AN IMAGE PROCESSING ALGORITHM FOR ACCURATE EXTRACTION OF THE CENTERLINE FROM HUMAN METAPHASE CHROMOSOMES

Akila Subasinghe Arachchige[*], *Jagath Samarabandu*[*], *Joan Knoll*[†], *Wahab Khan*[†], *Peter Rogan*[†]

[*]Image Recognition & Intelligent Systems Laboratory,University of Western Ontario,ON,Canada
[†]Schulich School of Medicine & Dentistry,University of Western Ontario,ON,Canada

ABSTRACT

The study of human metaphase chromosomes is an important aspect in clinical diagnosis of genetic disorders. Although many image processing techniques have been developed for chromosomal karyotyping to assist in laboratory diagnosis, they fail to provide reliable results in segmenting and extracting the centerline of chromosomes due to their shape variability when placed on microscope slides. In this paper we propose a hybrid algorithm that uses Gradient Vector Flow active contours, Discrete Curve Evolution based skeleton pruning and morphological thinning to provide a reliable centerline that is robust to shape variations of the chromosomes. Effective identification of the chromosome outline with its centerline provides a basis for further operations such as automated chromosome classification and centromere identification.

Index Terms— Chromosome segmentation, gradient vector flow, discrete curve evolution, medial axis

1. INTRODUCTION

Early detection of genetic disorders using human metaphase chromosomes is a critical stage in clinical diagnosis. Karyotype analysis is one of the main research areas in image processing. It aims at producing annotated karyograms with the least user involvement and therefore effectively reducing the diagnosis time period. Methods available for karyotyping or other chromosome analysis are mainly limited by the shape variability caused by non-rigid nature of the chromosome structure. Therefore, the effectiveness of these image processing techniques are highly limited by the inability to provide proper results irrespective of the shape of the chromosome [1]. Proper segmentation and extraction of the center line of the chromosome plays a vital role in many of the available karyotyping methods [2]. In this research, the image processing techniques were applied to DAPI (4',6-Diamidino-2-Phenylindole) stained chromosomes in contrast to Geimsa banded chromosomes used in many karyotype analysis methods in literature. The image processing techniques discussed in this paper are a part of an algorithm developed to accurately locate FISH (Fluorescent in situ hybridization) probes relative to landmarks on DAPI stained metaphase chromosomes. FISH uses fluorescence DNA probes to detect chromosome sequence rearrangements in genetic diseases.

The commonly used segmentation method is global or local thresholding [3]. Thresholding often yields acceptable results since the intensity histograms of chromosome images are typically bi-modal with a good separation of the peaks. However, the presence of noise and other artifacts in fluorescence microscopy can cause inaccurate segmentation due to thresholding being a point process.

Parametric deformable models are another widely tested segmentation method. The gradient vector flow (GVF) based active contours deliver better results in chromosome image segmentation [4]. GVF was first introduced into parametric active contours(referred to as snakes), to address a main limitation in the traditional active contour model [5] by improving its capture range drastically.

The centerline is a shape descriptor based on the topological skeleton of the object, which produces a longitudinal axis of symmetry. Medial Axis Transformation (MAT) is commonly used in order to achieve this. In practice, using MAT or other morphological operations such as object thinning tend to produce poor results due to the shape variability of chromosomes. Such variations often yield spurious branches during the skeletonization process. Numerous other approaches have been proposed, but most of these avoid using skeletonization or thinning while preventing unwanted branching [1],[2]. Those methods also suffer from the unpredictable shape variations of the chromosome preparations.

In this paper, we propose an algortihm based on GVF snakes, discrete curve evolution (DCE) based pruning and morphological thinning that can effectively yield accurate segmentation as well as centerline extraction which is independent of the chromosome shape. Additionally, the centerline extraction is designed to be robust against object boundary noise present in fluorescence microscope images.

The proposed algorithm is explained in detail in section 2 and its results on actual chromosome images are analyzed in section 3. Concluding remarks are given in section 4.

2. THE PROPOSED ALGORITHM

2.1. Pre-Processing and Segmentation

The initial processing stages prior to the secondary segmentation are listed below,

- Step1: The fluorescent chromosomal images were first normalized and then subjected to a global thresholding process based on Otsu's method [6]. This thresholding algorithm attempts to segment the histogram of the image into two clusters by minimizing their intra-class variance.
- Step2: The contour of the above segmentation result was extracted by removing all 4-connected pixels of the binary image thus leaving only the boundary pixels intact.
- Step3: The inverted DAPI chromosome image was further subjected to a 2D median filtering stage to remove any noise and artifacts present in the image while preserving information regarding the chromosome boundaries.

The rationale for adopting a parametric active contour model was due to the availability of a close approximation of the chromosome shape through thresholding and the presence of strong edges around the chromosomes. Yet, due to unequal illumination across the fields of view, these initial object shapes can be either under-approximated or over-approximated. Therefore, the snake model in question had to be able to either expand or contract into the chromosome object depending on the initial contour positioning. Also, due to the shape variability of chromosomes, we required a snake model that could converge into concave boundaries.

GVF based snakes [5] successfully address all of the above issues and have a significantly higher convergence and capture range compared to the traditional snake model. GVF snakes also have the ability to shrink or expand depending on the gradient vector field which is a diffused field based on the edge map of the image. Unlike Gaussian smoothing which is commonly used to increase the capture range, the use of GVF doesn't blur the edge map and thus leads to more accurate boundary positioning at convergence. GVF based snakes have been used in many segmentation algorithms including human chromosome segmentation [4] and yield satisfactory results compared to other segmentation methods. Therefore, the contour extracted from the reduced binary image was parameterized and those control points were allowed to move into the object boundaries by iterating until convergence.

2.2. Finding the Centerline

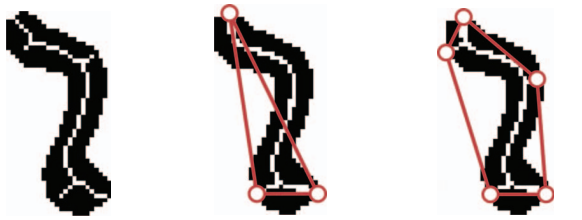
The chromosome centerline is necessary in many operations like classification performed on segmented chromosomes [1],[7]. Many shape and structure-related features such as the chromosomal banding pattern, width and density profiles can be extracted using the centerline. Small deviations in the extraction of these authentic features could result in classification errors [2]. The majority of centerline extracting methods reported in the literature are based on MAT (Medial Axis Transform) and different thinning methodologies [8]. Skeletonization or thinning produces spurious branches frequently at bend locations particularly towards the telomere regions of the chromosomes. The methods that are not based on MAT mainly have problems with handling objects with sharp bends which are commonly present in metaphase chromosomes [1],[2].

We have adopted a skeleton pruning method based on Discrete Curve Evolution (DCE) [9] which in our algorithm was applied only to chromosomes with skeletons longer than a particular length (35 skeletal points) and shorter chromosomes were processed using the thinning algorithm described by Lam [10]. The rationale behind this hybrid application is to use DCE based pruning only on chromosomes which are highly likely to be bent while utilizing thinning on relatively shorter chromosomes for which the skeleton deviates from the centerline. The DCE based skeleton pruning process is based on partitioning the object contour into polygonal sections and then evolving them according to DCE. Furthermore, pruning is achieved by removing all skeletal points of which all the generating points (the points where the maximal disks touch the object boundary) lie on the same polygon partition. Results in this skeletal pruning method tend to be highly dependent on the contour partitioning itself. Therefore the skeleton pruning problem can be viewed as a contour partitioning problem. DCE provides an ideal solution for this by effectively evolving polygon partitions by vertex deletion based on any given relevance measurement[11]. For the implementation, any digital image boundary can be approximated to a polygon without a loss of information by taking each boundary pixel as a vertex on the polygon and similarly considering the distance between each pixel as an edge. DCE was used to evolve the polygon iteratively by removing the vertex which had the least value for the relevance value $K(v,u,w)$ defined in equation 1, where d_{uv} & d_{vw} are the Euclidian length between the vertices and θ is the turn angle at vertex v. This relevance function was selected so that it is dependent on features of its neighbors and thus makes DCE able to evolve using global features of the shape information.

$$K(v,u,w) = (\theta * d_{uv} * d_{vw})/(d_{uv} + d_{vw}) \qquad (1)$$

Also, as DCE is simply deleting vertices of the polygon partitions, the topology information is guaranteed to be represented at the skeleton ends. Furthermore we have considered only the convex polygon combinations in order to prune spurious branches effectively [9],[12]. Figure 1 depicts the reliability and accuracy of the DCE based pruning method

compared to standard pruning. Figure 1(b) and figure 1(c) depict two DCE based pruning results for different number of vertices for the end convex polygon.



(a)Skeleton (b)DCE triangle (c)DCE pentagon

Fig. 1. Comparison between standard skeleton with DCE based solutions on a bent chromosome.

In the case of obtaining the medial axis of a chromosome, the ideal result would be a pruned skeleton with no extra branches. Yet, as the minimum convex polygon being a triangle and DCE being modeled as polygons, the resulting skeleton will at minimum have one spurious branch. This issue was resolved by tracing all branches and pruning the shortest branch completely. The DCE result was then processed by a modified thinning algorithm to ensure single pixel thickness of the skeleton. Our modified thinning algorithm consisted of the application of a set of masks to the skeleton on the basis of the morphological hit & miss algorithm followed by a thinning process described by Lam [10]. A curve fitting step was then introduced to obtain a smooth curve from the skeleton of the previous step. This step was based on cubic spline interpolation which attempts to fit a 3^{rd} order polynomial between each of its control points (knots) while keeping continuity at its end point connections. Therefore cubic spline interpolation is an appropriate method for approximating the centerline of any bent metaphase chromosome. The control points for curve fitting were provided by sampling the skeleton result obtained through DCE pruning step and registering a control point for approximately every 7 skeleton points after excluding some portion of each end. The interval of '7' points above was selected empirically in order to avoid over fitting the data while representing the shape information adequately. The end section clipping was performed to remove the skeletal portion that deviated at the telomere regions from the actual centerline.

Finally, a methodology was developed to correct the end points of the centerline using a heuristic gradient based end point detection method. This was achieved by creating a template which matched telomere regions of a chromosome (see table 1). The line segments used in this were extended segments of the end points of the sampling points used for spline curve fitting stage. The lengths of these segments were selected to be 20% of the centerline.

0	0	0	0	0	0	0
0	0	-	-	-	0	0
0	0	-	-	-	0	0
0	-	-	+	-	-	0
0	-	+	+	+	-	0
-	+	+	0	+	+	-
0	+	0	0	0	+	0

Table 1. A 7x7 represention of the original 20x20 template used for end point correction, where the coefficients were set as '0' - ignored, '+' as +1 and '-' as -1.

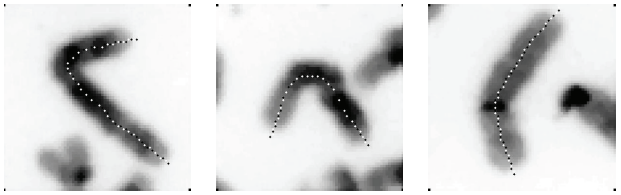
3. RESULTS AND DISCUSSION

The proposed algorithm was tested on 120 chromosomes extracted from 15 inverted DAPI stained lymphocyte cell images captured using an epifluorescence microscope. The chromosomes selected for the DCE based methodology did not overlap or touch each other and they met the minimum length criteria of 35 points (see section 2.2). The centerline extracted using our algorithm was compared with that obtained through a thinning approach [10]. Centerlines drawn by a geneticist were used as the gold standard. Two metrics (MAD - 'mean absolute distance' & MAXD - 'maximum absolute distance') were used to measure error from this gold standard and the results are shown in Table 2. These results show that in general, the DCE based method performs better than thinning in nearly all chromosome groups. Results using DCE were independent of the source of the chromosomal material, i.e. of the patient, slide or cell that was selected. The accuracy of the DCE based method was particularly high in regions of chromosome bends, which occur more frequently in longer chromosomes. Longer chromosomes are well represented in groups A,B & C in table 2, where the improvements of the DCE based method were apparent. In addition, longer chromosomes are found in pro metaphase chromosomes that are just beginning to condense in mitosis. Another category of relatively longer chromosomes can include chromosome rearrangements such as duplications or translocations.

The centerline derived from this algorithm was particularly sensitive to the parameters used for GVF segmentation and selection of sampling points during spline fitting. Firstly, the segmentation outcome was observed to be highly sensitive to the values set for the main internal parameters of the GVF snake such as α (elasticity factor), β (rigidity factor), μ (GVF regularization factor) and κ (external force weight). A sensible set of values for the above factors ($\alpha = 0.05,\ \beta = 0,\ \kappa = 2,\ \mu = 0.2$) provided satisfactory results in our experiments but fine tuning was possible through adjustment. Next, the selection of the sample point spacing and the starting and ending sample point offset (from the DCE result) were observed as critical parameters in the medial axis extraction and values were set based on empirical observations.

Table 2. Matric results (MAD and MAXD) mean values and standard deviation for each chromosome group. The number after each chromosome group name specifies the number of chromosomes present in the data set from that particular group.

Chrom. group	**Proposed method**		**Thinning method**	
	MAD	*MAXD*	*MAD*	*MAXD*
A-21	0.59±0.30	1.68±0.89	0.67±0.31	1.90±1.10
B-22	0.66±0.22	1.69±0.52	0.72±0.30	1.98±0.89
C-55	0.62±0.24	1.86±1.17	0.73±0.28	2.11±1.26
D-11	0.68±0.26	1.52±0.51	0.79±0.37	1.86±0.95
E-09	0.73±0.48	1.73±0.62	0.89±0.56	1.95±1.49
F-01	0.30	1.41	0.49	1.41
G-01	0.62	1.41	0.58	1.41



(a) (b) (c)

Fig. 2. Centerline results of representative bent chromosomes using the proposed algorithm.

The algorithm was also able to successfully handle bent chromosomes (as seen in Fig 2) and extract the centerline that closely represents the shape information of the chromosome. The sharpness of the centerlines in Fig 2 is mainly due to the pixelation effect of the magnified image.

4. CONCLUSION & FUTURE WORK

An algorithm based on GVF active contours, DCE based skeleton pruning and morphological thinning is presented in this paper. The accuracy of the centerline is important as consequences of errors in measuring chromosome lengths could result in inaccurate structural assignments.This algorithm is observed to be robust against image boundary noise as well as the high variability of the chromosome shapes. The proposed algorithm can be utilized as a part of operations such as centromere detection, chromosome classification, DNA probe localization etc.

Future work includes a suitable method for splitting overlapping chromosomes along with more improvements for the end point correction method to make this algorithm applicable to any metaphase chromosome. Potential applications of this algorithm can extend to non-cytogenetic fields as well. For an example, this can be used in geographical locations of topological contours to determine optimal paths between two physical objects at the same altitude.

5. REFERENCES

[1] Jim Piper & Erik Granum, "On fully automatic feature measurement for banded chromosome classification," *Cytometry*, vol. 10, pp. 242–255, 1989.

[2] Kao et al, "Chromosome classification based on the band profile similarity along approximate medial axis," *The Journal of Pattern Recognition Society*, vol. 41, pp. 77–89, 2008.

[3] Graham et al., "Automatic karyotype analysis," *Chromosome Analysis Protocols*, vol. 29, pp. 141–185, 1994.

[4] Prabu Britto & G. Ravindran, "Novel findings in chromosome image segmentation using discrete cosine transform based gradient vector flow active contours," *Information Technology Journal*, vol. 6, no. 1, pp. 1–7, 2007.

[5] Jerry L. Prince Chenyang Xu, "Gradient vector flow: A new external force for snakes," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.

[6] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[7] Popescu et al, "Automatic karyotyping of metaphase cells with overlapping chromosomes," *Computers in Biology and Medicine*, vol. 29, no. 1, pp. 61–82(22), January 1999.

[8] Mehdi Moradi & S. Kamaledin Saterahdan, "New features for automatic classification of human chromosomes : A feasibility study," *Pattern Recognition Letters*, , no. 27, pp. 19–28, 2006.

[9] Bai et al, "Skeleton pruning by contour partitioning with discrete curve evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 29, no. 03, March 2007.

[10] Louisa Lam & Seong-Whan Lee, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 09, pp. 869 – 885, September 1992.

[11] Longin Jan Latecki & Rolf Lakamper, "Polygon evolution by vertex deletion," in *Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision*. 1999, pp. 398 – 409, Springer-Verlag London, UK.

[12] Longin Jan Latecki & Rolf Lakamper, "Convexity rule for shape decomposition based on discrete contour evolution," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 441 – 454, March 1999.