Automated Karyotyping of Metaphase Cells with Touching Chromosomes

Mousami V. Munot
Assist. Prof., Dept of E&TC, Pune Institute of Computer Technology, Pune, India

Dr. Madhuri A. Joshi
Dean R&D, Prof, Dept of Electronics, Government College of Engineering, Pune

Nikhil Sharma
Student, Dept of E&TC, Maharashtra Institute of Technology, Pune, India

ABSTRACT
Since the birth of the automated karyotyping systems by the aid of computers, building a fully automated chromosome analysis system has been an ultimate goal. Along with many other challenges, automating chromosome classification and segmentation has been a major challenge especially due to overlapping and touching chromosomes. The earlier reported methods have limited success as they are sensitive to scale variations, computationally complex, use only color information in case of multispectral imaging and had challenges in segmentation. The proposed technique addresses the challenge of separating the touching chromosomes using initially 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. Contribution and novelty of this work lies in the ability of the algorithm to successfully separate the clusters of any number of touching chromosomes. System performance was tested and analyzed using a variety of metaphase images exhibiting various levels of touching chromosomes giving an overall accuracy of 100 % for resolving the cluster with 2 touching chromosomes and 95 % for separating a cluster of 3, 4 touches. The overall time was 2.4 seconds.

Keywords
Karyotyping, Metaphase, Combinatorial Computational Geometry, Snake algorithm.

1. INTRODUCTION
Human chromosome analysis is an essential task in cytogenetics, especially in prenatal screening and genetic syndrome diagnosis, cancer pathology research and environmentally induced mutagen dosimetry [1, 4]. Metaphase is the stage of cell division at which the chromosomes are most suitable for analysis. [2,3,5]. A normal human diploid cell contains 22 pairs of chromosomes, autosomes of classes 1-22 and 2 sex chromosomes, either XX or XY. The karyotype displays chromosomes in standard positions based on their length, centromeric index and band pattern as seen in figure 1.

The process of manual karyotyping is usually carried out by expert clinicians who initially identifies each chromosome in the picture using hierarchical chromosome identification and then finally using expert knowledge view the pictures, identify the chromosomes, cut and place them in their specified locations in the karyotype and classifies them into smaller groups [3]. The

visual inspection is thus tedious, time consuming, laborious and an expensive procedure. Hence, many attempts have been made to automate the process of karyotyping [5, 10]. Automated Karyotyping systems allows countless clinical advantages such as interactive and graphical environment, faster in the accomplishment of the samples, allowing quality printing, being self explanatory, better interpretation of the image, and it still makes possible the storage of the information in a database for future analysis [15].

The results of the automated systems though encouraging, have limited success, still needing human interaction. This is mainly due to non rigid nature of chromosomes thus requiring special to separate touching and overlapping chromosomes. Methods described in the literature for the separation of touching chromosomes include various segmentation methods, such as thresholding or region growing, heuristic search edge linking methods and shape decomposition method that uses fuzzy subset theory and trainable shape models. But these general methods have reported to give inferior results, are sometimes sensitive to scale variation, computationally complex or depend highly on the curvature and fail in case of cluster [16-18] .Another greater reported disadvantage of fuzzy based method is that it also yields erroneous decomposition of single chromosomes in cases of bent chromosomes or when the contraction near the centromere is too sharp [6, 7]. In mid of 1990s, Multicolor fluorescence in-situ hybridization (MFISH), a multispectral combinatorial labeling technique in which minimum of five fluorophores are used to stain each chromosome was a boon for the automated karyotyping systems [8, 11]. MFISH provides color karyotyping by assigning a pseudo colour to each pixel based on the spectral combination, thus making possible easy identification of each chromosome. The touching cases can be easily discerned when the pixel memberships are presented by two distinct colours as in MFISH. Colour information is itself sufficient for the segmentation of the chromosomes but when only colours are used, touchings and overlaps of the same kinds of chromosomes cannot be segmented and segmentation accuracy highly relies on initial pixel classification. The maximum likelihood decomposition methods for MFISH had limited success in separation of touching chromosomes [9-12].

Thus resolving the touches and overlaps in a metaphase image is still an open issue and a major hindrance in the development of automated karyotyping systems due to scale variation, computational complexity , segmentation challenges and also because only color information is used in MFISH Imaging.
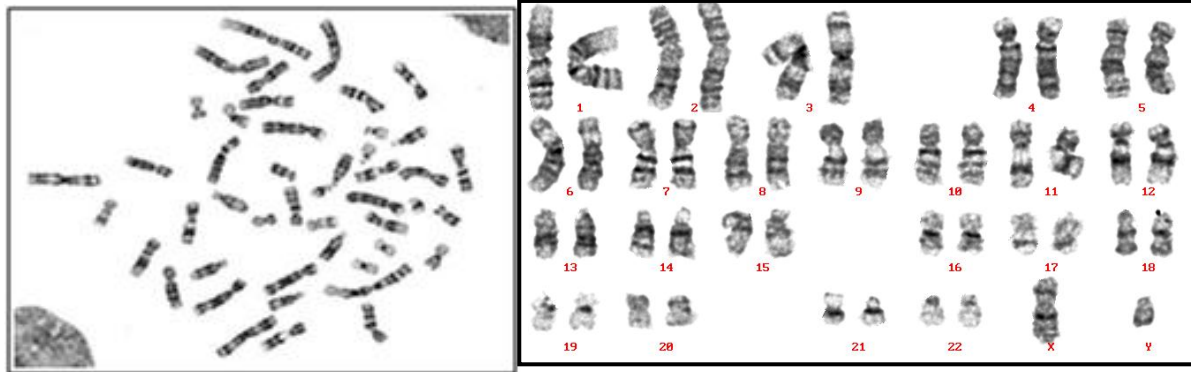
<div align="center">(a) (b)</div>

Fig. 1. (a) A metaphase cell spread (b) A Karyotype of the chromosome in (a)

2. PROPOSED APPROACH

In this paper, a novel and an efficient approach for separating the touching chromosomes by using computational geometry is proposed.
The basic steps in this approach are

- Segmenting the cluster of touching chromosomes from the metaphase image using a modified snake algorithm
- Detection of pixels on the boundary of the cluster
- Finding the touching point /points between two or more chromosomes in the cluster.
- Connecting the boundary of the chromosomes at the touching pixels found

2.1 Problem formulation:

The problem of automatically karyotyping a metaphase image is devised by defining C_i as the set of all pixels belonging to class *i*. since there are 24 classes of chromosomes, each non background pixel may be classified as one of the 24 classes [12]. Here in this approach, the problem of separating the touching chromosome is formulated as to find the potential touching points and then cutting the chromosome t that point.

This problem is modeled as a geometric query problem of combinatorial computational geometry where the target is finding the closest pair of points. In geometric query problems, commonly known as geometric search problems, the input consists of two parts: the search space part and the query part, which varies over the problem instances. The search space typically needs to be preprocessed, in a way that multiple queries can be answered efficiently [20]. The closest pair of points problem or closest pair problem is a problem of computational geometry: given *n* points in metric space, find a pair of points with the smallest distance between them.

In the problem of separating the touching chromosomes, the 'n' points are the pixels on the boundary of the cluster. To find the touching points in the cluster, the distances between all the $n (n − 1) / 2$ pairs of points must be computed and finally the pair with the smallest distance is selected as the touching point. The same can be described as :

minDist = infinity
for each *p* **in** *P*:
 for each *q* **in** *P*:
  if $p \neq q$ **and** *dist(p, q)* < *minDist*:
   minDist = *dist(p, q)*

 closestPair = (*p*, *q*)
return *closestPair*

In our application, choice of distance is Manhattan distance because it requires less computations compared to Euclidean distance. This distance measure required for the comparison need not be very accurate and is more of a relative comparison thus the easier and the measure requiring less time and processing is chosen.

The Manhattan distance, d_1, between two vectors p, q, in an *n*-dimensional real vector space with fixed Cartesian coordinate system, is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes and is stated as: [21]

$$d_1 (p, q) = \| p\text{-}q \|_1 = \sum_1^n | p_i\text{-}q_i | \qquad \ldots\ldots\ldots\ldots\ldots (i)$$
$$\text{where} \quad p = (p_1, p_2, \ldots p_n) \text{ and } q = (q_1, q_2, q_3 \ldots q_n) \ldots\ldots\ldots\ldots (ii)$$

2.2 Segmentation using modified snakes algorithm:

The reported limitation in active shape models using snake algorithm problem is when trying to locate the contour of an object which has boundary concavity no snake is able to move into the concavity to find the counter. The reason that the snake did not stop sooner is that new points keep being inserted in the horizontal stretch above the cavity. These new points move out to the sides and snake never converges according to stopping criteria and unable to move in cavity. This behavior is caused by a combination of the snake's internal energies and the image energy. Image energy pulls the snakes downwards and the elastic energy keeps the snake from stretching. This is shown in figure 2. [22].

Chromosomes are non rigid structures and so shape variability is natural. Moreover in this particular case of resolving the clusters of touching chromosomes, high degree of boundary concavity is expected and the original snake algorithm will be inefficient to segment the cluster. We propose a modification in the snake algorithm as highlighted in algorithm, figure 3 to successfully overcome this limitation.
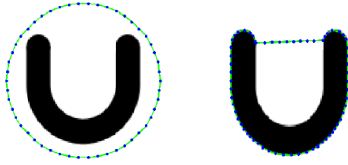
fig 2. Illustrating an object with a boundary concavity and the results of applying snakes algorithm
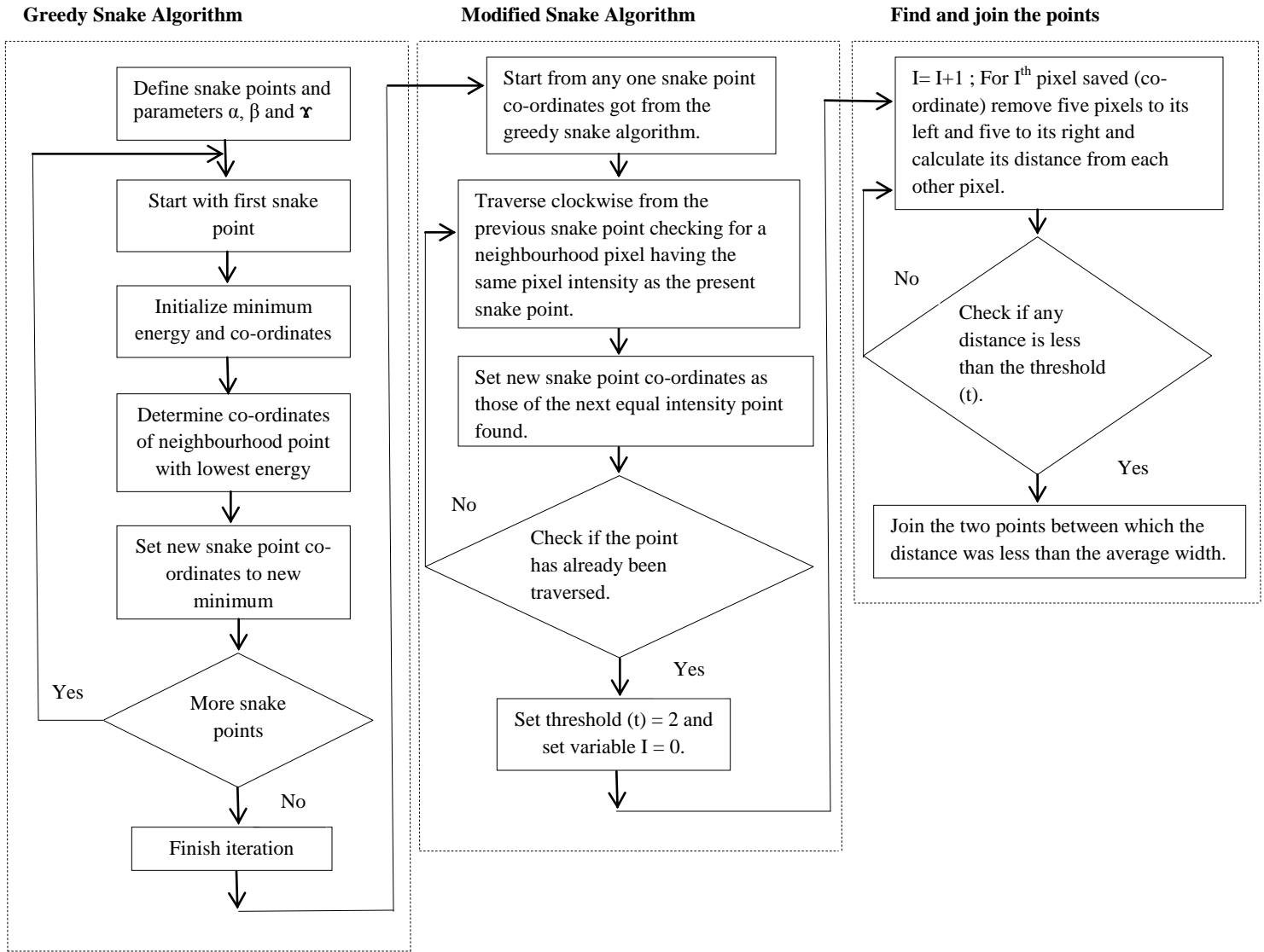
Greedy Snake Algorithm

Modified Snake Algorithm

Find and join the points

Fig 3: The proposed algorithm indicating the intermediate steps

2.4 Evaluation of the potential touching points and disentangling the touching chromosomes:

Upon the successful separation of the cluster of touching chromosomes, the Manhattan distance between every pixel on the boundary and 'nth' particular pixel is calculated. Image distances graph for one such pixel, which is actually the touch point of the chromosomes is shown in figure 4. Two pixels with the minimum distance between them will be the touch points of the chromosome and will have to be cut so to disentangle the touching chromosomes. Here the average width is heuristically taken as '2'. Thus the point under consideration and the point having distance less than the average width are connected. A cluster may have any number of touches and any number of chromosomes so the problem is dealt as a local search and not global which would have resulted in always searching out only one touch point.
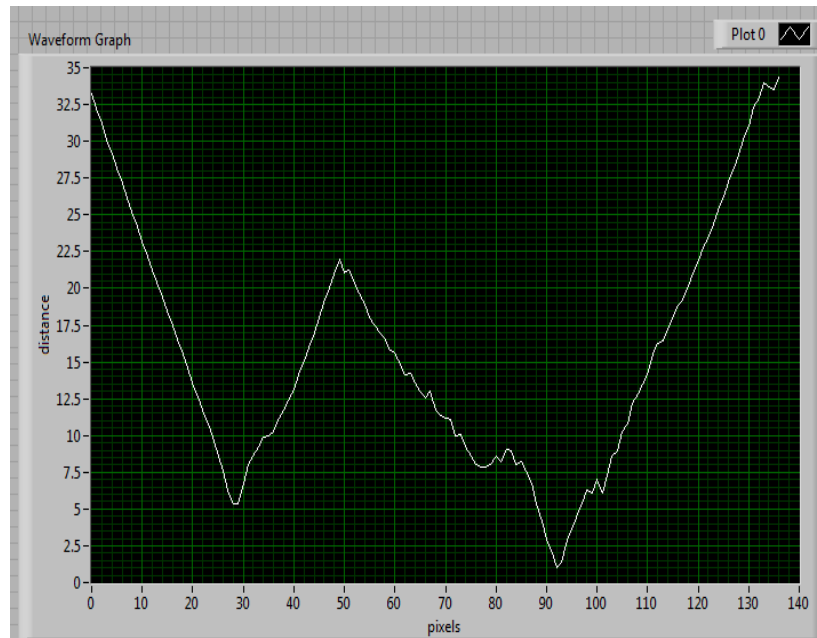
Fig. 4 : Graph showing the distance of one pixel (touch point) point from rest all the other pixels on the boundary of the cluster

3. EXPERIMENTS AND RESULTS

3.1 Data Base:
System performance was tested and analyzed using a variety of metaphase images exhibiting various levels of touching chromosomes

The data base used to test the proposed approach contains metaphase images from

- Publically available database from BioImLab, Laboratory of Biomedical Imaging, department of Information Engineering, University of Padova, Italy, which has the original metaphase images and their respective karyotyptes available for testing and verifying the results.
- Denanath Mangeshkar Hospital and Research Centre's genetic laboratory The images are acquired by conventional photography using Evolution™ VF color cooled camera.
  photography using Evolution™ VF color cooled camera. Magnification factor for microscope is 1000X. The gray scale resolution of the resulting digitized pictures was set to 256 levels. Features of camera are High Resolution 1.4 pixel sensor, 12 bit digitization, 165fps maximum frame rate and ROI (Region of Interest). Chromosomes were manually segmented and karyotyped by an expert in the

genetic laboratory. These karyotyped images were used as ground truth to test the accuracy of the proposed approach

3.2 Experiments
The proposed approach was implemented in LABVIEW 2009. The snap shots of the intermediate steps to separate the clusters of touching chromosomes are shown in figure 5 a and b. Assuming that chromosomes and clusters of chromosomes form shapes with long contours, the high level noise removal is achieved by deleting objects having relatively short counters.

In the approach to find the closest pair of pixels which are the candidates for the touching points, neighbouring pixels were likely to be identified as the potential touching points because the neighbouring pixels will obviously be the nearest one with the minimum distance between them.

To avoid the neighbouring pixels on the boundary to get identified as the touching points, the distance from the ten pixels on the either side of the nth pixel are neglected and the image distance matrix is computed for all the remaining pixels on the boundary.
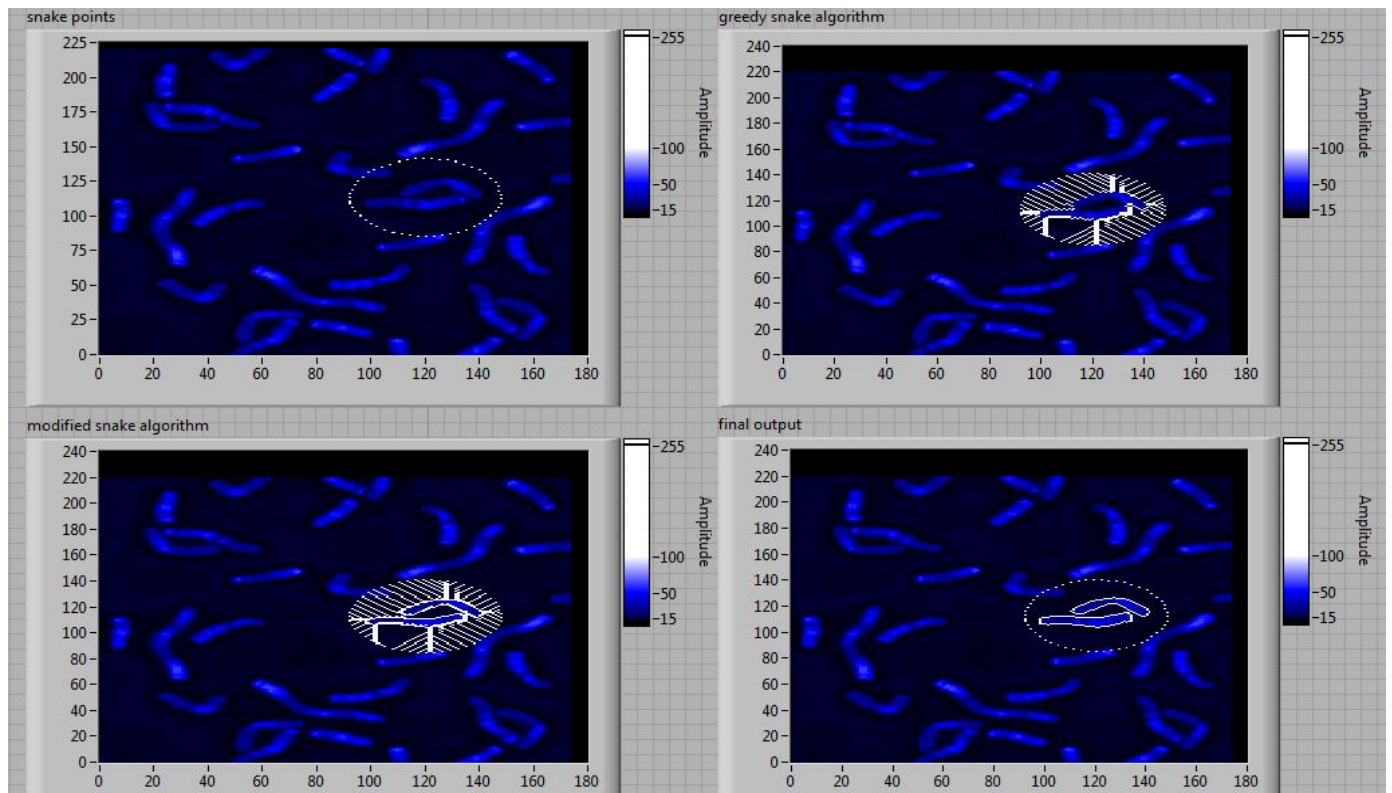
(a)

(b)

Fig 5: a) Intermediate steps for segmenting the touching chromosomes   b) Results of the algorithm for segmenting a cluster of 2, 3 and 4 touching chromosomes

4. DISCUSSIONS AND CONCLUSIONS

In this paper, a novel and an efficient approach for separating the touching chromosomes by using computational geometry is proposed. The proposed technique initially uses 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. Contribution and novelty of this work lies in

- The ability of the modified snake algorithm to successfully separate the clusters of chromosomes with boundary concavity.
- Successfully separating the clusters of any number of touching chromosomes.

The algorithm is independent on the scale variations and is not computationally complex Since the problem is modeled as a query based problem where the search space is fixed, the computational complexity for this class of problems is usually estimated by the time and space required . The time required is 2.4 sec.

Moreover the algorithm is also capable of separating the touching chromosomes in MFISH images of the same class. because colour is not use to resolve the touches.

The details are summarized in Table 1.

Table I : Decomposition Results

N_{CC}	N_C	N_{WD}	Accuracy (%)
2	40	0	100
3	35	2	94.2
4	20	1	95

N_{CC}: Number of chromosomes in a cluster, N_C = Number of Clusters, N_{WD}: Number of Wrong Decomposition

The algorithm outperforms even in critical cases of the cluster with 4 and more number of touching chromosomes in it and has also decomposed all the clusters with two touching chromosomes in it. But because the threshold of the distance is heuristically selected to be   2 , in one case wherein the chromosomes were actually not touching but were very close to each other was also identified as touching and the closest point boundaries were connected. This is shown in figure 5.
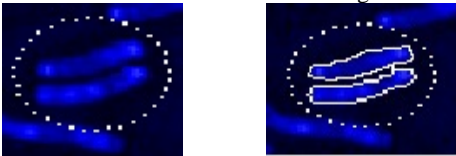

Fig 5. The two chromosomes are not touching but are at a distance of one pixel from each other but the algorithm erroneously joins the two points. Similar two cases of three touching chromosomes also failed for the same reason.

This work contributes in addressing one of the major challenges of separating the touching chromosomes   in automated karyotyping systems. Our further work will address the issue of separating the overlaps in the gray and the MFISH chromosome images.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Standing Committee on human Cyotgenetic Nomenclature, ISCN: an international system for human cytogentic nomenclature. Karger and cytogentics and cell Genetics, 1995.

[2] E. Grisan, E. Poletti, C Tomelleri, A Ruggeri, " Automatic Segmentation of Chromosomes in Q Band images", Proceddings of 29th annual International Conference of the IEEE EMBS , France, August 2007, pp 5513-5516.

[3] F. Cloopet, A. Boucher, " Segmentation of overlapping / aggregating nuclei cells in biological images" 2008

[4] W. Srisang, K, Jaroensutasinee, M. Jaroensutasinee, "Segmentation of overlapping chromosomes images using computational Geometry," Walailak Journal of Science and Technology, 2006, pp. 181-194.

[5] B. Lerner, " Toward A completely automatic neural network based human chromosome analysis", IEEE Transactions on system, Man, cybernetics-part B: cybernetics, vol. 28, no:4, August 1998, pp. 544-552.

[6] B. Lerner, H Guterman, I. Dinstein, " A classification driven partially occluded object segmentation (CPOOS) Method with application to chromosome analysis", IEEE Transactions on signal Processing, vol. 46, no. 10, October 1998 , pp 2841-2847.

[7] G. Agam, I. Dinstein, " Geometric Separation of Partially Overlapping Non rigid Objects applied to Automatic Chromosome Classification" , IEEE Transactions on pattern analysis and machine Intelligence, vol.19. no 11, November 1997.

[8] G. Charters, J Grahman, " Disentangling Chromosome overlaps by combining Trainable Shape models with Classification evidence" IEEE Transactions on signal Processing , vol. 50. no.8. August 2002.

[9] W. Schwwartzkopf, B. Evans, A. Bovik, "Minimum entropy Segmentation Applied to Multispectral Chromosome Images", pp 865-868.

[10] W. Schwwartzkopf, B. Evans, A. Bovik," Entropy Estimation for segmentation of Multi Spectral Chromosome Images", Fifth IEEE Southwest symposium on image analysis and interpretation (SSIAI 02), 2002, pp-

[11] H. Choi, A. Bovik, K. Castleman, " Maximum Likelihood Decomposition of Overlapping and Touching MFISH chromosomes using Geometry, size and color information", Proceedings of the 28th IEEE EMBS Annual International Conference, USA, 2006

[12] W. Schwartzkopf, A. C. Bovik, B. Evans, " Maximum Likelihood Techniques for joint Segmentation - classification of Multispectral chromosome Images, " IEEE Transactions on Medical Imaging, vol 24, No. 12, December 2005.

[13] L. Ji, "Intelligent Splitting in the chromosome domain," Pattern Recognition, vol. 22, pp 519-532, 1989

[14] L. Ji, "Fully Automatic Chromosome Separation," Cytometry, vol. 17.pp 196-208,1994

[15] M. Ppoescu, P. Gader, J. Keller, C. Klein, J. Stanley, C. Caldweli, "Automatic Karyotyping of Metaphase cells with overlapping chromosomes," Computers in Biology and Medicine , vol.29,pp.61-82,1999

[16] A. Carothers, J Piper, "Computer Aided Classification of Human Chromosomes: A review," Statistics and Computing, vol. 4, pp. 161-171,1994

[17] B. Lerner, "Toward a Completely Automatic Neural Network Based Human Chromosome Analysis," IEEE transactions systems, man, and cybernetics-part B: cybernetics, vol. 28, no. 4, pp. 544-552, 1998.

[18] J. Piper, E. Granum, "On Fully Automatic Feature Measurement for Banded Chromosome Classification," cytometry vol. 10, 1989, pp. 242- 255.

[19] A Khemlinskii, R. Ventura, J. Sanches, " A Novel Metric for Bone Marrow Cells Chromosome Pairing," IEEE Transactions on biomedical Engineering, Vol. 57, no. 6, June 2010.

[20] T. Cormen, C. Leiserson, R. Rivest, C. Stein. *Introduction to Algorithms*, Second Edition. MIT Press and McGraw-Hill, 2001. ISBN 0-262-03293-7. Pages 957–961 of section 33.4:

[21] Paul E. Black, *Dictionary of Algorithms and Data Structure-* Manhattan distance*s*, NIST

[22] Nikolas Tiilikainen, "A Comparative Study of Active Contour Snakes" 2007