SOUTHEAST EUROPE JOURNAL OF SOFT COMPUTING
Available online at www.scjournal.com.ba

# Extracting Gray Level Profiles of Human Chromosomes by Curve Fitting

Sadina Gagula-Palalic[a], Mehmet Can[b]

[a] Faculty of Engineering, International University of Sarajevo,
Hrasnička Cesta 15, 71200 Sarajevo
Bosnia and Herzegovina,
sadina@ ius.edu.ba

[b] Faculty of Engineering, International University of Sarajevo,
Hrasnička Cesta 15, 71200 Sarajevo
Bosnia and Herzegovina,
mcan@ius.edu.ba

***Abstract*— In this paper, a unified algorithm for extracting gray level profiles of Human chromosomes is presented. It is a unified approach since we do not discriminate chromosomes as straight and bended. This is a very helpful procedure which extends the domain of success of most of the previously reported algorithms to highly curved chromosomes. The gray image of the chromosome is thresholded at the gray level 0.9, and the matrix of gray image is transformed into a list of pixel coordinates whose gray level is less than 0.9. To the list of two dimensional points, the most appropriate smooth curve is fitted. Then this smooth curve subdivided into *n* arcs of equal lengths, and straight lines are drawn that are normal to the curve at the end points of the subdivision. The points of the list are classified into *n* bins according to their distance to these *n* straight lines. The average of gray levels of each bin gives the gray levels at the points of the gray level profile of the chromosome. It is seen that the gray level profiles of the bended chromosomes have a high similarity with the straight counterparts.**

**Keywords— Chromosome gray level profiles, Feature extraction, bended chromosomes, thresholding, curve fitting.**

## INTRODUCTION

To detect genetic abnormalities, and fatal diseases like leukemia, karyotyping human chromosomes is a standard tool in today's medicine (Hong, and Mark 2000). Karyotyping starts by segmentation, this consists of picking up 23 pairs of chromosomes from the cell nucleus picture in metaphase stage. Second stage is the extraction of features for classification. The most important features are obtained from the gray level profiles of chromosomes (Piper, and Granum 1989). Then segmented chromosomes are classified into 23 types. Although there are devices and computer softwares to automate the process, still it is done manually by human experts in laboratories (Neurath et. al. (1972).

Usually cell nuclei in metaphase stage are photographed under a light microscope as seen in Figure 1. In metaphase stage, the chromatin is condensed inside the chromosomes making their bands to be easily observed with a light microscope (Lerner et al 1995). Bands on the chromosomes are clearly distinguishable from their neighbors by their darker or brighter appearances. Each of 23 chromosomes has specific band patterns of its own (Lerner 1998).

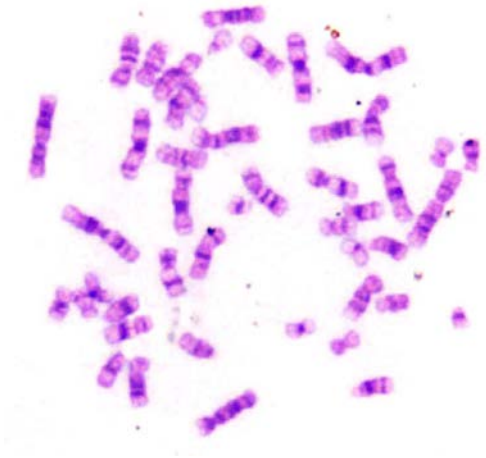

Figure 1. Cell nucleus in metaphase stage

In Figure 2 the result of the karyotyping process of a human cell is presented.

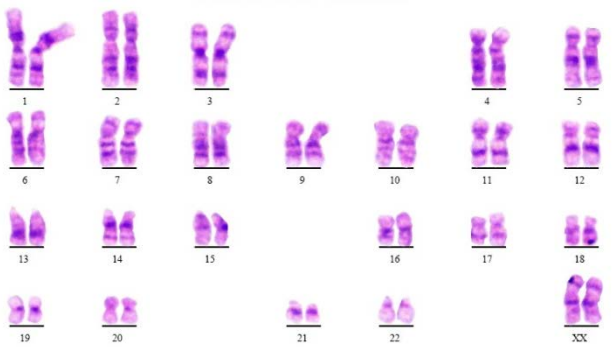


Figure 2. The result of the karyotyping process

## FEATURE EXTRACTION AND SELECTION

To developing a computer-assisted classification system it is important to choose among all possible extracted features, the best subspace that preserves class separability as much as possible in the lowest possible dimensional space. An optimal and small feature set is one of the important factors determining the performance and robustness of a classification system.

Major research effort have been spending into defining and searching for optimal features that can be extracted from chromosome images. Small size and limited resolution of banded chromosomes make finding effective features from original images quite difficult,.

As a result, certain kinds of pre-processing procedures, such as image processing and transformation, are often used to enhance chromosome image features or generate new types of transformed image features.

Lengths of chromosomes in one cell are considered one of the most important features to classify chromosomes. Although chromosome lengths gradually decline from type 1 to 22, there are overlapping in sizes. That is a type 16 chromosome may be longer than a type 14 chromosome.

Subtle variations among cells, such as preparation technique and image quality, can affect the computational accuracy of chromosome length (Stanley et al. 1998). In order to accurately detect the length of a chromosome, some scientists apply medial axis transformation to extract and protect the skeleton of a chromosome.

Using this transformation and the thinning algorithm, computer-assisted schemes can iteratively delete edge points of a region subject to constraints. This process does not remove end points, does not break connectedness and does not cause excessive erosion of the region. Then two length related morphological features, relative length, the ratio of the length of the i-th chromosome to the total length of all 46 chromosomes in one cell, and centromeric index, the ratio of the length of the short arm to the whole length of a chromosome, are computed (Ryu et al 2001). These two features provide a significant amount of chromosome delimiting capability (Stanley et al. 1998, Stanley 1998).

Since each of the 24 chromosome classes possesses unique banding patterns, computing these banding patterns or corresponding features attracted extensive research interests. One of the simplest approaches to represent the banding pattern is using a density profile that computes the mean grey levels along perpendicular lines to the medial axis of a chromosome (Piper, and Granum 1989). Studies have demonstrated that an automated karyotyping system relying primarily on the number of bands and their features could be useful tools in classification of chromosomes (Zimmerman 1986).

Currently, from the density profile as many as 100 feature data can be sampled and extracted (Ryu et al 2004). Because many of these features are redundant, it is required to compress the feature data with certain feature selection techniques, such as knock-out algorithm (Lerner et al 1994), and principal component analysis PCA. A study demonstrated

that the optimal performance could be achieved using a vector with only 10 features computed from the density profile (Ruan 2000) .

Surface features refer to the vector representation of a group of pixel value based image statistics, such as intensity, localized mean, and variance of a particular pixel location in the image, that have been tested in the classification of chromosomes.

In an attempt to further improve classification performance, local energy features which is based on physiological evidence suggesting that human visual system responds strongly to points in an image where phase information is highly ordered (Morrone, and Burr 1988). The local energy can be computed via a set of wavelet transform (Pudney et al 1996). A study demonstrated that combining intensity and local energy based surface features improved the performance of a Kohonen's self-organizing feature map (SOFM) in classification of chromosomes (Kyan et al. 1999).

In addition to the features computed in the space domain, features in the frequency domain have also been explored to classify banded chromosomes. One study investigated and compared features extracted from wavelet and Fourier descriptors in chromosome classification (Sweeney et al 1997). After computing density profile of each chromosome, the discrete wavelet transform and discrete Fourier transform were applied to the density profile (Qiang, and Castleman, 2000). Then, the transformed densitometry signals were equally sampled and used as analytic features (Guimaraes et al 2003). Testing results demonstrated that using Fourier transform based features could achieve higher accuracy compared with using wavelet transform based features (Sweeney et al 1997).

## GRAY LEVEL PROFILE

To extract the gray level profiles of chromosomes, fist we transform the gray picture into a list of points by thresholding. A threshold of 0.9 is chosen, and the coordinates of pixels darker than the gray level 0.9 are recorded by a search algorithm.

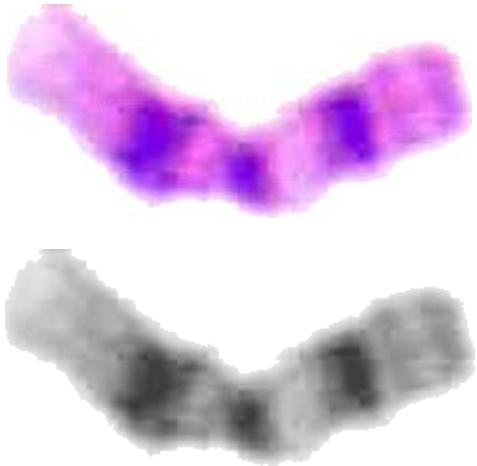

Figure 3. Segmented chromosome and its grayscale version

Search is started at an interior point, called center, of the gray picture. Algorithm first searches eight pixels surrounding the center. Picks the coordinates of the pixels darker than the gray level 0.9, then passes to one of the pixels recently recorded as dark enough, and repeat the same procedure to the neighbors which are not visited before. If none of the new neighbors are not darker than 0.9, search ends.

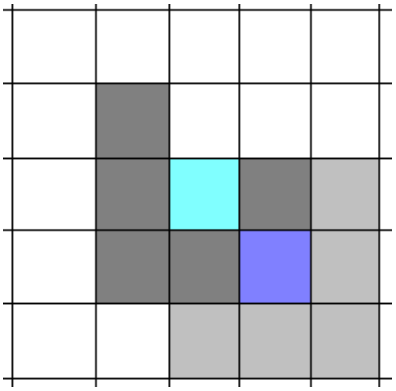

Figure 4. Coordinates of pixels darker than the gray level 0.9 are recorded.

The record of coordinates of pixels darker than the gray level 0.9 is a list of points. The plot of this list gives the profile of the chromosome, where the pixels are darker than the threshold 0.9.

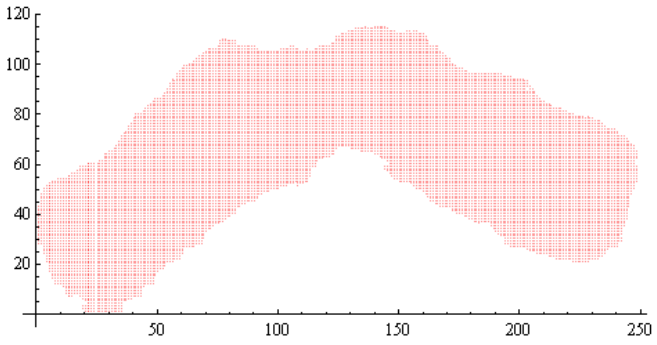


Figure 4. The plot of the pixels of the chromosome, those are darker than the threshold 0.9.

*Skeleton of the Chromosome by Curve Fitting*

For the above bended chromosome, it is seen that a downward parabola is a good fit. A usual curve fitting program fits the parabola

$$y = 14.2965 + 0.999743\,x - 0.00372539\,x^2$$

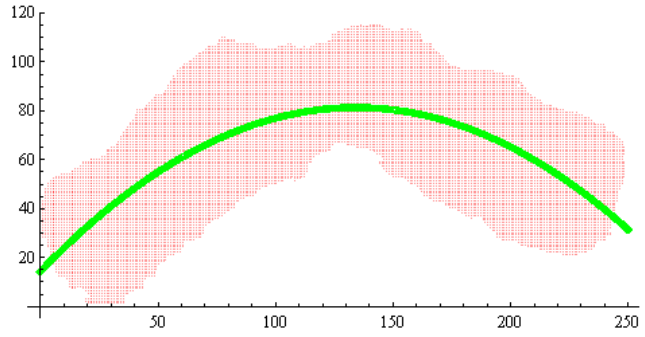


Figure 5.Parabola fitted the list of points

To slice the chromosome into *n* ribbons orthogonal to this smooth curve, the curve is subdivided into *n* arcs of equal length, and straight lines L are drawn that are normal to the curve at the end points of these subdivision points.

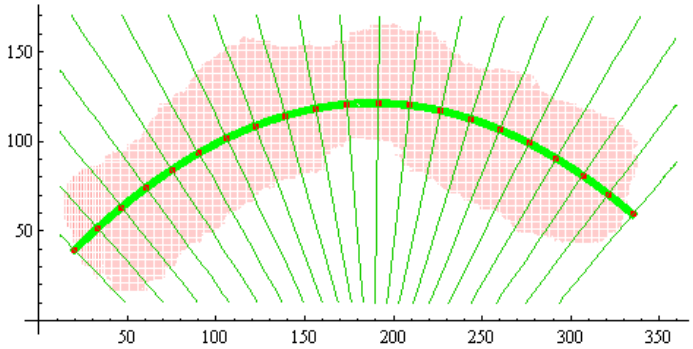


Figure 6. Normals to the parabola fitted

## *The Bins of Points Nearest To Normals*

Algorithm calculates the distance of a point in the list L to all normals, and put this point in a bin of points that are nearest to this normal. This procedure is applied to all points of the list L, and L is partitioned into n bins.

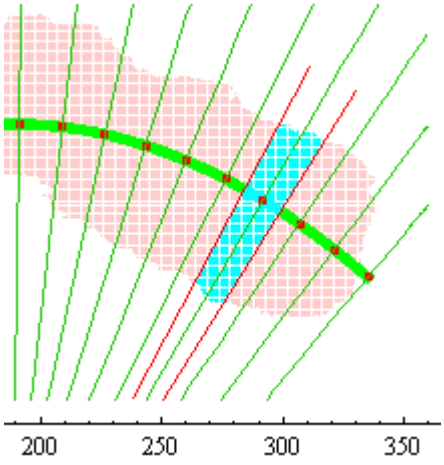


Figure 7. Bin of points that are nearest to this chosen normal

## *Gray Level Profile of the Chromosome*

In the bins there are the pixel coordinates of the chromosome points. The gray levels of the chromosome points in bins are called back using the addresses kept in coordinates, and gray levels of points in bins are averaged to give the gray level profile of the chromosome.

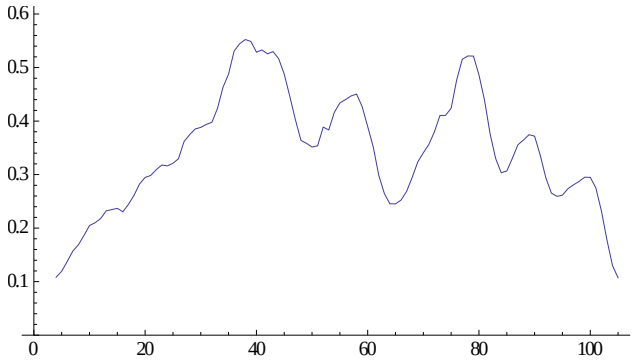


Figure 8. Gray level profile of the chromosome chosen

To see the success of extraction of the gray level profile of the bended chromosome, we compare it with the gray level profile of the other straight pair of the chosen chromosome.

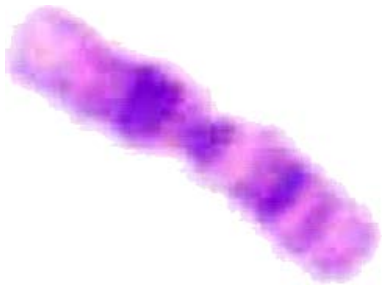

Figure 9. Straight pair of the chosen chromosome and its grayscale version

Straight line

$$y = 14.0191 + 0.60213\,x$$

is a good curve fit for this straight chromosome.

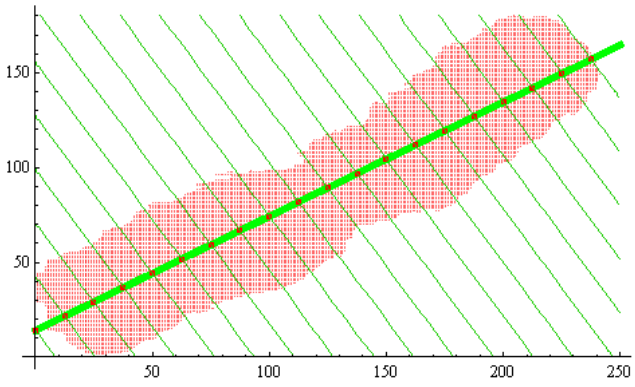


Figure 10. Straight line is a good fit for the straight pair chromosome chosen

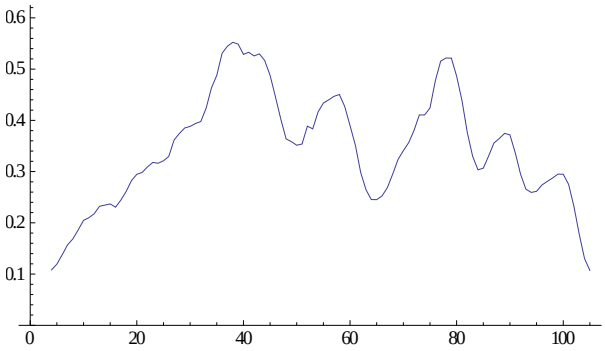


Figure 11. Gray level profile of the straight pair of the chromosome chosen

The comparison of the gray level profile in Figure 8 of the bended chromosome, and gray level profile of its straight pair in Figure 11 proves the success of the method of curve fitting.

*Extra Bended Chromosomes*

The chromosome in Figure 3 is bended further using Microsoft Paint.

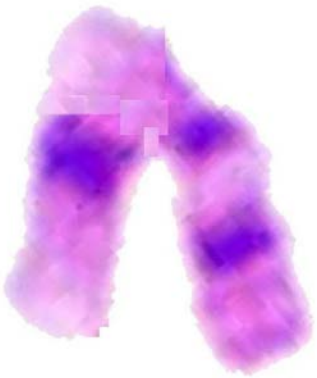

Figure 12. Extra bended chromosome

A usual curve fitting program fits the parabola

$$y = 94.3906 + 2.29016\,x - 0.005535\,x^2$$

to the data points.

To slice the chromosome into *n* ribbons orthogonal to this smooth curve, the curve is subdivided into *n* arcs of equal length, and straight lines are drawn that are normal to the curve at the end points of these subdivision points as before.

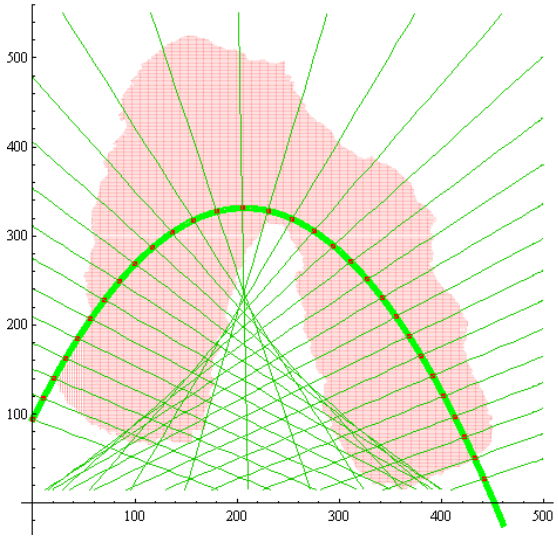


Figure 13. Normals to the parabola fitted

Algorithm calculates the distance of a point in the list L to all normals, and put this point in a bin of points that are nearest to this normal. This procedure is applied to all points of the list L, and L is partitioned into n bins. In the bins there are the pixel coordinates of the chromosome points. The gray levels of the chromosome points in bins are called back using the addresses kept in coordinates, and gray levels of points in bins are averaged to give the gray level profile of the chromosome.

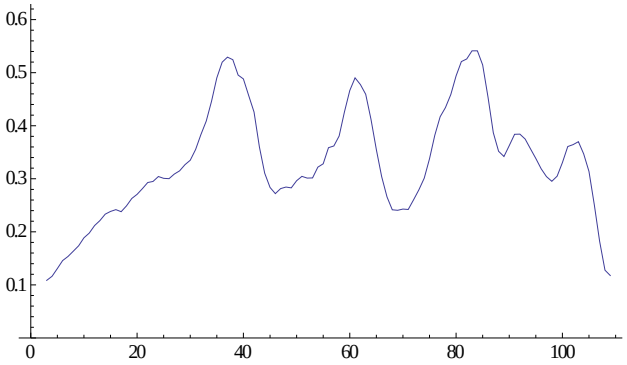


Figure 14. Gray level profile of the extra bended chromosome

The comparison of the gray level profile in Figure 14 of the extra bended chromosome in Figure 12 with the gray level profile in Figure 8 of the bended chromosome in Figure 3, and gray level profile of its straight pair in Figure 11 proves the success of the method of curve fitting.

## CONCLUSION

Papers dealing with the feature extraction for bended human chromosomes mostly exaggerate the difficulty posing by bended chromosomes (Roshtkhari, and Setarehdan 2007,

Barrett, and de-Carvalho 2003). In this paper a simple, yet very effective algorithm for obtaining gray level profiles bended human chromosomes is presented. And it has been shown that gray profiles of chromosomes do not have sensitive dependence on bending.

All previously reported methods for automatic Karyotyping can benefit from the proposed algorithm without any discrimination between straight and bended chromosomes (Ji 1994, Carothers, and Piper 1994, Moradi, and Setarehdan 2006).

## REFERENCES


Barrett SD, and de-Carvalho CR (2003) A software tool to straighten curved chromosome images", Chromosome Research, No. 11, pp. 83-88.

Carothers A, and Piper J (1994) Computer-Aided Classification of Human Chromosomes: A Review", J. Statistics and Computing, Vol. 4, No. 3, 1994, pp. l6l-l7l.

Guimaraes LV, Schuck A, Elbern A (2003) Chromosome classification for karyotype composing applying shape representation on Wavelet packet transform. In: Proc. 25th Annual Internat. Conf. of the IEEE EMBS, pp. 941–943.

Hong F, and Mark L (2000) Medical Cytogenetics, Marcel Dekker, pp. 21-49.

Ji L (1994) Fully Automatic Chromosome Segmentation Cytometry 17: pp. 196 -208

Kyan MJ, Guan L, Amison MR, and Cogswell CJ (1999) Feature extraction of chromosomes from 3D confocal microscope images," presented at 1999 International Conference on Image Processing.

Lerner B, Guterman H, Dinstein I, and Romem Y (1995) Medial axis transform-based features and a neural network for human chromosome classification", Pattern Recognition, Vol. 28, No. 11, pp. 1673-1683.

Lerner B (1998) Towards a Completely Automatic Neural Network-Based Human Chromosome Analysis", IEEE Trans. On Sys., Man. and Cybernetics, Vol. 28, No.4, pp. 544-552.

Lerner B, Levinstein M, Rosenberg B, and Guterman H (1994) Feature selection and chromosome classification using a multilayer perceptron neural network, Neural Networks," presented at IEEE International Conference on Computational Intelligence.

Moradi M, and Setarehdan SK (2006) New features for automatic classification of human chromosomes: A feasibility study", Pattern Recognition Letters, Vol. 27, pp. 19-28.

Morrone MC,and Burr DC (1988) Feature detection in human vision: A phase dependent energy model," presented at Proceedings of the Royal Society, London.

Neurath PW, Lin PS, and Low D (1972) Karyotyping: a model twin study. Cytogenetics 11 pp: 457-474

Piper J, and Granum E (1989) On Fully Automatic Feature Measurement for Banded Chromosome Classification," J. Cytometry, No. 10, pp. 242-255.

Pudney C, Robins V, Robins B, and Kovesi P (1996) Surface detection in 3D confocal microscope images via local energy and ridge tracing," presented at J Computer-Assisted Microscope.

Qiang, W., Castleman, K.R., (2000) Automated chromosome classifier using wavelet-based descriptors. In: Proc. IEEE Symp. on Computer Based Medical Systems, pp. 189–194.

Roshtkhari MJ, and Setarehdan SK (2007) A New Approach to Automatic Classification of the Curved Chromosomes, Proceedings of the 5th International Symposium on image and Signal Processing and Analysis, pp. 19-24.

Ruan X A (2000) A classifier with the fuzzy Hopfield network for human chromosomes, Intelligent Control and Automation," presented at Proceedings of the 3rd World Congress on Intelligent Control and Automation.

Ryu SY, Cho JM, and Woo SH (2001) A study for the feature selection to identify giemsa-stained human chromosomes based on artificial neural network," presented at Proceedings of the 23rd Annual International Conference of the IEEE, Istanbul, Turkey.

Ryu SY, Cho JM, and Woo SH (2004) A study for the feature selection to identify giemsa-stained human chromosomes based on artificial neural network, presented at Proc. 23rd Annual EMBS Int. Conf., San Francisco, CA, USA, Sept. 2004.

Stanley RJ, Keller JM, Gader P, and Caldwell CW (1998)Data-driven homologue matching for chromosome identification," IEEE Trans Med Imaging, vol. 17, pp. 451-462.

Stanley RJ (1998) Data-driven homologue matching for chromosome identification," vol. Ph.D.: University of Missouri – Columbia.

Sweeney N, Becker RL, and Sweeney B (1997) A comparison of wavelet and Fourier descriptors for a neural network chromosome classifier," presented at Proceedings of 19th International Conference – IEEE / EMBS, Chicago IL.

Zimmerman SO, Johnston DA, Arrighi SA, and Rupp ME (1986) Automated homologue matching of human G-banded chromosomes," Comput. Biol. Med., vol. 16, pp. 223–233.