Intensity Integrated Laplacian-Based Thickness Measurement for Detecting Human Metaphase Chromosome Centromere Location

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

***Abstract*—Accurate detection of the human metaphase chromosome centromere is an important step in many chromosome analysis and medical diagnosis algorithms. The centromere location can be utilized to derive information such as the chromosome type, polarity assignment, etc. Methods available in the literature yield unreliable results mainly due to high variability of morphology in metaphase chromosomes and boundary noise present in the image. In this paper, we have proposed a multistaged algorithm which includes the use of discrete curve evolution, gradient vector flow active contours, functional approximation of curve segments, and support vector machine classification. The standard Laplacian thickness measurement algorithm was enhanced to incorporate both contour information as well as intensity information to obtain a more accurate centromere location. In addition to segmentation and width profile measurement, the proposed algorithm can also correct for sister chromatid separation in cell images. The proposed method was observed to be more accurate and statistically significant as compared to a centerline-based method when tested with 226 human metaphase chromosomes.**

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

I. Introduction

THE centromere can be used in identifying the chromosome type, number, and also in diagnostic processes such as the chromosome dicentric assay (see Fig. 1). The centromere is characterized by a constriction of the width of the chromosome. Detecting the human metaphase centromere location presents unique challenges.

The morphology and lengths of chromosomes can vary significantly between different growth conditions and cytogenetic preparations. Detection of subtle structural chromosome abnormalities at high resolution requires methods (such as addition of DNA intercalating agents, reduced exposure to colcemid, cell cycle synchronization, 3–4 day lymphocyte culture) that reduce chromosome condensation or arrest chromosomes at prometaphase. Numerical chromosome abnormalities or low-frequency large structural acquired abnormalities (e.g., dicentric chromosomes) present in cancer or biodosimetry samples can be processed with methods (such as prolonged colcemid time and concentration; two days cell culture) that increase the number of cells in metaphase but result in shorter chromosome lengths. Shorter, more condensed chromosomes often have separated or distinct sister chromatids on each arm and fewer chromosomal bands (see Fig. 2).

The width constriction on higher banded chromosomes can be missed easily due to bends or noise on the chromosome boundary, while chromosomes with sister chromatid separation tend to mislead the width profile calculation near the telomeric region. This constriction can be identified using the chromosome width profile which can be defined as the sequential width measurements along the centerline or the axis of symmetry of the chromosome.

From an image analysis point of view, metaphase chromosomes pose numerous challenges. Morphological variability caused by nonrigid chromosome structures is one prominent factor. Furthermore, the stage of mitosis (cell division cycle) at which the cells were arrested and the microtubule

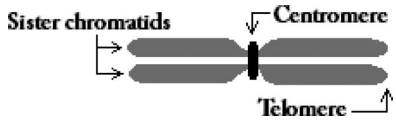

Fig. 1. Sample graphical representation of a human metaphase chromosome with key components and regions labeled for reference.

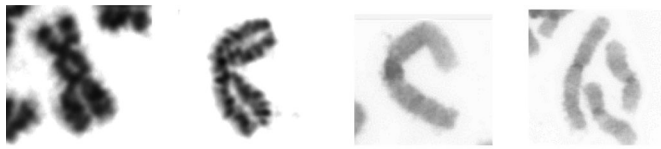

Fig. 2. Depiction of various degrees of sister chromatid separation present in some Geimsa stained chromosome cell images from a radiation biodosimetry laboratory (a and b) as well as some lengthy chromosomes characteristic to ones prepared at a clinical cytogenetic laboratory (c and d) which are prepared with small, but significant differences in the laboratory cell preparation procedures.

Manuscript received June 18, 2012; revised November 21, 2012 and January 29, 2013; accepted February 2, 2013. Date of publication February 20, 2013; date of current version June 24, 2013. This work was supported in part by the Western Innovation Fund (University of Western Ontario), in part by the Natural Sciences and Engineering Research Council of Canada, and in part by the DART-DOSE CMCR (5U01AI091173-02 from the U.S. Public Health Service). *Asterisk indicates corresponding author*.

*A. S. Arachchige is with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: akilamike@gmail.com).

J. Samarabandu is with the Department of Electrical and Computer Engineering, Western University, London, ON N6A 5B9, Canada (e-mail: jagath@uwo.ca).

J. H. M. Knoll and P. K. Rogan are with the Departments of Pathology and Biochemistry, Schulich School of Medicine & Dentistry, Western University, London, ON N6A 5B9, Canada (e-mail: Joan.Knoll@schulich.uwo.ca; progan@uwo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2248008

polymerization inhibitors significantly affect chromosome appearance. These factors, which can dictate the presence and the degree of sister chromatid separation as well as the length and width of chromosomes in that cell, are usually standardized within but not necessarily between laboratories. Therefore, effective chromosome image processing techniques must accurately account for these sources of variability in detecting cardinal features, such as centromeres, in order to distinguish different chromosomes from one another and from abnormalities. Fig. 2 depicts a sample of chromosomes with high shape variability.

Published methods for centromere identification can be inconsistent due to variable morphologies and boundary noise present in images of metaphase chromosomes. Chromosome segmentation methods for centromere identification are fundamentally distinguished as either centerline-based methods or methods based on other features.

A. Centerline-Based Methods

Many published algorithms attempt to detect the centromere location by detecting the constriction along the centerline of the chromosome. Although Piper and Granum [1] approached this by taking the second moment along the centerline, a common approach is to calculate feature profiles along scan lines perpendicular to the centerline [2]. In a similar approach, Wang *et al*. used these scan lines or trellis structures which are perpendicular to the centerline of the chromosome to extract the shape profile, the width profile (collection of width measurements), and the banding patterns of chromosomes [3]. All these methods are prone to having spurious branches in the centerline. We previously proposed an algorithm to overcome this problem and yield a reliable centerline [4]. Yet this method could also give false positives as noise on the centerline (introduced by a noisy object boundary) can result in the scan lines missing the actual constriction at the centromere location. Mohammad [5] used the centerline (using our previous approach [6]) to detect the centromere locations which were derived using the degree of concavity of the object boundary which is a scale variant feature. However, this method also suffers from boundary noise in chromosome cell images which are reflected both on the centerline as well as on the concavity measuring algorithm. Furthermore, the presence of high degree of sister chromatid separation can also introduce error into the centerline detection as well as to the centromere localization.

B. Methods Based on Other Features

Some of the methods in the literature do not use the centerline for localizing the centromere location. Mousavi and Ward assigned a membership value for each pixel of DAPI (4′,6-diamidino-2-phenylindole) and FITC (fluorescein isothiocyanate) images (with centromere probes) based on an iterative fuzzy algorithm [7]. Yet, this method has limited scope of application as it depends on special specimen preparation and information in the form of centromere probes and FITC images. Moradi *et al*. [8] and Faria *et al*. [9] took the horizontal and vertical projection vectors of the binary segmented chromosomes. These projection vectors were obtained by summing up the number of object pixels in the binary segmented image in each horizontal and vertical directions and the centromere was located by finding the global minimum in these vectors. Chromosomes with a bend greater than 90° as well as acrocentric chromosomes could not be handled accurately using this method.

Despite the fact that centerline-based methods perform better than their counterpart, these approaches are still susceptible to errors introduced by noisy object boundaries. Therefore, we have proposed to utilize the centerline of a chromosome not as a means of deriving the scan lines (trellis structure), but for merely dividing the chromosome into two symmetric partitions. This proposed algorithm can automatically detect sister chromatid separation near the telomere region of the chromosome and then correct for that artifact. The intensity information present in the chromosome images was utilized in order to obtain improved results for chromosomes with various staining methods. This provides a basic framework for incorporating additional feature into the standard Laplacian-based thickness measurement algorithm. This paper extends our previous work [10] and provides extensive statistical analysis to examine the performance of the proposed method on chromosomes obtained through different staining methods.

II. Method

The following section will describe the proposed algorithm for calculating the width profile of human metaphase chromosomes. This algorithm can be functionally categorized into four tasks as follows:

1) segmentation and contour extraction;
2) centerline extraction;
3) telomere shape analysis and correction for sister chromatid separation; and
4) thickness measurement and centromere detection.

Chromosomes are first subjected to global thresholding and gradient vector flow (GVF) [11] active contours for obtaining an accurate outline of the chromosome boundary. This information is utilized for width profile acquisition, and therefore, the accurate representation of the chromosome outline is important. Then the centerline of the chromosome is extracted using a discrete curve evolution (DCE)-based [12] skeleton pruning algorithm. The proposed algorithm has the ability to detect sister chromatid separation by performing chromosome telomere shape analysis through functional approximation. This information along with the centerline is effectively used as a method of dissecting the chromosome into two approximately symmetric contour segments. Next, the width profile of the chromosome is calculated by creating a Laplacian vector field between the two contour segments. We have modified the Laplacian vector field further to include intensity information in addition to the contour segments. By doing so, the vector field can be guided using intensity localities within the chromosome. Therefore, the integration of intensity aids in reducing the effects of boundary noise on the width profile.

Each of the aforementioned steps will be discussed further in the subsequent sections.

A. Segmentation and Contour Extraction

The image preprocessing stage of the algorithm consisted of intensity normalization followed by median filtering to suppress noise while preserving edges.

The algorithm used for obtaining the segmentation and the centerline of the chromosome (based on our previous approach [4]) first obtains a binary output through global thresholding using the Otsus method. A major issue in using a thresholding technique such as this is the need to carefully select the threshold parameter to account for intensity differences among images. To overcome this, we use the GVF active contour model [11] using the outline from the thresholded image as the starting point. This relieves us from having to adjust the threshold parameter for each image and yields a smoother contour [4].

The main internal parameters of the GVF were set at $\alpha = 0.05$ (elasticity factor), $\beta = 0$ (rigidity factor), $\mu = 0.2$ (GVF regularization factor), and $\kappa = 2$ (external force weight). This set of values were obtained empirically and yielded satisfactory segmentation results across the entire dataset.

B. Centerline Extraction

The centerline is a shape descriptor based on the topological skeleton of the object, which produces a longitudinal axis of symmetry. The chromosome centerline is necessary in many operations like classification performed on segmented chromosomes [1], [13]. Many shape and structure-related features such as 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 [14]. In the literature, medial axis transformation (MAT) has been commonly used 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. Hassouna and Farag have proposed a method for obtaining a robust skeleton for 3-D objects using a novel skeletonization algorithm which seems to include built-in pruning abilities [15] and clearly warrants future investigation. However, in our approach, we have adopted a separate skeleton pruning method based on DCE [12]. This algorithm was selected to obtain reliable centerline with prometaphase chromosomes as well as chromosomes with sister chromatid separation. The DCE algorithm evolves polygon partitions by vertex deletion based on any given relevance measurement [16]. 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 then evolves the polygon iteratively by removing the vertex which had the least value for the relevance value $K(v, u, w)$ defined in (1), where d_{uv} and d_{vw} are the Euclidean lengths 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. Since DCE is simply deleting vertices of the polygon partitions, the topology information is guaranteed to be preserved

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

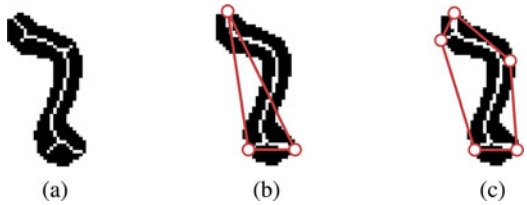

Fig. 3. Comparison between standard skeleton and DCE-based solutions on a bent chromosome. (a) Skeleton. (b) DCE triangle. (c) DCE pentagon.

Fig. 3 depicts the reliability and accuracy of the DCE-based pruning method compared to standard pruning. Fig. 3(b) and (c) depicts two DCE-based pruning results for different number of vertices for the end convex polygon.

In the case of obtaining the medial axis of a chromosome, the ideal result would be a pruned skeleton with no extra branches (two vertices). Yet, as the minimum convex polygon being a triangle and DCE being modeled as polygons, we have to set the DCE to terminate with three vertices. Therefore, the resulting skeleton will at minimum have one spurious branch. This issue was resolved by tracing all branches and pruning off the shortest branch completely. The DCE result was then pruned by 10% of the total skeletal length at each end in order to avoid effects of skeletal bifurcation near telomere region. Next, the centerline was sampled with a 7 pixel interval. The centerline ends were recreated by extending the ends of the pruned centerline based on the orientation of the extreme sample point pairs.

C. Correcting for Sister Chromatid Separation

Sister chromatid separation during cell preparation can adversely affect the centromere detection process. The separation can cause the centerline of the chromosome to traverse into one of the sister chromatids and, therefore, yield false minima on the width profile. In a preliminary study performed on detecting 66 centromere locations in Geimsa stained chromosomes (with chromatid separation), 11 out of 13 false detections were due to sister chromatid separation. In order to correct for this, the sister chromatid separation has to be detected for each chromosome. We propose an automated contour partitioning and shape analysis process for this task. The proposed algorithm will automatically partition and sample the telomere regions and label accordingly to reflect presence of sister chromatid separation.

1) Contour Partitioning: One of the clear visual cues for detecting telomere region is the high curvature of the object contour in the vicinity of the telomere region. In one attempt, Xu and Kuipers proposed a method in which the contour was divided at maximum-curvature points after tracing for curvature and position continuity [17]. High curvature points in general tend to mark salient points along a contour as a distinct visual cue. Yet, high curvature points can also be introduced by boundary

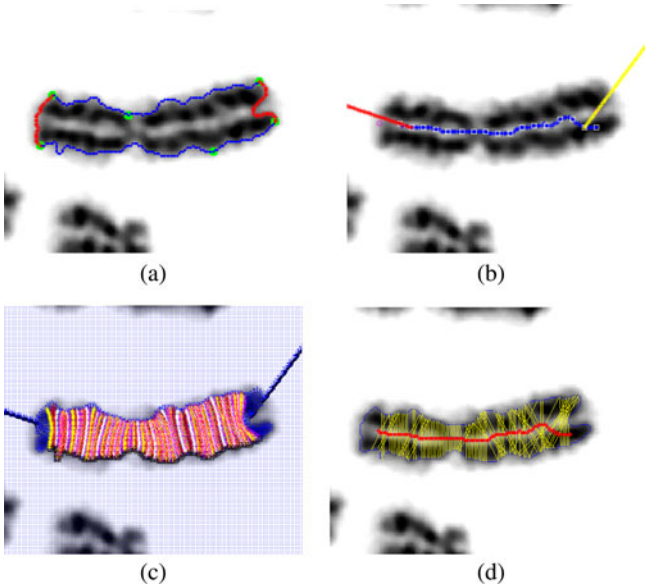

Fig. 4. (a) Demonstration of an example of the telomere segment partitioning where the red segments are the extracted telomere segments, while the yellow squares are the six original DCE points. (b) Instance where correction for sister chromatid separation is applied to the right-hand side telomere region (yellow line) while retaining the original line segment for the other telomere red line. The sample points along the centerline are depicted by the blue squares. (c) and (d) Depiction of the width profile sampling of the proposed method and our previous approach [4], respectively.

noise in the segmentation result. The addition of unnecessary contour segments can complicate the detection process. We have identified the following as three important conditions in the partitioning algorithm:

1) The partitioning has to include key visual cues or features of the contour completely within each segment.
2) Partitioning results must be reproducible and less sensitive to noise present in the contour.
3) The partitioning algorithm must be effective for all morphologies of the object.

The DCE process used earlier for obtaining the centerline of the chromosome satisfies all of the aforementioned conditions. Therefore, the same algorithm was used for obtaining the salient points related to the telomere region. Due to high morphological changes in chromosomes, the highest four relevance points [see (1)] may not capture (include) all telomere end points. Therefore, the stopping criteria for the DCE process were altered to yield six salient locations (as opposed to the three vertices polygon used in Section II-B) along the object contour to reduce the probability of missing true positives (see Fig. 4). Next, the true telomere end points (four points) are selected from these six high relevance DCE points. The two centerline end points are a subset of this set of candidate points. Therefore, telomere point detection can be further simplified into a problem of selecting two points out of four candidates, with respect to two known telomere points. The selection criteria of the shortest distance along object contour were observed to yield satisfactory results. Once these four points are selected, the telomere regions can be easily extracted along the object boundary.

2) Shape Information Extraction: The partitioned telomere curve segments carry information relating to the presence of sister chromatid separation. Traditionally, this is achieved by incorporating high-level shape information using a set of well-defined features. These features should represent the shape characteristics of a sister chromatid separated telomere curve segment. Yet, defining such shape features can be undesirable due to following reasons.

1) The classification results will be highly dependent and limited by the set of features selected.
2) High variability of contour segment due to the amount of chromatid separation and intensity fading in between those sister chromatids would yield a large set of possibilities for the telomere curve segments. This will increase the difficulty in finding a suitable set of features.
3) Defining a significant number of features and then trying to optimize a set of features using any technique such as PCA is an added computational step. The proposed functional approximation method includes this step inherently.

Therefore, we have used functional approximations for each of the curve partitions and used their orthogonal basis coefficients as an alternative to various geometrical features. These coefficients directly represent the shape information of the telomere contour segment in each coordinate axis. The concept of using orthogonal coefficients for matching shapes has been used in the field of hand writing recognition with satisfactory results [18], [19]. Therefore, we have adopted a similar approach for the telomere contour shape matching problem based on deriving coefficients for Legendre polynomials.

In practice, it is not possible to calculate all (infinite number) coefficients for each coordinate axis. Therefore, empirically, we have selected to calculate coefficients up to the order 10. The lower order coefficients tend to give low-frequency information while the higher order coefficients yield high-frequency wiggles. Therefore, for each segment, we calculate 20 coefficients (10 coefficients for each coordinate axis). Next, a support vector machine (SVM) was trained using 90 labeled set of telomere coordinate curve segments [20]. With twofold cross validation, the SVM achieved higher than 92% accuracy of classification. When applied to the class of chromosome images with sister chromatid separation, the extension of the sampled points was altered to satisfy the coordinates of the telomere midpoint. This correction is not meant to correct the centerline to reflect symmetry along the object longitudinal axis. Since the centerline is not directly used for getting the width profile, it is sufficient to simply partition the contour of the chromosome symmetrically.

D. Laplacian-Based Thickness Measurement

Laplacian-based thickness measurement is an algorithm used for cortical thickness measurements in a number of brain mapping applications [21], [22]. The Laplacian operator (Δ) yields the divergence of the gradient of a function in the Euclidean point space. This is written as follows, where $\bigtriangledown$ is the first derivative or the differential operator in any given direction

$$\Delta f = \mathrm{div}(\bigtriangledown f) = \bigtriangledown . \bigtriangledown f. \quad (2)$$

This operator is used to obtain the steady state of heat flow or voltage distribution between two heated/charged contour segments in these applications. The vector field created at steady state can be easily used to trace the thickness along the object contour. This method gives a uniform sampling of the width profile better than techniques based on the centerline. Yet, due to the sole dependence on the contour information, the Laplacian-based method can still be susceptible to contour noise embedded during the segmentation stage. With different staining methods and imaging conditions, the object boundary noise content may vary significantly and in return will change the effectiveness of this algorithm.

Since banding information present in chromosome texture is visible with many staining methods and is in general oriented normal to the object contour, it can be utilized for assisting the thickness measurement process. Therefore, we have proposed a novel algorithm to overcome the aforementioned limitation by incorporating intensity information into static vector filed calculation. This algorithm modifies the standard Laplacian-based method by 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. The addition of this intensity information can directly influence the accuracy of the thickness measurement process and can be adjust for obtaining accurate centromere localization. The intensity bias minimizes the effects of boundary noise on the width profile.

The proposed thickness for measuring algorithm requires the following information as inputs to the system:

1) The single-pixel wide contour of the segmented object of interest.
2) A separation of the object contour using the longitudinal axis of symmetry of the object. Correcting this for sister chromatid separation was performed through the shape analysis process.

The Laplacian static vector field-based thickness calculation method guides a set of high-potential contour points toward their unique closest set of points on the other contour [23]. The following description of intensity integration is reproduced from our previous work [10] for better readability.

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

Given the intensity image (I) which contains the object of interest, a total of eight matrices (digital images) were created based on connectivity and directional intensity gradients with identical dimensions to I as follows:

$$\nabla\vec{I}_{(i,j)} = \text{abs}[I(x,y) - I(x+i, y-j)]$$
$$(i,j) = \{i, j \in (-1, 0, 1), (i,j) \neq (0,0)\}. \quad (3)$$

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 [refer (4)]. 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 is modified by simply removing the inverting step for all eight matrices. By doing so, the weighting factors will bias toward higher intensity differences instead of homogenous regions

TABLE I
KERNEL THAT INTEGRATES INTENSITY INFORMATION INTO THE LAPLACIAN CALCULATION FOR LOCATION (x, y)

$-\frac{\nabla\vec{I}_{(-1,1)}(x,y)}{8}$	$-\frac{\nabla\vec{I}_{(0,1)}(x,y)}{8}$	$-\frac{\nabla\vec{I}_{(1,1)}(x,y)}{8}$
$-\frac{\nabla\vec{I}_{(-1,0)}(x,y)}{8}$	$+1$	$-\frac{\nabla\vec{I}_{(1,0)}(x,y)}{8}$
$-\frac{\nabla\vec{I}_{(-1,-1)}(x,y)}{8}$	$-\frac{\nabla\vec{I}_{(0,-1)}(x,y)}{8}$	$-\frac{\nabla\vec{I}_{(1,-1)}(x,y)}{8}$

$$\nabla\vec{I}_{(i,j)} = \frac{\nabla\vec{I}_{(i,j)}}{\max(\nabla\vec{I}_{(i,j)})}. \quad (4)$$

The intensity-based weighting matrices were then rescaled according to (5), 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 $(b,\ 1)$

$$\nabla\vec{I}_{(i,j)} = \nabla\vec{I}_{(i,j)} * (1-b) + b. \quad (5)$$

The purpose of the control variable b is to scale or control the influence from the intensity variation onto the standard 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. The practical range of values would lie between the limited range of $(0.7,\ 1)$ for this experiment. Empirically, the control variable b was set to 0.9 for all our experiments.

Once these sets of intensity weighting factor matrices are calculated, those values can be directly used to change the way the Laplacian static field is calculated at each iteration. Therefore, instead of the standard Laplacian kernel, we propose to use the kernel given by Table I, which is now defined for each $(x,\ y)$ coordinate location in the image. Therefore, now we have a static vector field generation process that includes both nonuniform and local shape features depending on the intensity variation in the region and the control variable b which controls the amount of biasing. This approach provides the ability for each pixel to affect the neighboring pixels based on the intensity similarity or difference between them. It is also important to realize that these weight matrices are static in nature and do not change with

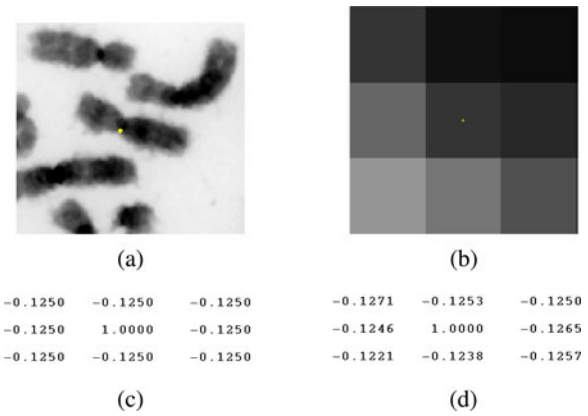

Fig. 5. (a) Demonstration of the difference between the kernel of the proposed method and the standard Laplacian kernel. (b) Enlarged view of the 3×3 neighborhood of the pixel location marked by yellow in (a). (c) and (d) Representation of one the standard Laplacian and intensity biased Laplacian kernels calculated for the neighborhood of interest.

each iteration. Therefore, the proposed algorithm is comparable with the standard Laplacian calculations in computational cost.

Fig. 5 depicts the difference between the standard Laplacian kernel and (one instance of) the proposed intensity-based Laplacian kernel. The instance of the proposed kernel [see Fig. 5(d)] clearly depicts the biasing of the Laplacian field in directions where similar intensities are present.

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 where $B(x, y)$ is the steady-state Laplacian image

$$\frac{\Phi(x,y)}{\Delta x} = \frac{(B(x+\Delta x, y) - B(x-\Delta x, y))}{2}$$
$$\frac{\Phi(x,y)}{\Delta y} = \frac{(B(x, y+\Delta y) - B(x, y-\Delta y))}{2}. \quad (6)$$

Then, each of these gradient components was 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 the x- and y-axis directions.

Once the proposed intensity integrated Laplacian static field is derived, the corresponding contour points and the distance between them have to be calculated. The same thickness measures can be obtained by using starting points from either contour segments or even the centerline points of the chromosome. Euler's method was used for the aforementioned task. This is a simple and yet effective way of traversing through a vector field as given by (7), based on the local vector field direction and magnitude. For implementation, the direction of traverse has to be adjusted (by flipping the polarity of the vector field when necessary) in order to make sure that the thickness is measured within the chromosome body

$$x_n = x + \Delta x$$
$$y_n = y + y'\Delta x. \quad (7)$$

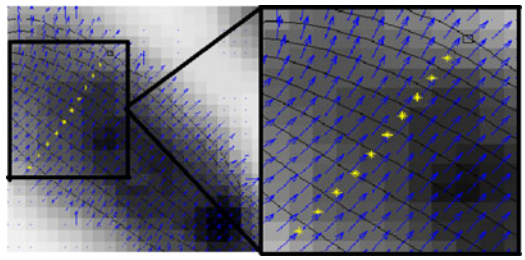

Fig. 6. Steps of tracing the thickness (yellow stars) at one contour location of the chromosome where the arrows indicate the Laplacian vector field. The black square indicates the starting point on the contour of the object. The final thickness value is calculated by getting the sum of all the lengths of these small steps.

First, the gradient values at each pixel location (Ψ) is calculated using each tangent vector component (N_x and N_y) as given by

$$\Psi(x,y) = y'(x,y) = N_y(x,y)/N_x(x,y). \quad (8)$$

Computation of the thickness involves the following iterative process using Euler's method for each high-potential contour point:

1) Based on the local direction of the vector field gradient ($\Psi(x,y)$), select the direction (axis) for incrementing.
2) Apply Euler's equation and calculate the new pixel location along the direction of the vector field.
3) Calculate the Euclidean distance between the new and current location and accumulate with the current total distance.
4) If the new location is within the object of interest, start from the first step onward. Once the calculated location placed outside the object, the algorithm will move to the next contour point.

The collection of these accumulated Euclidean length values is considered as the thickness/width profile of that object. Fig. 6 depicts the steps of tracing the thickness at one contour location of the chromosome. 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 reduces the chance of detecting a telomere of a chromosome as a centromere location.

III. Results

The centromere and centerline detection method described by Arachchige *et al*. [4], [6] was used for comparing the results of the proposed method. The centerline-based method was selected since it can handle any chromosome morphology without yielding spurious branches in the centerline and also attempts to find the width profile similar in principle to most existing methods. The centromere location manually recorded by an expert was used as the "gold standard" in the analysis. The analysis of intraobserver variability of ground truth was not attempted at this time due to limitations in resources. 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 in pixels from the centromere

TABLE II
BREAKDOWN OF CHROMOSOME CELL IMAGES AND CHROMOSOMES BASED ON THE STAINING METHOD AND THE SISTER CHROMATID SEPARATION (SC SEP.)

Abbr.	Label	Images	Chromosomes
D-NSC	DAPI-No SC Sep.	4	72
D-WSC	DAPI-With SC Sep.	3	59
G-WSC	Geimsa-With SC Sep.	5	95
	Total	12	226

location given by the algorithm (midpoint of the scan line with the minimum width) to the user drawn line segment is denoted as the error of detection. In this study, we are more interested in errors in the vicinity of the center of the chromosome rather the orientation of the scan line. Therefore, the experiment was set up in a way that any displacement of the detected centromere location along the drawn ground truth centromere line would not influence the accuracy of detection. Furthermore, we have not normalized pixel error values since the centromere structure mostly remains fixed despite chromosome morphology or chromosome number. These error values will be denoted by E_L and E_C for the error of the Laplacian-based proposed method centromere and the centerline method result, respectively, and will be referred here after. The following section will provide an in-depth analysis of the preliminary results provided in our previous work [10].

A total of 226 human lymphocyte chromosomes from 12 chromosome cells were randomly selected for the study given that they have no overlaps or touches with neighboring chromosomes. Table II provides the breakdown of these cell images based on the staining method as well as the presence of sister chromatid separation (judged visually).

The algorithm performs well on chromosomes regardless of the staining method and the shape of the chromosome. Yet, the algorithm can fail in the presence of high sister chromatid separation in the binary segmentation of the chromosome. The DCE algorithm in the contour partitioning algorithm can select a high curvature point within the telomeric region and, thus, yield the correction for sister chromatid separation ineffective. Fig. 7 shows some of the sample results for multiple staining methods used commonly in cytogenetic studies and analysis. Fig. 7(f) depicts an instance where the correction for sister chromatid separation has failed to yield the expected result. In Fig. 7(e), we have presented a case where the centerline-based method has failed to yield an accurate centromere location. This is caused by a noisy centerline which misses the actual width constriction at the centromere location. But the proposed method yields better results due to uniform sampling despite object boundary noise [see Fig. 4(d)].

A. Statistical Analysis

A preliminary analysis of the two data distributions was performed using the summary of the error metric values (in units of pixels) obtained for the dataset (see Table III) [10]. The proposed algorithm yields a smaller error mean value while maintaining a smaller standard error of mean. This observation is further supported by the skewness and kurtosis values obtained for the proposed method as opposed to the centerline method.

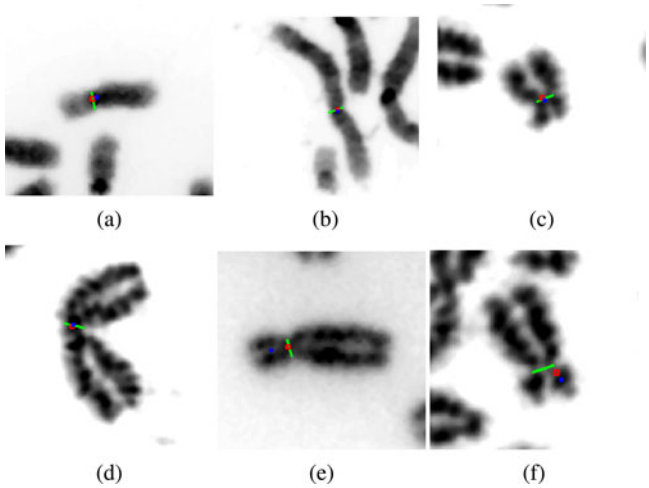

Fig. 7. (a) Demonstration of some sample results of the algorithm where the detected centromere location is depicted in a red color circle against our previous approach [4] in a blue color star, while the expert-drawn centromere line is depicted in white. (a) and (b) Results of DAPI stained chromosomes; (c)–(f) results of Geimsa stained chromosomes. (e) Instance in which the proposed algorithm has outperformed the state of the art significantly; (f) instance in which the proposed algorithm has failed to yield the accurate centromere location due to high degree of sister chromatid separation.

TABLE III
DESCRIPTIVE VALUES FOR THE DETECTION ERROR DATASET WHEN ANALYZED WITH THE PROPOSED LAPLACIAN-BASED METHOD (E_L) AND CENTERLINE-BASED METHOD (E_C) [4] (REPRODUCED FROM THE PREVIOUS WORK [10])

	N	Mean		Kurt-sis	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 higher kurtosis value suggests a tight clustering of error values around the peak while the higher skewness depicts an asymmetric distribution biased toward lower error value.

Since the same set of chromosomes were used to analyze the results of both algorithms, this experiment falls under the category of repeated measurement analysis. Therefore, a "*t* statistic" cannot be utilized to obtain the significance of the results. Therefore, the data distribution was then checked for normality within each category using the Kolmogorov–Smirnova normality test. The test proved that we can reject the null hypothesis ($p < 0.05$) that the distributions are normal (test statistics for $E_L = 0.28$ and $E_C = 0.25$). This deviation from normality is further supported by the high skewness and kurtosis values (see Table III). This provides evidence toward the conclusion that the proposed method yields a better grouped distribution toward a lower mean error value when compared to the centerline method.

The significance of these finding was explored using the "Wilcoxon signed rank test" which can be directly applied to repeated measurements without the normality constraint. The results of this test are given in Table IV.

The data in Table IV analyze cases based on their signs after comparing each corresponding pair. Therefore, it can be observed that the sum of positive ranks is significantly higher than that for negative ranks. This corresponds to the cases in which

TABLE IV
WILCOXON SIGNED TEST RANK ANALYSIS RESULTS

		N	Mean Rank	Sum of Ranks
E_C - E_L	(-) Ranks	80	89.74	7179.00
	(+) Ranks	146	126.52	18472.00
	Ties	0		
	Total	226		

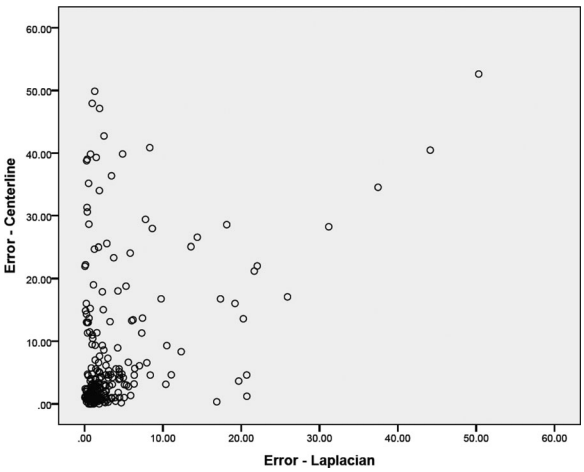

Fig. 8. Scatter plots for demonstrating the correlation between the two detection error distributions in which the "x" axis is the detection error of the proposed Laplacian-based method (E_L) and the "y" axis is the centerline-based method (E_C).

the Laplacian-based proposed method yields lower error values compared to our previous approach.

Based on this analysis, we conclude that the proposed Laplacian-based method elicits statistically significant improvement in centromere localization ($Z = -5.738, p < 0.05$) compared to the centerline-based method.

The Games–Howell post hoc test was selected to analyze variability of error measurements between the three labels given in Table II for the proposed method since the Levene test proved that the null hypothesis of equal variance can be rejected (test statistics for $E_L = 9.763$ and $E_C = 23.362$).

The post hoc analysis demonstrates that the performance of the proposed method varies significantly ($p(= 0.016) < 0.05$) only between D-NSC (DAPI without SC Sep.) and G-WSC (Geimsa stained). In the meantime, the results of the centerline-based method vary significantly ($p(= 0.000) < 0.05$) between groups D-NSC (DAPI without SC Sep.) and G-WSC (Geimsa stained) as well as groups D-NSC (DAPI without SC Sep.) and D-WSC (DAPI with SC Sep.) ($p(= 0.001) < 0.05$). This shows that the proposed method varies less in performance based on the image group type in comparison to the centerline-based method. The cases with visible sister chromatid separation were also handled better through the proposed method compared to the centerline-based method.

The scatter plot shown in Fig. 8 shows the correlation of measurement error between the two methods. The outliers of the error values (a common shortcoming shared by the proposed and prior approaches) are observed to be contributed by the inaccuracy of the contour partitioning and correcting for sister chromatid separation. The same phenomenon was observed between the error values for the proposed method with and without integrating intensity information. A nonparametric correlation test using the Spearman coefficient revealed a strong correlation between the error values obtained through both these methods ($\rho_{(X,Y)} = 0.250$). The clustering of data points in Fig. 8 corresponds to the majority of this correlation measurements been significant.

To study the benefit of the proposed algorithm, we examined 11 chromosomes from the dataset which had the most positive impact from the addition of intensity into the algorithm. They showed an improvement of error by 20.9 pixels in average. From this set of chromosomes, three were affected by the presence of high degree of sister chromatid separation. The chromosomes that were not affected by this phenomenon showed the presence of a lighter intensity band close to the centromere location. The proposed method has guided the thickness measuring process more accurately by retaining the scan line within this intensity band. On the other hand, only four chromosomes were present in the dataset with a considerable negative bias (average of 26.4 pixels in error). A close examination of these four chromosomes revealed the presence of high degree of sister chromatid separation. We observed that such separation tends to increase the error of centromere detection. Conversely, improved contour partitioning which correctly recognizes sister chromatid separation will result in the proposed integration of intensity onto the Laplacian framework yield better accuracy.

IV. CONCLUSION

We have presented a novel intensity integrated Laplacian-based method for detecting centromere locations in human metaphase chromosomes more accurately. The statistical analysis demonstrated significant improvement in the accuracy of centromere localization through the proposed method in comparison to a centerline-based method [4]. We have presented a framework for incorporating additional features into the Laplacian thickness measurement process and have demonstrated using the intensity feature. We will further explore other possible features such as boundary concavity which can aid the detection of centromere locations as well as other similar measurement problems. Furthermore, an improvement in the accuracy detecting boundary salient points for contour partitioning can increase the accuracy of the proposed algorithm significantly. Further analysis is required including various other existing staining methods to test the performance of the proposed algorithm. We also need to explore inter- and intraobserver variability in manually detecting centromere locations. In addition, we plan to further develop the algorithm to detect dicentric chromosomes.

The proposed method requires 3.85 s per chromosome using MATLAB on an Intel core i5 3.30-GHz processor. Although this is noticeably slower than our previous approach (0.91 s per chromosome), we opted to sacrifice the loss of speed for a substantial gain in accuracy. The main factor for increased

computational burden was observed to be the iterative Laplacian field calculation. We are currently exploring methods for parallelizing the proposed method and effectively reducing the processing time.

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 Recognit. Lett.*, no. 27, pp. 19–28, 2006.

[3] X. Wang, B. Zheng, S. Li, J. J. Mulvihill, and H. Liu, "A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes," *Comput. Methods Programs Bio. Med.*, vol. 89, pp. 33–42, 2008.

[4] A. S. Arachchige, J. Samarabandu, J. Knoll, W. Khan, and P. Rogan, "An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes," in *Proc. Int. Conf. Imag. Process.*, Sep. 2010, pp. 3613–3616.

[5] R. M. Mohammad, "Accurate localization of chromosome centromere based on concave points," *J. Med. Signals Sens.*, vol. 2, no. 2, 2012.

[6] A. S. Arachchige, J. Samarabandu, J. Knoll, W. Khan, and P. Rogan, "An accurate image processing algorithm for detecting fish probe locations relative to chromosome landmarks on dapi stained metaphase chromosome images," in *Proc. 7th Can. Conf. Comput. Robot Vis.*, May 2010, pp. 223–230.

[7] P. Mousavi and R. Ward, "Feature analysis and centromere segmentation of human chromosome images using an iterative fuzzy algorithm," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 4, pp. 363–371, Apr. 2002.

[8] M. Moradi, S. K. Setarehdan, and S. R. Ghaffari, "Automatic locating the centromere on human chromosome pictures," in *Proc. IEEE 16th Symp. Comput.-Based Med. Syst.*, Jun. 2003, pp. 56–61.

[9] E. R. Faria, D. Guliato, and J. C. S. Santos, *Segmentation and Centromere Locating Methods Applied to Fish Chromosomes Images*. Heidelberg/Berlin, Germany: Springer-Verlag, 2005.

[10] A. S. Arachchige, J. Samarabandu, P. K. Rogan, and J. H. M. Knoll, "Intensity integrated Laplacian algorithm for human metaphase chromosome centromere detection," in *Proc. IEEE 25th Can. Conf. Electr. Comput. Eng.*, May 2012, pp. 1–4.

[11] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Imag. Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.

[12] X. Bai, L. J. Latecki, and W. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 449–462, Mar. 2007.

[13] M. Popescu *et al.*, "Automatic karyotyping of metaphase cells with overlapping chromosomes," *Comput. Biol. Med.*, vol. 29, no. 1, pp. 61–82, Jan. 1999.

[14] J. H. Kao, J. H. Chuang, and T. Wang, "Chromosome classification based on the band profile similarity along approximate medial axis," *J. Pattern Recognit. Soc.*, vol. 41, pp. 77–89, 2008.

[15] M. S. Hassouna and A. A. Farag, "Variational curve skeletons using gradient vector flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2257–2274, Dec. 2009.

[16] L. J. Latecki and R. Lakämper, "Polygon evolution by vertex deletion," in *Proc. 2nd Int. Conf. Scale-Space Theor. Comput. Vis.*, 1999, pp. 398–409.

[17] C. Xu and B. Kuipers, "Object detection using principal contour fragments," in *Proc. Can. Conf. Comput. Robot. Vis.*, May 2011, pp. 363–370.

[18] O. Golubitsky and S. M. Watt, "Online stroke modeling for handwriting recognition," in *Proc. 18th Annu. Int. Conf. Comput. Sci. Softw. Eng.*, 2008, pp. 72–80.

[19] V. Mazalov and S. M. Watt, "Digital ink compression via functional approximation," in *Proc. 12th Int. Conf. Frontiers Handwriting Recognit.*, 2010, pp. 688–694.

[20] M. F. Valstar and M. Pantic, "Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics humancomputer interaction," in Proc. *IEEE Int. Conf. Human-Comput. Interact.*, 2007, pp. 118–127.

[21] M. I. Miller, A. B. Massie, J. T. Ratnanather, K. N. Botteron, and J. G. Csernansky, "Bayesian construction of geometrically based cortical thickness metrics," *NeuroImage*, vol. 12, no. 6, pp. 676–687, 2000.

[22] H. Haidar and J. Soul, "Measurement of cortical thickness in 3D brain MRI data: Validation of the Laplacian method," *NeuroImage*, vol. 16, pp. 146–153, 2006.

[23] S. E. Jones, B. R. Buchbinder, and I. Aharon, "Three-dimensional mapping of cortical thickness using Laplaces equation," *Hum. Brain Mapp.*, vol. 11, pp. 12–32, 2000.

Akila Subasinghe Arachchige received the B.Sc. degree in engineering from the University of Moratuwa, Moratuwa, Sri Lanka, and the M.Sc. degree in engineering from Western University, London, ON, Canada. He is currently working toward the Ph.D. degree in electrical and computer engineering at the Western University.

His research interests include medical image analysis, machine learning, and robotic systems. He is currently working on a research project to automate radiation dosimetry using metaphase lymphocyte chromosome cell images.

Jagath Samarabandu received the M.S. and Ph.D. degrees at SUNY Buffalo, Buffalo, NY, USA.

Since 2000, he has been with Western University, London, ON, Canada, where he is currently an Associate Professor of Electrical and Computer Engineering. He was a Fulbright Scholar. His research activities include image analysis, computer vision, and pattern recognition. He has more than 20 years of academic and industrial experience in this domain. His research interests include computer vision, pattern recognition, image understanding, machine learning, and cyber security.

Joan H. M. Knoll received the B.Sc. degree from Western University, (WU) London, ON, Canada, and the Ph.D. degree from the University of Manitoba, MB, Canada, and the Postdoctoral degree from Harvard University, MA, USA.

She is a Professor at WU and a Medical Leader in Molecular Diagnostics, London Health Sciences Centre, ON, Canada, and a Chief Scientific Officer of Cytognomix. She is board certified in cytogenetics and molecular genetics in the USA and Canada. She has 21 years of experience in directing two CLIA-certified and CAP-accredited cytogenetic laboratories and 30 years of experience in human cytogenetics. She with Rogan founded Cytognomix based on their inventions of cytogenetic technologies and they hold patents on multiple genome-related technologies. Their work has been supported by NCI, NSF, and Canadian federal and provincial agencies. The NSF project led by Dr. Rogan developed a remote automated metaphase microscopy system for chromosome image capture and prioritization.

Peter K. Rogan received the B.S. degree from Johns Hopkins, Baltimore, MA, USA, the Ph.D. degree from Yale University, CT, USA, and the postdoctoral degree from the U.S. National Cancer Institute.

He is a President of Cytognomix and a Professor of Biochemistry, Computer Science, and Biomedical Engineering at Western University. He holds the Canada Research Chair in Genome Bioinformatics and is a Member of the American Society of Human Genetics, and a Prolific Inventor and Serial Entrepreneur (previously founding CEO of Phylogenetix Laboratories Inc., a Pennsylvania company). He has more than 20 years of experience in developing multiple software applications that are widely used by molecular diagnostics laboratories. He previously directed a U.S. National Science Foundation-funded project at the University of Missouri (in which Knoll was a coinvestigator) which built an automated molecular cytogenetics microscope system, capable of capturing and ranking optimal metaphase cell images, stored them in an online database, which had the facility to recapture metaphase cells through remote control of the microscope.