INTENSITY INTEGRATED LAPLACIAN ALGORITHM FOR HUMAN METAPHASE CHROMOSOME CENTROMERE DETECTION

Akila Subasinghe Arachchige[*]*, Jagath Samarabandu*[*]*, Peter K. Rogan*[†] *and Joan H.M. Knoll*[‡]

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

ABSTRACT

Centromere localization in human metaphase chromosomes is an essential task in many cytogenetic diagnosis procedures. The centromere location can be utilized to derive information such as the chromosome type, polarity assignment etc. Methods available in literature yield unreliable results mainly due to high variability of morphology in metaphase chromosomes and boundary noise in the image. In this paper we have proposed a multi-staged algorithm which utilizes both contour information as well as intensity information to obtain a more accurate centromere location. The width information along the axis of symmetry is obtained using a novel Laplacian based thickness measurement algorithm. The proposed method was observed to be more accurate compared to the state of the art when tested with 226 human metaphase chromosomes.

Index Terms— Centromere detection, chromosome analysis, Laplacian based thickness measurement

1. INTRODUCTION

The centromere is the most constricted region of a chromosome, to which the spindle fiber is attached during mitosis (cell division). The centromere location can be used for deriving information such as chromosome type, number and also in diagnostic processes such as chromosome dicentric assay. The width profile which can be defined as the sequential width measurements along the centerline or the axis of symmetry of the chromosome, is an important measurement used for deriving the centromere location. The high morphological variability in the way chromosomes sit on a microscope slide can introduce error into the measurement of the centromere locations for most existing methods. These methods can be divided into two categories based on the fundamental approach. Each of those categories can be described as follows.

The author would like to thank Wahab Khan and Heather Tarnowski for their help lent in collecting cell images used for this research.

1.1. Centerline based methods

These methods derive the width profile mainly by sampling the width along the chromosome centerline. A majority of existing methods fall under this category. Piper and Granum's approach towards this was to achieve the second moment of the profile of the chromosome along its centerline [1]. Moradi and Saterahdan took the average of image intensity along scan lines perpendicular to the centerline and used wavelet de-noising to remove sharp perturbation in the density profile (DP) [2]. They then classified the chromosomes with the use of a trained artificial neural network (ANN). Similarly, Wang et al. extracted the shape profile, density profile and the banding patterns using scan line sampling and then used a rule based setup to detect the chromosome centromere which claimed to have improved the reliability of the result [3]. The width profile in these methods is obtained after drawing scan lines or trellis structures which are perpendicular to the centerline of the chromosome. Since noise on the object boundary is represented in the centerline, the scan lines tend to miss the actual constriction at the centromere location. This can lead to high false positives in the centromere localization process. Furthermore, all these methods are prone to having spurious branches in the centerline of the chromosome. We previously proposed an algorithm for obtaining a more accurate and reliable centerline for the chromosome [4]. However this method also could yield false positives due to noisy centerline data points.

1.2. Other methods

Several algorithms have been published, that do not use the centerline for finding the constriction that marks the centromere. For example, Moradi et al. [5] took the horizontal and vertical projection vectors of the binary segmented chromosomes to detect the centromere location. This method did not perform satisfactorily for both acrocentric chromosomes as well as for any chromosome with a bend greater than 90 degrees.

Centerline based methods generally perform better than those which are not. Yet, the centerline based width profile is highly susceptible to noise in the centerline. Therefore, we

2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)
978-1-4673-1433-6/12/$31.00 ©2012 IEEE

propose a method which utilizes the centerline not as a base for width profile measurement but as a means to divide the chromosome into two symmetric partitions. This algorithm is also capable of utilizing intensity information present in the chromosome images through various staining methods for obtaining better results.

2. PROPOSED METHOD

The proposed method first performs the segmentation and then extracts the centerline of each chromosome. Next, an improved Laplacian based thickness measurement method was used for obtaining the width/intensity profile, which in turn is used for chromosome centromere localization. Each of the above stages will be discussed further in the subsequent chapters.

2.1. Segmentation and centerline extraction

Images were pre-processed using intensity normalization followed by median filtering to suppress noise while preserving edges.

The algorithm discussed in our previous publication [4] was used for obtaining the segmentation and the centerline of the chromosome. This algorithm performs the segmentation by first thresholding the image based on Otsu's method and then using the contour of the binary object as the starting point for Gradient Vector Flow (GVF) active contour model. The GVF is a segmentation model that could converge into concave boundaries, while having a high capture range [6]. This is highly useful when dealing with chromosomes which in general possesses highly variable morphologies.

The centerline is a shape descriptor based on the topological skeleton of the object, which produces a longitudinal axis of symmetry. The centerline was obtained using a skeleton pruning method based on Discrete Curve Evolution (DCE) [7]. DCE algorithm evolves polygon partitions by vertex deletion based on any given relevance measurement [8]. Subasinghe A. et al. [4] provide specifics regarding the utilization of the DCE based method for obtaining the centerline.

The binary segmentation of the chromosome is used solely to obtain the contour of the object while the centerline is merely used for partitioning the chromosome contour into approximately symmetrical halves near the telomere regions of the chromosome. By doing so, we can prevent the noise along the centerline from adversely affecting the centromere localization process.

2.2. Laplacian based thickness measurement & Centromere detection

The Laplacian operator (Δ) can be used to obtain the steady state of heat flow or voltage distribution between two heated or charged surfaces. By retaining the two longitudinal contour sides at two different potentials or temperatures, we can derive a set of equipotential lines in the static vector field created by the heat flow in steady state according to the Laplacian equation [9]. Then using simple incremental methods, thickness can be traced from one side to the other by traversing normal to these equipotential lines. The Euclidean length of all these small segments sum to the thickness at each cross-section of the object. This method gives a uniform sampling of the width profile better than techniques based on the centerline.

Yet, since this set up solely depends on the contour information for deriving the vector field, it can be susceptible to boundary noise in chromosome images. Most chromosome images contain some amount of intensity band information (depending on the staining technique) which can be used to assist the thickness measurement process. Therefore we propose a new algorithm by adding a flexible framework for incorporating intensity information into the standard Laplacian based thickness measurement process.

2.2.1. Intensity integration

Intensity information can be utilized to assist thickness measurements of textured objects in chromosome images. Banding information in chromosomes comes with many staining methods and is in general oriented normal to the object contour. Therefore, we have proposed to incorporate intensity information into static vector filed calculations using a local weighting scheme based on image intensity. The objective of this is to guide the Laplacian static field across the breadth of the object, based on neighboring pixel intensity values. By taking intensity information into consideration, the thickness measurement process can be adjusted to yield more accurate results. This inclusion minimizes the effects of boundary noise on the chromosome width profile and in turn on the centromere detection process.

The standard Laplacian static vector field based thickness calculation method guides a set of high potential contour points towards their unique closest set of points on the other contour [10]. The application of the standard Laplacian equation to a digital image can be easily performed by averaging the immediate neighborhood of a pixel and then subtracting that average from the pixel value.

The intensity information in the proposed method was simply used to bias the field towards the desired intensity pattern. This was achieved by using the weighting scheme described below.

Given the intensity image (I) which contains the object of interest, a total of 8 matrices (digital images) were created based on connectivity and directional intensity gradients (as a vector defined in (x, y) space) with identical dimensions to I as follows,

$$\nabla \vec{I}_{(i,j)} = abs[I(x,y) - I(x+i, y-j)] \quad (1)$$
$$(i,j) = \{i,j \in (-1,0,1), (i,j) \neq (0,0)\}$$

For simplicity and clarity, remaining steps will be described using the generic term $\nabla \vec{I}_{(i,j)}$. Next, all the matrices were normalized to the interval $(0\,,\,1)$, using the maximum absolute intensity difference in that direction. Then, the matrix values were inverted within the same range of $(0\,,\,1)$ by subtracting each matrix value from 1. The matrix $\nabla \vec{I}_{(i,j)}$ will now yield values close to unity where intensity level in the neighborhood is similar. Similarly this will also give smaller values (close to 0) for pixels with high intensity gradients. To address cases where intensity patches are parallel to the object contour, the proposed algorithm can be modified by simply removing the inverting step for all 8 matrices. By doing so, the weighting factors will bias towards higher intensity differences instead of homogenous regions.

The normalized intensity based weighting matrices were then re-scaled according to equation 2, where b is a scalar value between $(0\,,\,1)$ which will be referred to as the 'control variable' henceforth. Therefore the values in the weighting matrix $\nabla \vec{I}_{(i,j)}$ will vary in the interval of $(b\,,\,1)$. Empirically the control variable b was set to 0.9 for all our experiments.

$$\nabla \vec{I}_{(i,j)} = \nabla \vec{I}_{(i,j)} * (1-b) + b \quad (2)$$

The purpose of the control variable b is to scale or control the influence of the intensity variation on the proposed Laplacian calculation. A lower value for b will increase the influence of the intensity information and vice versa. Therefore, a value of 1 for the control variable will calculate the standard Laplacian vector field with no influence from the intensity values. This value has to be set based on how prominent and consistent the intensity patterns are in a given image. Practical range of values would lie between the limited range of $(0.7\,,\,1)$ for this experiment.

Once these sets of intensity weighting factor matrices are calculated, those values can be directly used to change the Laplacian static field calculated at each iteration. Therefore instead of the standard Laplacian kernel, we propose to use the intensity integrated kernel, which is now defined for each $(x\,,\,y)$ coordinate location in the image. Now we have a static vector field generation process that includes both non-uniform and local shape features depending on the intensity variation in the region and the control variable b which controls the amount of biasing. When using the standard Laplacian kernel, every pixel influences the 8 connected neighbors uniformly. Whereas, in the proposed method, each pixel affects the neighboring pixels based on the degree of similarity or differences in the respective intensity values. Its also important to realize that these weight matrices are static in nature and do not change with each iteration. Therefore, the proposed algorithm is comparable with the standard Laplacian calculations from in computational cost.

Next, the gradients at each pixel location (Φ) were calculated along the two major axes (x and y) using neighborhood pixel values as given below,

$$\frac{\Phi(x,y)}{\Delta x} = \frac{(B(x+\Delta x, y) - B(x-\Delta x, y))}{2}$$
$$\frac{\Phi(x,y)}{\Delta y} = \frac{(B(x, y+\Delta y) - B(x, y-\Delta y))}{2} \quad (3)$$

Then each of these gradient components were normalized and stored in matrices N_x and N_y using the magnitude of the vector at each pixel. The matrices N_x and N_y contain the intensity biased Laplacian static field vector components for x and y axis directions.

Once the proposed intensity integrated Laplacian static field is derived, the corresponding contour points and the distance between them were calculated using Euler's method to calculate the thickness corresponding for all contour points. In order to avoid incorporating the telomere region width measurements, the profile was pruned on either side by 10% (selected empirically) of the total number of points on the contour segment. This reduce the chance of detecting a telomere of a chromosome as a centromere location.

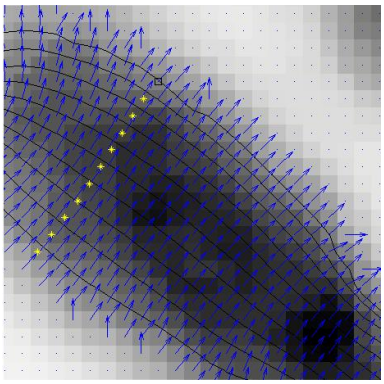

Fig. 1. The steps of tracing the thickness (yellow stars) at one contour location of the chromosome. The final thickness value is calculated by getting the sum of all the lengths of these small steps.

The collection of these thickness values constitutes the width profile of the chromosome. Figure 1 depicts the steps of tracing the thickness at one contour location of the chromosome. The centromere location is detected by obtaining the global minimum of this width profile.

3. RESULTS

We have tested the accuracy of the proposed method in detecting centromere locations in 226 human metaphase chromosomes against that of a state of the art method [11]. These

chromosomes, with no overlaps or touches of neighboring chromosomes, were selected from 12 lymphocyte metaphase cells. The centromere locations were detected using the same methodology after obtaining the width profile measurements using the two different methodologies. The centromere location manually recorded by an expert was used as the 'gold standard' in the analysis. Since the chromosome centromere is a region as opposed to a single location, the expert was instructed to draw a line across the centromere region. Then, the perpendicular distance from the pixel location given by the algorithm to the user drawn line segment is calculated as the error of detection. By doing so, any displacement of the detected centromere location along the drawn centromere line would be trivial. These error values will be denoted by E_L and E_C for the error of the Laplacian based proposed method centromere and the state of the art centerline method result respectively and will be referred here after. Table 1 provides a summary of the error metric values (in units of pixels) obtained for the data set.

Table 1. Descriptive values for the detection error data set when analyzed with proposed Laplacian based method (E_L) and the state of the art Centerline based method (E_C).

Descriptive Statistics

	N	Mean		Kurt--osis	Skew--ness
		Stat.	Std. Error		
E_L	226	4.0243	.4535	17.859	3.839
E_C	226	8.7819	.7749	2.657	1.834

The proposed algorithm yields a smaller error mean value while maintaining a smaller standard error of mean. This behavior can be further supported by the skewness and kurtosis values obtained for the proposed method as opposed to the centerline method. The higher kurtosis value suggests a tight clustering of error values around the peak while the higher skewness depicts an asymmetric distribution biased towards lower error value. The direction of the bias was observed through a simple histogram plot of error values.

4. CONCLUSION AND FUTURE WORK

We have presented a novel intensity integrated Laplacian based method for detecting centromere locations in human metaphase chromosomes more accurately. The results compared to a state of the art centerline based method have demonstrated encouraging results and warrant further investigation.

We will attempt to generalize the algorithm in future and make it applicable to other similar measurement problems. Thorough schemes of statistical analyzes are required to validate the accuracy improvements of the proposed method as well as to explore inter and intra observer variability in manually detecting centromere locations. In addition, we plan to develop the algorithm further to detect dicentric chromosomes.

5. REFERENCES

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

[2] M. Moradi and S. K. Saterahdan, "New features for automatic classification of human chromosomes : A feasibility study," *Pattern Recognition Letters*, , no. 27, pp. 19–28, 2006.

[3] X. Wang et al, "A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes," *Computer Methods and Programs in Bio Medicine*, vol. 89, pp. 33–42, 2008.

[4] A. Subasinghe A. et al., "An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes," in *International Conference on Image Processing (ICIP)*, September 2010.

[5] M. Moradi et al., "Automatic locating the centromere on human chromosome pictures," in *16th IEEE Symposium on Computer-Based Medical Systems*, 2003.

[6] Chenyang Xu and Jerry L Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transaction on Image Processing*, vol. 7, no. 3, 1998.

[7] X. 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.

[8] L. J. Latecki and R. Lakämper, "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.

[9] Haidar H. & Soul J.S., "Measurement of cortical thickness in 3d brain mri data:validation of the laplacian method.," *NeuroImage*, vol. 16, pp. 146 – 153, 2006.

[10] J. Stephen E. et al., "Three-dimensional mapping of cortical thickness using laplaces equation," *Human Brain Mapping*, vol. 11, pp. 12–32, 2000.

[11] A. Subasinghe A. et al., "An accurate image processing algorithm for detecting fish probe locations relative to chromosome landmarks on dapi stained metaphase chromosome images," in *Seventh Canadian Conference on Computer and Robot Vision (CRV)*, May 2010.