value is traced. This process is repeated till a boundary point is reached. The direction of search is changed after three points. Like this the gradient paths are traced for all the cut points. To segment the touching and overlapping chromosomes the lines of the gradient paths split the chromosomes into multiple regions.

Now this technique of separation of the chromosome clusters into multiple regions result into over segmentation. To overcome this drawback the region merging approach is used. The region merging of the over segmented parts of the clusters is performed by matching the chromosomes with their homologous pairs. By trying different combinations the over segmented chromosomes are reconstructed.

## 4. Results and Discussion

The proposed approach has been tested on ADIR dataset. The ground truth for the number of overlapping and touching chromosomes has been calculated and is illustrated in Table 1.

Table 1. Ground truth for touching and overlapping chromosomes in ADIR dataset.

| Count of Touching Chromosomes | Count of Overlapping Chromosomes |
|---|---|
| **1178** | **189** |

The segmentation of touching group of chromosomes is said to be correct if the method is able to produce the desired number of individual chromosomes, from the touching cluster of chromosomes. The results of separation for touching group of chromosomes for Schwartzkopf's method [22], Ji's method [10], Karveli's method [12] and the proposed method are presented in the Table 2 below.

Table 2. Comparative analysis of proposed method with [10,12,22] method's for segmentation of touching chromosomes.

| Method | Accuracy |
|---|---|
| **Schwartzkopf's** | 77% |
| **Ji's** | 84.2% |
| **Karveli's** | 90.6% |
| **Proposed** | 96.7% |

The segmentation of the overlapping chromosomes is the most tedious of all the steps. The results of the proposed method for the segmentation of the overlapping chromosomes is compared with the results of segmentation of Schwartzkopf's method [22] and Karveli's method [12] in the Table 3.

Table 3. Comparative analysis of proposed method with [12,22] for segmentation of overlapping chromosomes.

| Method | Accuracy |
|---|---|
| **Schwartzkopf's** | 34% |
| **Karveli's** | 80.4% |
| **Proposed** | 81% |

As evident from the analysis and the results proposed by various researchers touching and overlapping chromosomes are difficult to segment, after long period of research the accuracy of the results is still not 100%. The segmentation results for the touching chromosomes are quite encouraging as compared to previous methods. But the segmentation results for overlapping chromosomes are not so satisfactory. Second significant contribution is that, it is applicable to any type of chromosome images like G Banded Images or Q Banded Images, but for the purpose of comparison the results are shown for MFISH images.

## 5. Conclusions

A region based active contours method has been proposed for the segmentation chromosomes, from the metaspreads of human chromosomes. The proposed

method was able to resolve the issue of touching chromosomes to a great extent, as compared to previous methods. However the domain of overlapping chromosomes could not achieve much performance improvement. The main contributions of the proposed work are that it is able to handle the metaspread images having non homogeneous background. Secondly this segmentation technique is a general purpose technique as it can be applied to any type of images. Thirdly only three features have been considered for the purpose of classifying the chromosomes as either single chromosomes or cluster of touching or overlapping chromosomes. It is hoped that the proposed technique will be quite fruit-full in the domain of segmentation of touching and overlapping chromosomes.

## References


[1] Agam G. and Dinstein I., "Geometric Separation of Partially Overlapping Nonrigid Objects Applied to Automatic Chromosome Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1212-1222, 1997.

[2] Arora T. Dhir R., "Segmentation of Human Metaspread Images Using Region Based Active Contours," *in Proceedings of in International Conference on Recent Trends in Engineering and Material Science*, Jaipur, pp. 1-5, 2016.

[3] Arora T. and Dhir R., "An Efficient Segmentation Method for Overlapping Chromosome Images," *International Journal of Computer Applications*, vol. 95, no. 1, pp. 29-32, 2014.

[4] Arora T. and Dhir R., "A Review of Metaphase Mhromosome Image Selection Techniques for Automatic Karyotype Generation," *Medical and Biological Engineering and Computing*, vol. 54, no. 8, pp. 1147-1157, 2015.

[5] Arora T. and Dhir R., "Segmentation Approaches for Human Metaspread Chromosome Images Using Level Set Methods," *in Proceedings of International Conference on Mass Data Analysis of Images and Signals*, New York, pp. 13-30, 2017.

[6] Bickmore W., *Karyotype Analysis and Chromosome Banding*, Wiley Online Library, 2001.

[7] Castleman K., Choi H., and Bovik A., "Maximum-likelihood Decomposition of Overlapping and Touching M-FISH Chromosomes using Geometry, Size and Color Information," *in Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Society*, New York, pp. 3130-3133, 2006.

[8] Granlund G., "Identification of Human Chromosomes by Using Integrated Density Profiles," *IEEE Transactions on Biomedical Engineering*, vol. 23, no. 3, pp. 182-192, 1976.

[9] Grisan E., Poletti E., and Ruggeri A., "An Improved Segmentation of Chromosomes in Q-Band Prometaphase Images Using a Region Based Level Set," *in Proceedings of World Congress on Medical Physics and Biomedical Engineering*, Munich, pp. 748-751, 2009.

[10] Ji L., "Intelligent Splitting in the Chromosome Domain," *Pattern Recognition*, vol. 22, no. 5, pp. 519-532, 1989.

[11] Ji L., "Fully Automatic Chromosome Segmentation," *Cytometry*, vol. 17, no. 3, pp. 196-208, 1994.

[12] Karvelis P., Likas A., and Fotiadis D., "Identifying Touching and Overlapping Chromosomes Using the Watershed Transform and Gradient Paths," *Pattern Recognition Letters*, vol. 31, no. 16, pp. 2474-2488, 2010.

[13] Kumar T. and Reddy K., "A Technique for Burning Area Identification Using IHS Transformation and Image Segmentation," *The International Arab Journal of Information Technology*, vol. 12, no. 6A, pp. 764-771, 2015.

[14] Li C., Huang R., Ding Z., Gatenby J., Metaxas D., and Gore J., "A Level Set Method for Image Segmentation in the Presence of Intensity Inhomogeneities with Application to MRI," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2007-2016, 2011.

[15] Li C., Kao C., Gore J., and Ding Z., "Minimization of Region-Scalable Fitting Energy for Image Segmentation," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1940-1949, 2008.

[16] Minaee S., Fotouhi M., and Khalaj B., "A Geometric Approach to Fully Automatic Chromosome Segmentation," *in Proceedings of IEEE Signal Processing in Medicine and Biology Symposium*, Philadelphia, pp. 1-6, 2011.

[17] Moallem P., Karimizadeh A., and Yazdchi M., "Using Shape Information and Dark Paths for Automatic Recognition of Touching and Overlapping Chromosomes in G-Band Images," *International Journal Image, Graphics and Signal Processing*, vol. 5, no. 4, pp. 22-28, 2013.

[18] Nickolls P., Piper J., Rutovitz D., Chisholm A., Joitnstone I., and Robertson M., "Pre-Processing of Images in an Automated Chromosome Analysis System," *Pattern Recognition*, vol. 14, no. 1-6, pp. 219-229, 1981.

[19] Oosterlinck A., Daele J., Boer J., Dom F., Reynaerts A., Berghe H., "Computer-Assisted Karyotyping with Interaction," *The Journal of Histochemistry and Cytochemistry*, vol. 25, no. 7, pp. 754-762, 1977.

[20] Piper J. and Granum E., "On Fully Automatic Feature Measurement for Banded Chromosome

Classification," *Cytometry*, vol. 10, no. 3, pp. 242-255, 1989.
[21] Poletti E., Zappelli F., Ruggeri A., and Grisan E., "A Review of Thresholding Strategies Applied to Human Chromosome Segmentation," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 679-688, 2012.
[22] Schwartzkopf W., Bovik A., and Evans B., "Maximum-Likelihood Techniques for Joint Segmentation-Classification of Multispectral Chromosome Images," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1593-1610, 2005.
[23] Shemilt L., Verbanis E., Schwenke J., Estandarte A., Xiong G., Harder R., Parmar N., Yusuf M., Zhang F., and Robinson K., "Karyotyping Human Chromosomes by Optical and X-Ray Ptychography Methods," *Biophysj*, vol. 108, no. 3, pp. 706-713, 2015.
[24] Somasundaram D. and Kumar V., "Separation of Overlapped Chromosomes and Pairing of Similar Chromosomes for Karyotyping Analysis," *Measurement*, vol. 48, pp. 274-281, 2014.
[25] Srisang W., Jaroensutasinee K., and Jaroensutasinee M., "Segmentation of Overlapping Chromosome Images Using Computational Geometry," *Walailak Journal of Science and Technology*, vol. 3, no. 2, pp. 181-194, 2006.
[26] Tjio J. and Levan A., "The Chromosome Number of Man," *Genetics*, vol. 42, no. 1-2, pp. 1-6, 1956.

**Tanvi Arora** is currently working as associate professor in the department of Computer Science & Engineering at Baddi University of Emerging Sciences and Technology, Baddi, Himachal Pradesh, India. Her teaching and research interests include Image Processing, Pattern Recognition.