PIXEL SENSIBLE LOCAL BAND ANALYSIS IN MICROSCOPIC CHROMOSOME IMAGES USING CSPA

D. SOMASUNDARAM

[1] Department of ECE, Sri Shakthi Institute of Engineering and Technology, Coimbatore
E-mail: somgce@gmail.com

In chromosome analysis, local band analysis plays the main role to identify the perfect matched chromosome in metaspread images to attain the karyotyping. Literature investigations are narrow in chromosome image band analysis due to the higher complexities. In this paper, Pixel level based Conditional Seed Point Algorithm (CSPA) is proposed. This simulation algorithm separates the weak band region to the strong band region, and the strong band region area evaluated was based on the Region of Seed condition Points. This algorithm works well for different intensity levels and adopts the structural changes to identify the bands in image. This algorithm was simulated in more than 450 individual chromosomes to identify the local bands in the chromosome images and provided the accuracy more than 96%.

Keyword*: chromosome, local bands, seed point algorithm, karyotyping.*

Introduction. Chromosome image analysis is an important process in the field of cytogenetics to identify the chromosome abnormalities. Karyotyping [1] is a process used to analyse these chromosomes for different genetic problems like Down syndrome, turner syndrome, etc., so many techniques were proposed for karyotyping in image processing. In metaphase cell spread images, segmentation of overlapped and non-overlapped chromosome identification and segmentation are necessary to build the chromosome karyotype. In which, normal translocation-based labeling method [2, 3] for M-fish chromosomes is analyzed. Fuzzy c-means based segmentation [4] provides the background correction and performs the segmentation. Genetic algorithm based segmentation [5] analyzed search and optimization technique and contour method as used to identify the boundaries. Watershed segmentation [6] applied to M-fish chromosomes for segmentation process. This method performs better than other method not in the boundary regions of the overlapped chromosomes. K-means [7] segmentation is an initiative to other segmentation methods. In which the cluster deformation is high compare to watershed segmentation. This leads the data inaccuracy in the segmentation. Fuzzy subset based segmentation performed segmentation based on shape breakdown. From which many automated and semi-automated segmentation and karyotyping methods are proposed earlier. In which, automatic and semi-automatic karyotyping of chromosomes [8–10, 18] are detailed in studies. Computer based karyotyping initially introduced [11] based on the virtual reality on the shapes. These literatures focused only about the segmentation and classification of human chromosomes based on the length, centromere position. Only limited literatures are explained the concepts for local bands in chromosomes. Presented a technique for local band descriptor based classification [12], explained the concept of band profile analysis for chromosomes [13].

In this paper, we proposed a novel approach for local bands extraction in chromosomes based on the pixel sensibility (Fig. 1). The band profile and the centromere location are identified along the medial axis. Based on the centromere position p-arm and q-arm lengths are considered. To improve the accuracy of the local band extraction, pixel wise conditional analysis is carried out over the chromosome image. Section 2, focused on the medial axis extraction and centromere location identification. Section 3 proposed a Conditional Seed Point Algorithm for band extraction. Section 4 presents the experimental results and section 5 present the conclusion of this paper.

Methods and materials. *Chromosome medial axis extraction*. The chromosome midline is to obtain a curve to represent the approximate curvature of the chromosome. This is the pre-processing step for chromosome shape alteration. The human chromosome shape in metaphase is basically linear with the length much larger than the width; therefore, in practice, the width may be neglected so that a midline can be created to represent the chromosome curvature (Fig. 2, *f*). The medial axis curve

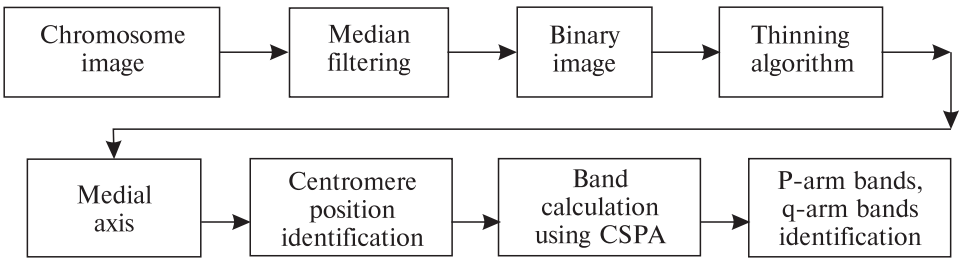

Fig. 1. Proposed method

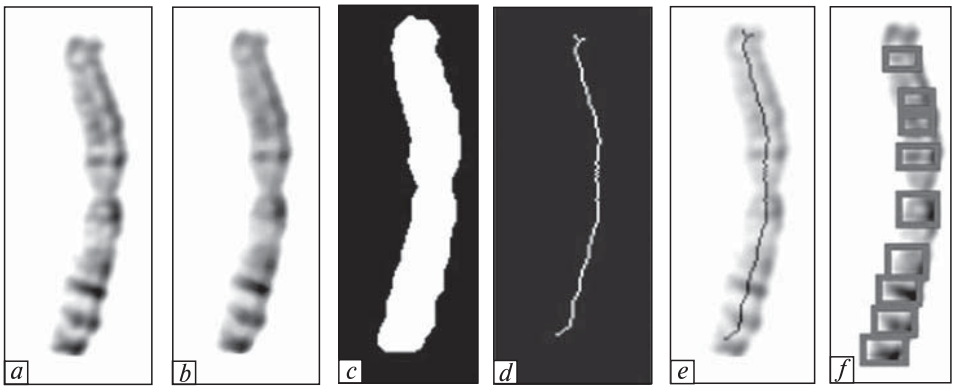

Fig. 2. Chromosome image (*a*), filtered image (*b*), binary image (*c*), thinned image (*d*), medial axis (*e*), band patterns (*f*)

is in parallel to the two sides with the same distances. The midline curve can be obtained in three steps: boundary detection, thinning, and curve fitting. Boundary detection is to acquire the boundary of chromosome image. Thinning is to convert binary shape obtained from boundary detection to a 1-pixel wide curve. Curve fitting is used to find the «best fit» line or curve for a series of data points. Medial axis has to be extracted from thebinary converted chromosome image, to obtain the skeleton the chromosome image for this ketler thinning algorithm is used.

Centromere identification. Centromere is a part of chromosome and helps in linking the sister chromatid. Centromere position helps in homologous chromosome classification. The centromere positions are identified by several steps. The initial input is taken as an individual chromosome image. The second step is the conversion of input image to binary image. In binary image if there is any opening that can be closed by filling process so a good binarized image can be obtained. The explanation of the above steps is already done for segmentation operation. The third step is the Euclidean distance transform. In this medial axis is identified in the image for centromere location. For extracting the medial axis a distance transform is considered. Finally the centromere of the chromosome is obtained. With this the segmentation output is compared and the homologous chromosome is identified.

Centromere extraction. The centromere is identified by a simple distance transformation technique. The distance transform provides a metric or measure of the separation of points in the image. It computes the Euclidean distance transform of the binary image. For each pixel in the image, distance transform assigns a number that is the distance between that pixel and the nearest nonzero pixel for any dimensional image. The input image is initially converted to binary image. The image is then processed by distance transform technique. A profile is drawn perpendicular to in the transformed image which gives a plot through which the location of centromere can be identified. Two different type of profile is considered in this work namely density profile, shape profile. Density profile is the average grey scale value of all perpendicular line across the medial axis of a chromosome image. It is computed by the given formula

$$D(x) = [\sum_{i=1}^{n} g_i(x)/n],$$

where $g_i(x)$ is the Gray value of each pixel in a perpendicular line and n is the number of all pixels

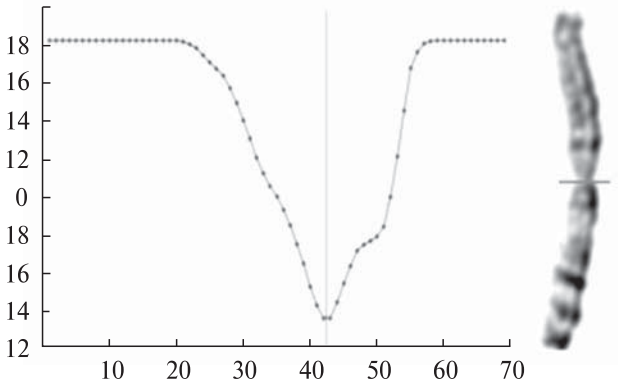

Fig. 3. Centromere identification using Density Profile of chromosome

in each perpendicular line. The computer scheme applies a median filter to reduce possible impulses and noise in the density profile. A shape profile gives the weighted width for all perpendicular line across the medial axis of a chromosome image. It is defined by the formula

$$s(x) = \sum_{i=1}^{n} [g_i(x) \cdot d_i(x)^2] \; / \sum_{i=1}^{n} d_i(x)^2.$$

Shape profile corresponds to the sum of the product of the grey scale value $g_i(x)$ and its corresponding Euclidean distance $d_i(x)$ away from the medial axis of the perpendicular line, divided by the sum of the distance (Wang et al., 2008). With the shape and density profile the centromere position is calculated. The segmented output image is taken and verified with the outputs obtained for the homologous identification. The homologue chromosome images are identified by the centromere position. The original chromosome image centromere position is identified and the same algorithm is used for the separated chromosome image.

Conditional Seed Point Algorithm (CSPA). In this seed point Based analysis, Pixel wise analysis is carried out in the chromosome image. In that, each colour spaces are separated and any one of the colour space is considered. This does not make any deviation in the image. Because of all the colour space have same pixel values in the chromosome image. So that, in RGB colour space any one of the colour space is considered. Here Red region is considered for the process. For the seed point analysis, certain properties and conditions are considered and evaluated to confirm the band regions in the chromosome image.

Algorithm flow: read the input image; 3X3 sliding window created over the image; centromere position is identification; conditional seed point selection region and strong band region conditions applied; band regions extracted based on the seed point area; P-arm and q-arm bands are calculated

$$\begin{matrix} P(X_1, Y_1) & P(X_1, Y_2) & P(X_1, Y_3) \\ P(X_2, Y_1) & P(X_2, Y_2) & P(X_2, Y_3) \\ P(X_3, Y_1) & P(X_3, Y_2) & P(X_3, Y_3) \end{matrix}$$

Conditions for the Conditional Seed point selection region:

1. Consider 3X3 window elements in a chromosome image, any two pixels should be identical then respective pixel is taken as seed point.

2. Deviation between the pixels should not exceed the value of 10. i.e., $P(X_m, Y_n) - P(X_m, Y_n) \leq 10$.

3. Elements should not 255 and 0 $P(X_m, Y_n) \neq \neq 255$, $P(X_m, Y_n) \neq 0$.

Conditions for the strong band region:

4. In the each seed point in the window deviation and the distance should be minimum, in each 3X3 window, any two identical pixels occurs in the window then the midpoint pixel is consider for band

Table 1. **Sample region in chromosome**

Region selection	Horizontal image Range (pixels)								
Vertical image Range (pixels)	141	97	84	93	103	116	146	160	187
	125	83	60	69	77	98	118	141	168
	116	80	60	61	77	96	114	136	148
	126	91	77	80	100	112	121	132	145
	128	102	99	110	123	134	142	153	155
	127	119	127	140	150	165	170	173	166

region seed point and the matched value pixels are considered for band region area pixels.

5. If the pixels are identical and having the large deviations with the neighborhood pixels then it's considered for the weak band region.

Tabl. 1 shows the region of chromosome image consist the band and non-band region.

In Tabl. 2 3X3 window, comparing the pixel values, conditional seed point is 83, the identical value with minimum deviation pixel value is 60. In the second table (2) seed point is 60 and the region of ROC pixel is 60.

In Tabl. 3, the pixel values are not identical and deviation is more than 10. So this window consider as non-band window. In this illustration, the seed points are, Conditional Seed points: 83, 60, 69, 77, 98, 80; Region of Band area pixels: 60, 77.

From these points, the band region is Tabl. 4.

This algorithm applied to the chromosome images and this algorithm extracted the band regions in the chromosome.

Table 2. **3X3 window, Seed Point**

Seed point	Horizontal region (pixels)		
	Region 1		
Vertical region (pixels)	141	97	84
	125	83	60
	116	80	60
	Region 2		
Vertical region (pixels)	97	84	93
	83	60	69
	80	60	61

Table 3. **Non-band region**

Non-band region 1	Horizontal region (pixels)		
Vertical region (pixels)	116	146	160
	98	118	141
	96	114	136

Experimental Results and Discussion. CSPA algorithm tested over 300 normal chromosomes and 200 bended chromosomes. In Fig. 2, *a* shows the individual chromosomes segmented from Meta spread chromosome image. Median filter is applied to the image to remove the spikes in the images and it preserves the edges of chromosome images. Filtered image respectively converted in to Binary image using Otsu method, to identify the external boundaries. Stentiford Thinning algorithm is applied to the chromosome image. This algorithm provides the medial axis of the chromosome. This axis combined with the original image, Fig. 2, *e* to identify the medial axis in the chromosome image. For this image the density profile is analyzed and identified that centromere position to identify the p-arm and q-arm in chromosome image shown in Fig. 3.

To identify the band patterns proposed algorithm is applied to the image and the bands are calculated above and below the centromere region i.e., in p-arm and q-arm regions. Fig. 2, *f* shows the Band Patterns in the chromosome.

Conclusion. This proposed algorithm works well for female and male chromosomes to identify the local bands in the chromosomes. This improves the accuracy of the identification in the bands in each chromosome. Local band based classification of each chromosomes are possible and it will improves the accuracy of classification. Proposed method applied to 500 individual male and female chromosomes. Compare to existing methods, proposed method gives better identification of bands in each chromosomes.

Table 4. **Band Region**

Band region	Horizontal region (pixels)								
Vertical region (pixels)	141	97	84	93	103	116	146	160	187
	125	**83**	**60**	**69**	**77**	**98**	118	141	168
	116	**80**	**60**	**61**	**77**	**96**	114	136	148
	126	91	**77**	**80**	100	112	121	132	145
	128	102	99	110	123	134	142	153	155

АНАЛИЗ ЛОКАЛЬНЫХ ПОЛОС НА УРОВНЕ ПИКСЕЛЕЙ В МИКРОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ ХРОМОСОМ С ПОМОЩЬЮ CSPA

D. Somasundaram

При изучении хромосом анализ отдельных полос играет основную роль в идентификации хромосом в метафазных пластинках для проведения кариотипирования. Данные литературы ограничены в анализе полос на фотографиях хромосом из-за их сложности. В настоящей работе предлагается алгоритм Conditional Seed Point Algorithm (CSPA) на уровне пикселей. Этот алгоритм моделирования отделяет участки слабых полос от ярких полос, а области ярких полос оцениваются на основе Region of Seed condition Points. Этот алгоритм хорошо работает на различных уровнях интенсивности и адаптирует структурные изменения для идентификации полос на изображении. Он моделирован в более чем 450 индивидуальных хромосомах и обеспечивает уровень точности на 96 %.

REFERENCES

1. Arthur, D.C., and Bloomfield, C.D. Partial deletion of the long arm of chromosome 16 and bone marrow eosinophilia in acute nonlymphocytic leukemia, *Blood*, 1983, vol. 61, no. 5, pp. 994–998.
2. Speicher, M.R., Gwyn Ballard, S., and Ward, D.C. Karyotyping: Human chromosomes by combinatorial multi-flour FISH, *Nat. Genet.*, 1996, vol. 12, no. 4, pp.368–375.
3. Schrock, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Ledbetter, D.H., Bar-Am, I., Soenksen, D., Garini, Y., and Ried, T. Multicolor spectral karyotyping of human chromosomes, *Science*, 1996, vol. 273, pp. 494–497.
4. Adams, R., and Bischof, L. Seeded region growing, *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1994, vol. 16, no. 6, pp. 641–647.
5. Panwar, P., and Gulati, N. Genetic algorithms for image segmentation using active contours, *J. Global Res. Comp. Sci.*, 2013, vol. 4, no. 1, pp. 34–37.
6. Karvelis, P.S., Fotiadis, D.I., Georgiou, I., and Syrrou, M. A watershed based segmentation method for multispectral chromosome images classification, *Proc. 28 IEEE EMBS Ann. Int. Conf.*, New York City, 2006, pp. 3009–3012.
7. Małyszko, D., and Wierzchoń, S.T. Standard and genetic k-means clustering techniques in image segmentation, *IEEE 6th Int. Conf. on Computer Information Systems and Industrial Management Applications (CISIM'07)*, 2007, 0-7695-2894-5/07.
8. Carothers, A. and Piper, J. Computer-aided classification of human chromosomes : A review, *Statistics and Computing*, 1994, vol. 4, no. 3, pp. 161–171.
9. Ledley, R.S., Ing, P.S., and Lubs, H.A. Human chromosome classification using discriminant analysis and Bayesian probability, *Comput. Biol. Med.*, 1980, vol. 10, no. 4, pp. 209–219.
10. Lerner, B. Toward a completely automatic neural-network based human chromosome analysis, *IEEE Trans. Systems Man and Cybernetics*, 1998, vol. 28, pp. 544–552.
11. Yang, X., Wen, D., Cui, Y., Cao, X., Lacny, J., and Tseng, C. Computer based Karyotyping, *IEEE Third Int. Conf. on Digital Soc.*, 2009, 978-0-7695-3526-5/09.
12. Groen, F.C.A., ten Kate, T.K., Smeulders, A.W.M. and Young, I.T. Human chromosome classification based on local band descriptors, *Pattern Recogn. Lett.*, 1989, vol. 9, no. 3, pp. 211–222.
13. Kao, J., Chuang, J. and Wang, T. Chromosome classification based on the band profile similarity along approximate medial axis, *Pattern Recogn.*, 2008, vol. 41, no. 1, pp. 77–89.

Received 29.09.14