Original Article

www.jmss.mui.ac.ir

Accurate Localization of Chromosome Centromere Based on Concave Points

Mohammaed Reza Mohammadi

Department of Biomedical Signal and Image Processing Lab., Sharif University of Technology, Tehran, Iran

Submission: 26-01-2012 Accepted: 02-04-2012

ABSTRACT

Analyzing the features of the chromosomes can be very useful for diagnosis of many genetic disorders or prediction of possible abnormalities that may occur in the future generations. For this purpose, karyotype is often used, for which to be made, it is necessary to identify each one of the 24 chromosomes from the microscopic images. Definition and extraction of the morphological and band pattern-based features for each chromosome is the first step to identify them. Centromere location is an important morphological feature. In this paper, a novel algorithm for centromere localization is presented. The procedure is based on the calculation and analyzing the concavity degree of the chromosome's boundary pixels. In this method, the centerline of the chromosome is computed and the score of each pixel on the centerline is considered as the sum of the concavity degree of two pixels on the chromosome's boundary that are perpendicular to it. Finally, location of the centromere is estimated as one pixel on the centerline which is corresponding to the maximum score. When applied the proposed algorithm on 50 images, an average error of 2.25 pixels for centromere localization is achieved.

Key words: *Centerline, chromosome centromere, concave points, karyotyping, polynomial fitting*

INTRODUCTION

In cytogenetic, the analysis of chromosomes is useful for many biological applications. Human inherited diseases, for example, are detectable by observing certain chromosomes of the existing 46 chromosomes in human body.[1] Karyotype, a systemized array of the chromosomes of a single cell prepared either by drawing or by photography,[2] is often used for this purpose. To make a karyotype it is necessary to identify each one of the 46 chromosomes (22 pair of autosomal and a pair of sex chromosomes).

Karyotyping consists of the identification, classification, and presentation of the 23 pairs of the chromosomes in a single picture. This process, which is usually done manually by a human expert, is a difficult and time-consuming task. In conventional karyotyping, giemsa-banded cells are photographed under a light microscope (an example picture is shown in Figure 1a) during the metaphase stage.[3] The result of the karyotyping process for Figure 1a, which is done manually by a cytogeneticist, is shown in Figure 1b. Two stages of this process are segmentation and classification of the chromosomes.

It is certainly of interest to have accurate techniques in doing karyotyping. Inaccuracy may lead to disastrous consequences, for example, when dealing with disease identification. It may cause a misleading observation on one's set of chromosomes that may lead to false diagnosis on the patient's condition. Of course, one would rather have traditional methods, yet somewhat still accurate, to diagnose the disease he has, than become paralyzed or dead due to unintentional medical malpractice.[4]

Features used in chromosome classification generally fall into two main categories of the geometrical features and the band pattern-based features.[5] The length of the chromosome and the centromeric index (CI) are the most important geometrical features. CI is the ratio of the length of the short arm of the chromosome to its long arm. These two arms are separated from each other in a point called centromere.

Based on the location of the centromere along the chromosomes, there are three classes defined for them.[3] In some chromosomes, which are called metacentric, the centromere is located in the middle of the long axis of the chromosome and the two arms are almost of the same length (therefore CI ≅1). Chromosomes number 1, 3, 16, 19, and 20

Address for correspondence:
Mr. Mohammaed Reza Mohammadi, Biomedical Signal and Image Processing Lab., Sharif University of Technology, Tehran, Iran.
E-mail: mrmohammadi@ee.sharif.edu

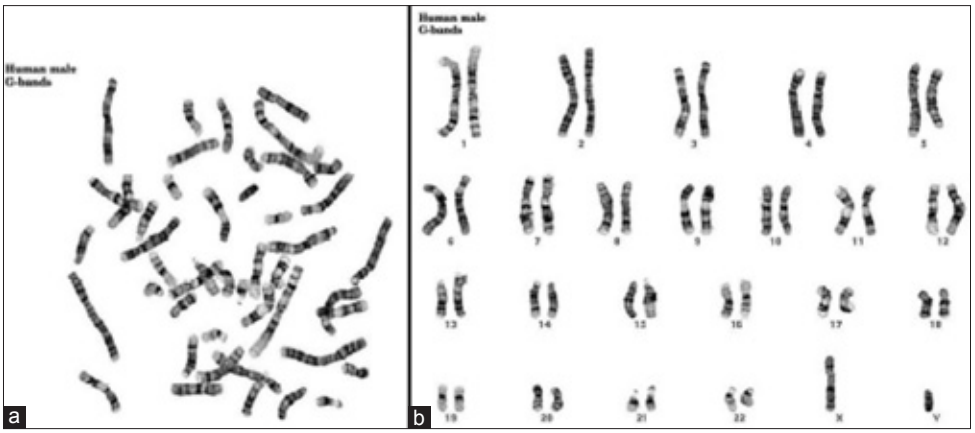

Figure 1: Karyotyping, (a) G-banded chromosomes as seen under a light microscope, and (b) corresponding karyotype (a male)[3]

are metacentric. A chromosome is called acrocentric when the centromere divides the chromosome into two arms of unequal length. Chromosomes number 13, 14, 15, 21, and 22 belong to this class. In the last class of the chromosomes named telocentric the short arm is very small and the centromere is located near to one of the two ends of the chromosome. From these explanations, it is clear that CI is an important feature for classification of the chromosomes.

Automatic classification of chromosomes has been a well-studied problem in the last four decades.[6-11] Natural complexity of the problem is caused by various unpredictable appearances of the chromosomes due to nonrigid nature of them.

The proposed algorithm[3] is based on the calculation and analyzing the vertical and horizontal projection vectors of the binary image of the chromosome. This algorithm cannot apply on the rotated or highly bent chromosomes. To remove this restriction, the projections are performed on the medial axis or skeleton of the chromosome in the algorithms.[5,12] In other words, the using feature in these algorithms is the length of the line segment that connects two boundary pixels and is perpendicular to the skeleton. This feature is raised from this sentence: "the centromere is located in the narrowest part of the chromosome along its longitudinal direction."[3] This feature is very effective, but it is sensitive to noise.

The block diagram of the proposed algorithm for centromere localization is shown in Figure 2. Input of this block diagram is an image of one chromosome which is cropped manually. The first step in the proposed algorithm is to segment the input image which a simple thresholding algorithm.[13] A sample result of this thresholding algorithm is shown in Figure 3. The focus of this paper is on the next steps. In these steps, the centerline of the chromosome region is computed and a polynomial curve is fitted to it. On the other hand, the proposed feature (concavity degree) is computed for the pixels of the chromosome's boundary. Then, the

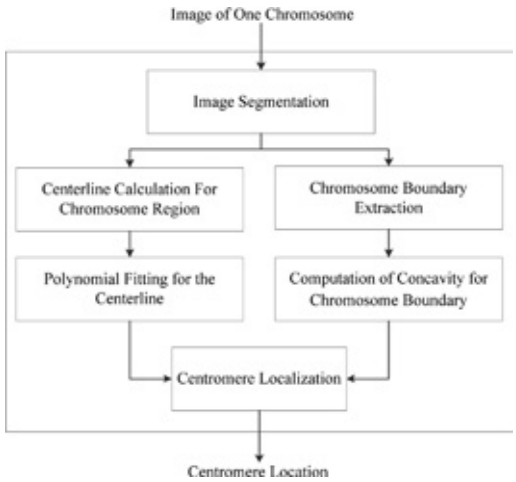

Figure 2: Block diagram of the proposed algorithm for centromere localization

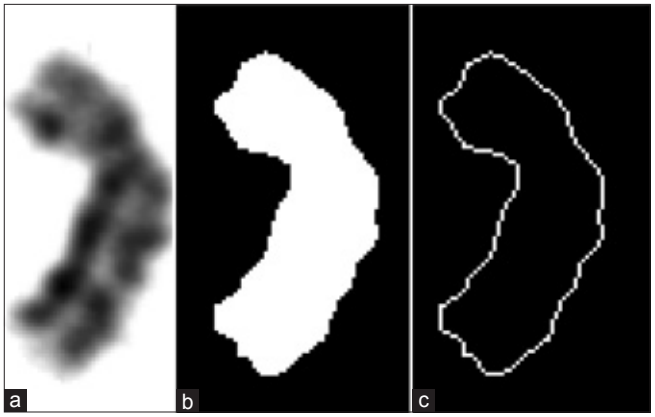

Figure 3: Segmentation of the chromosome, (a) original image, (b) segmented image by the method,[13] and (c) the chromosome's boundary

concavity degrees are projected to the centerline and the corresponding pixel to the maximum concavity is chosen as the centromere location.

The remainder of this paper is structured as follows. The chromosome centerline computation is discussed in the "Chromosome Centerline Computation" section.

The proposed feature, concavity degree, and the method of its computation are presented in the "Concavity Degree Calculation" section. Experimental results are given on real images in the "Centromere Localization" section. Finally, the "Experimental Results" section concludes the paper.

CHROMOSOME CENTERLINE COMPUTATION

Real chromosomes have nonrigid nature and may observe bend [Figure 3a]. Therefore, the search for the centromere location cannot be performed on a straight line (i.e., vertical or horizontal). So, the centerline of the chromosome region may be used to search for the centromere location. Skeleton of the Figure 3b is plotted in the Figure 4a. As can be seen, the raw skeleton consists of some undesirable branches and its search is very hard. To overcome this problem, some algorithms are proposed to prune the skeleton. In this paper, the proposed algorithm[14] is used whose result is depicted in Figure 4b.

In our proposed method of search the centerline pixels, the perpendicular line to each pixel is required. In addition, the centerline may be noisy. Thus, to reduce the noise effect and to obtain an equation for the slope of the centerline's pixels, two polynomial curves of degree 4 are fitted to the *x*- and *y*-coordinates of the centerline. If the polynomial coefficients of the *x*- and *y*-coordinates store to p_1 and p_2 vectors, respectively, (1) and (2) are the corresponding equations of the fitted curves:

$$x_t = p_{14}\, t^4 + p_{13}\, t^3 + p_{12}\, t^2 + p_{11}\, t + p_{10} \quad (1)$$

$$y_t = p_{24}\, t^4 + p_{23}\, t^3 + p_{22}\, t^2 + p_{21}\, t + p_{20} \quad (2)$$

where $t = 1, 2, \ldots, n$ and n is the number of centerline pixels. This fitted curve of Figure 4b is shown on the original chromosome image in Figure 4c. Further computations on the centerline will present in the "Centromere Localization" section.

CONCAVITY DEGREE CALCULATION

Figure 5 shows a synthetic image of a chromosome. From this figure, it is observed that the centromere line has the shortest length among all the perpendicular lines. Nevertheless, for real images, this feature has sensitivity to noise.

Another important feature of the chromosome line is the concavity of its endpoints [Figure 5]. Therefore, the focus of this paper is on this novel feature and development of an algorithm to centromere localization based on this feature.

Angles and curvature are probably the most widely used features for concavity calculation. However, both angle and curvature are vulnerable to noise, especially when the segmentation step cannot produce a neat and clean contour due to the noise. In this paper, the property of convex regions [Figure 6] is used to define a new concavity calculation method.

Based on the above discussion, for the concave points, the straight line segment between two near boundary points is outside of the region. This property is shown in Figure 7.

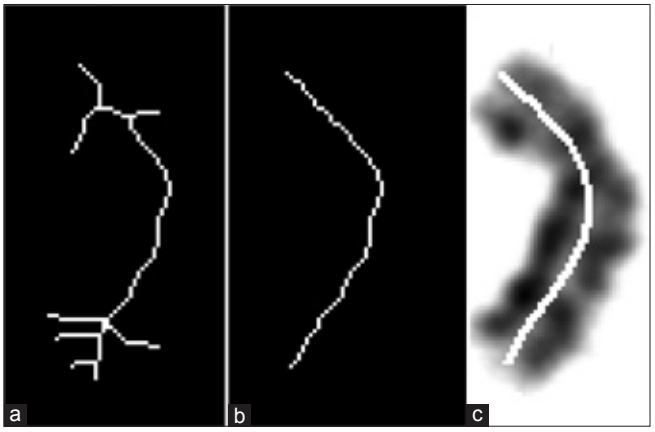

Figure 4: Computation of the chromosome's centromere, (a) initial skeleton, (b) obtained centerline by the method,[14] and (c) plot of the fitted curve to the centerline on the original image

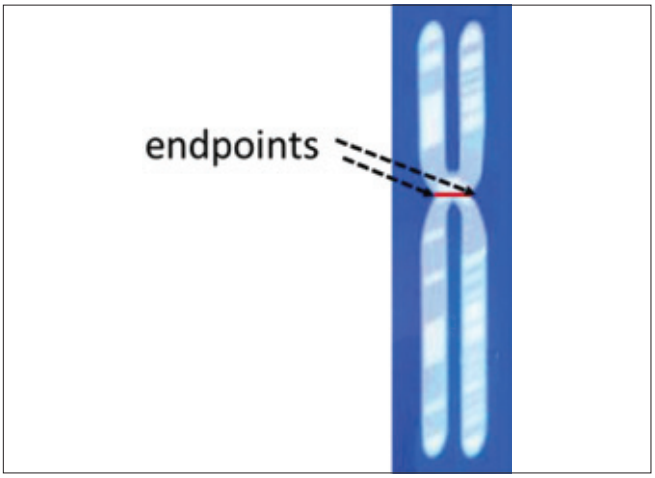

Figure 5: Two endpoints of the centromere line are concave points

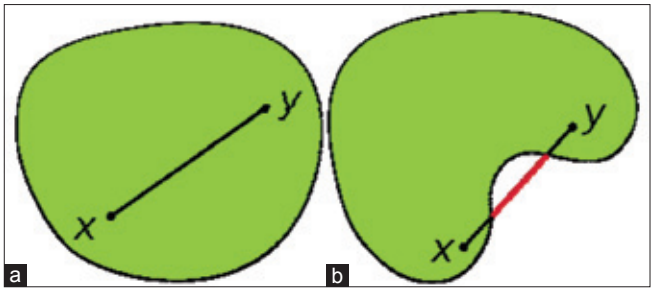

Figure 6: Demonstration of convex and nonconvex regions, (a) in a convex region, for every pair of points within the region, every point on the straight line segment that joins them is also within the region, (b) in a nonconvex region, previous condition is not valid for some pair of points[15]

To compute the concavity degree of each boundary pixel, its corresponding boundary points by distance h are considered as the two endpoints of its line segment. Then, this line is plotted and its points that are not on the region are enumerated. The concavity degree is defined as the ratio of outside points of the line segment to the total number of the line segment points (3):

$$C_i = \frac{\sum L_i \cap R'}{\sum L_i} \tag{3}$$

where R is the region of the chromosome and R' is its complement. Also, L_i is the region of the line segment and C_i is the concavity degree corresponding to the ith boundary pixel. Moreover, operator Σ calculates the number of zeros in the corresponding operand. Thus, the value of ΣL_i is equal to $2h$-1 (two endpoints are not considered as the line segment points). Figure 8 shows the proposed method on an image. In this figure, pixel 1 is a fully concave point and by considering its corresponding line segment, its C_i is 1. On the other hand, pixel 2 is a fully convex point and by considering its corresponding line segment, its C_i is 0.

Concavity degrees for the chromosome's boundary of Figure 3c are depicted in Figure 9. In this figure, the brighter pixels correspond to the larger concavity degrees. As can be comprehended visually, for the more concave pixels, the value of C_i is larger.

Thus, the boundary pixels that have larger C_i are candidates for the endpoints of the centromere line. In the next section, the proposed algorithm for the centromere line detection based on the concavity degrees is presented.

CENTROMERE LOCALIZATION

The search for the centromere location is performed on the centerline curve. On the other hand, the concavity feature is defined on the boundary pixels. Therefore, the boundary pixels that are corresponded to the centerline pixels should be found. To find these boundary pixels, the perpendicular line for any centerline pixel is obtained and intercrossed by the boundary pixels. For any pixel on the centerline of the chromosome, the perpendicular slope to it may compute by (4)

$$s_t = -\frac{1}{\frac{dy}{dx}} = -\frac{\frac{dx}{dt}}{\frac{dy}{dt}} = -\frac{4p_{14}t^3 + 3p_{13}t^2 + 2p_{12}t + p_{11}}{4p_{24}t^3 + 3p_{23}t^2 + 2p_{22}t + p_{21}}. \tag{4}$$

The equation of the line that passes through the current pixel with slope s is (5)

$$Y = s_t\left(X - x_t\right) + y_t. \tag{5}$$

Figure 10 illustrates three samples of these perpendicular

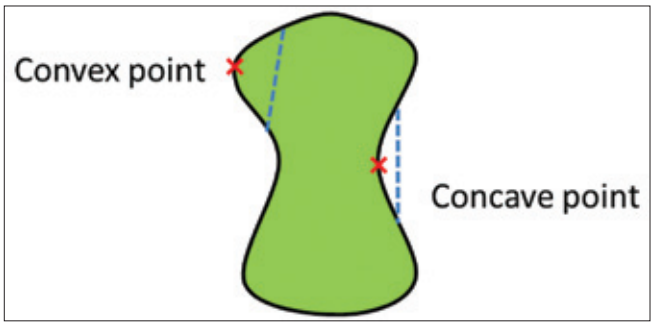

Figure 7: Illustration of concave and convex points

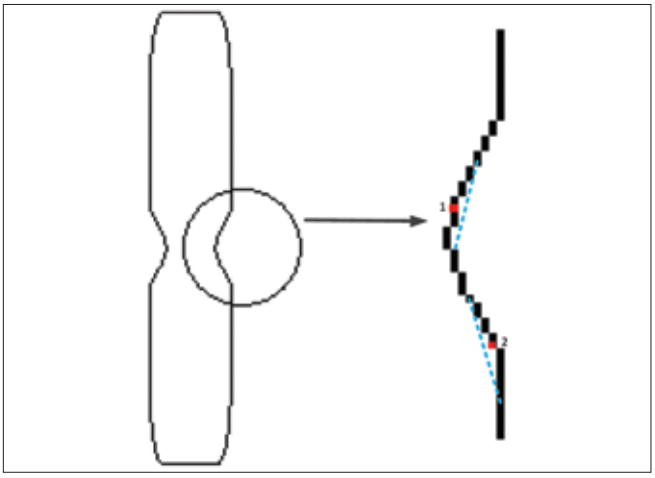

Figure 8: Demonstration of the proposed method to computation of concavity degrees of two pixels. For pixel 1, which is a concave point, the corresponding line (that connects two pixels by h distance before and after it) has no intersection by the region. On the other hand, for pixel 2, which is a convex point, all of the corresponding line overlies on the region

Figure 9: Concavity degrees of the boundary pixels (brighter colors correspond to the larger concavity degrees)

lines. Two pixels in the two sides of the chromosome boundary that this line passes through can be found (L_t and R_t). The score of this pixel on the centerline is the sum of the concavity degrees for two boundary pixels as in (6):

$$S_t = C_{L_t} + C_{R_t}. \quad (6)$$

So any pixel on the centerline that has larger S_t may be the location of the centromere and the corresponding line may be the centromere line.

Figure 11 shows the S_t for the chromosome of Figure 10. The centromere location is corresponded to the maximum value in this curve. In addition, the estimated centromere line is shown in Figure 12.

EXPERIMENTAL RESULTS

The estimated centromere lines for four sample chromosomes are shown in Figure 13. This figure may depict the high quality of the proposed method.

To quantify the accuracy, for 50 images of chromosomes the centromere location are marked manually. Then, the proposed algorithm is applied on them and distance between the results of algorithm and the markers are computed as the error of the algorithm. Eight samples of the manually marks (yellow square) and the results of the proposed algorithm (green triangle) are shown in Figure 14.

Average error of the proposed algorithm is listed in Table 1. The width of test images is about 100 and their height is about 200 pixels. Moreover, the average error using the shortest line and combination of the shortest line and larger concavity are listed in Table 1. From this table, it can be observed that the average

Table 1: Comparing the average and variance of the errors of the proposed method using three features: larger concavity, shortest line, and their combination

Features	Average (pixel)	Standard deviation (pixel)
Concavity degree	2.25	4.47
Line segment length	6.09	10.08
Combination of two features	1.41	1.61

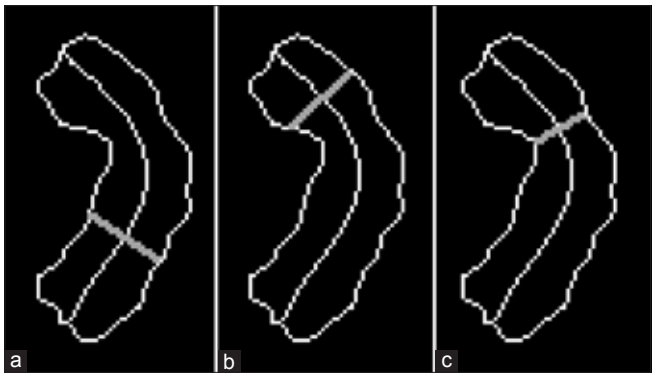

Figure 10: Three samples of the perpendicular lines to the centerline pixels and their corresponding boundary pixels

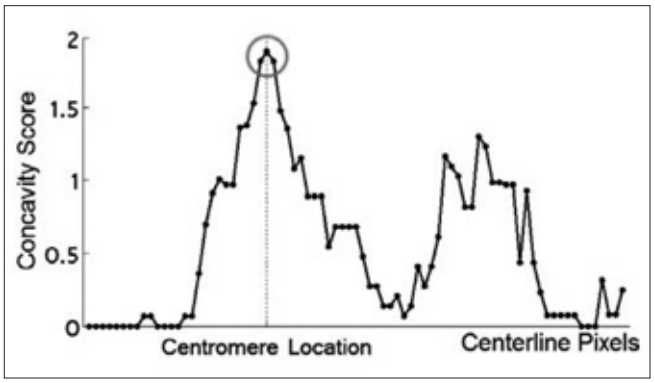

Figure 11: Concavity score of the centerline pixels. The maximum value is corresponded to the centromere location

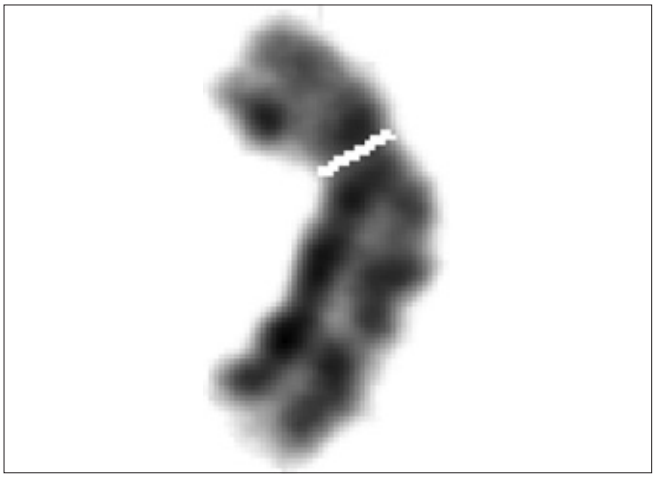

Figure 12: Estimated centromere line for a sample chromosome

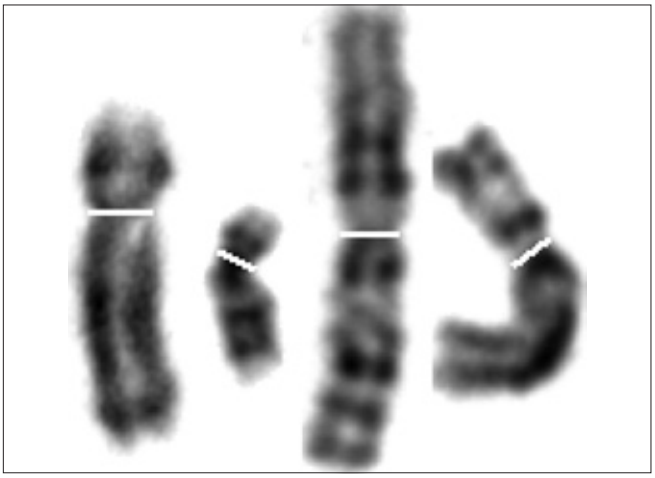

Figure 13: Estimated centromere lines for some sample chromosomes

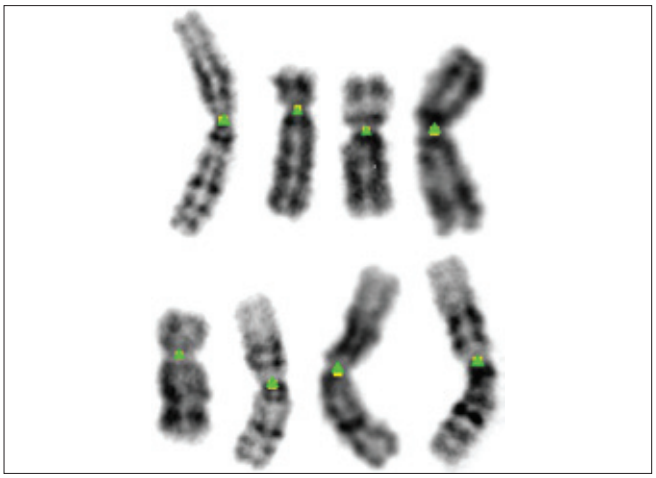

Figure 14: Eight samples of the test images. Manually marks are depicted by the yellow square and the results of the proposed algorithm are depicted by the green triangle

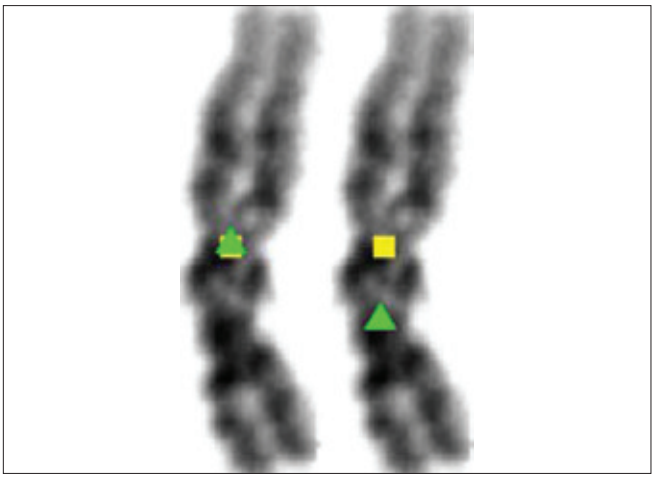

Figure 15: An example of the higher performance of the concavity degree from the line distance, (a) result of the proposed method using concavity degree, and (b) using the lie distance

error using concavity degree is smaller than the average error using the line segment length. Moreover, the combination of these two features results in a smaller average error.

Figure 15 is an example of the higher performance of the concavity degree from the line segment length. Parts (a) and (b) in this figure are corresponded to the concavity degree and line segment length features, respectively. So, for this example the performance of concavity degree is higher (from Table 1, average error of this feature also is smaller). This higher performance corresponds to the robustness of concavity degree to the noise and other nonideal conditions.

CONCLUSION

An accurate algorithm for locating the centromere in a microscopic image of a human chromosome was presented. Centromere locating is important for feature extraction and classification of the chromosomes, which is a necessary step toward automatic karyotyping. The algorithm is based on the calculation of the concavity degree for boundary pixels of the chromosome region and projecting them to the centerline. The algorithm was applied to 50 real chromosome images. The mean error (Euclidean distance between the reference and automatically extracted centromere locations) is about 2.25 pixels, which is small and the accuracy may be satisfactory.

The combination of the concavity degree and line segment length results in smaller error. In this paper, a weighted sum of these features was used. Thus, the method of their combination and definition of new features can be studied in the future works. In addition, the input of the proposed algorithm was an image of one chromosome which was cropped manually. Thus, another future work is to automatically crop the chromosome images.

REFERENCES

1. DeCherney AH, Nathan L. Current Diagnosis and Treatment Obstetrics and Gynecology. 10th ed. USA: McGraw-Hill; 2007.
2. Available from: http://www.pathology.washington.edu/galleries/Cytogallery/karyotype.html. [Last accessed on 2012].
3. Moradi M, Setarehdan SK, Ghaffari SR. Automatic locating the centromere on human chromosome pictures, in proceedings of the 16th IEEE conference on Computer-based medical systems: New York, New York; 2003.
4. Trimananda R. Chromosome centromere and chromatid's banding identification using pattern vector, in proceedings of the 2010 second international conference on advances in computing, control, and telecommunication technologies, 2010.
5. Jong Man C. Chromosome classification using backpropagation neural networks. IEEE Eng Med Biol Mag 2000;19:28-33.
6. Castleman KR, Melnyk JH. An automated system for chromosome analysis: final report: Jet Propulsion Laboratory. United States: California Institute of Technology; 1976.
7. Lerner B, Levinstein M, Rosenberg B, Guterman H, Dinstein L, Romem Y. Feature selection and chromosome classification using a multilayer perceptron neural network, in Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on, vol.6. 1994. p. 3540-5.
8. Guimaraes LV, Schuck A, Elbern A. Chromosome classification for karyotype composing applying shape representation on wavelet packet transform, in Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE. Vol.1. 2003. p. 941-3.
9. Wu Q, Liu Z, Chen T, Xiong Z, Castleman KR. Subspace-based prototyping and classification of chromosome images. IEEE Trans Image Process 2005;14:1277-87.
10. Kao JH, Chuang JH, Wang T. Chromosome classification based on the band profile similarity along approximate medial axis. Pattern Recogn 2008;41:77-89.
11. Oskouei BC, Shanbehzadeh J. Chromosome Classification Based on Wavelet Neural Network. International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2010, p. 605-10.

12. Moradi M, Setarehdan SK, Ghaffari SR. Automatic landmark detection on chromosomes' images for feature extraction purposes. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, Vol. 1. 2003. p. 567-70.
13. Otsu N. A Threshold selection method from gray level histograms. IEEE Trans Syst Man Cybern 1979;9:62-6.
14. Arachchige AS, Samarabandu J, Knoll J, Khan W, Rogan P. An image processing algorithm for accurate extraction of the centerline from human metaphase chromosomes, in Image Processing (ICIP), 2010 17th IEEE International Conference on Image Processing, 2010. p. 3613-6.
15. Otsu N. A Threshold selection method from gray level histograms. IEEE Trans Syst Man Cybern 1979;9:62-6.

How to cite this article: ***

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHY

Mohammad Reza Mohammadi was born in Qom in Iran, on July 25, 1987. He received undergraduate and MSc degrees from Amirkabir University of Technology in 2009 and 2011, respectively. He is currently a PhD student of Sharif University of Technology and researches in Biomedical Signal and Image Processing Lab (BiSIPL). His interests and research include biomedical image processing, machine vision, and machine learning.

E-mail: mrmohammadi@ee.sharif.edu