Trans. Nat. Acad. Sci.
Technol. 14: 231-242. 1992

COMPUTER-BASED METAPHASE-FINDING AND KARYOTYPING OF HUMAN CHROMOSOMES

Maricor Soriano and Caesar Saloma
Instrumentation Physics Laboratory, National Institute of Physics
University of the Philippines, Diliman 1101, Q.C.

ABSTRACT

Preliminary results of a simpler approach to the metaphase-finding and karyotyping of human chromosomes are presented. By reducing what initially was a two-dimensional (2D) pattern recognition problem into an effective 1-D task, the complexity of the numerical processing involved was reduced. This simplification made possible the numerical task of metaphase-finding and karyotyping using a personal computer in a reasonable amount of time.

The simplification resulted from 1) the use of an appropriate image magnification scheme in metaphase-finding, 2) an innovative implementation of the centromere-finding algorithm originally proposed by Gallus and Mendelssohn, and 3) the more efficient application of the three parameters (size, centromere location, and band patterns) in classifying human chromosomes.

Introduction

The ordered classification (karyotyping) of human chromosomes into 23 pairs from a given metaphase is an important pattern-recognition problem. The normal procedure of karyotyping is a very tedious job and often requires the services of trained personnel to image the metaphases of a given sample, develop the photographic print of the microscope image, and via a 'cut and paste' procedure arrange the chromosomes according to size, centromere location, and band patterns.

The type of chromosome aberration may reveal the kind of disorder present in a patient. The disorder may be Mongolism, sterility, cancer or mutation (9). Beyond diagnosis, gene manipulation to prevent or correct the occurrence of the mentioned disorders sans the harmful side-effects may become practical if simulated surgeries of chromosomes are first done in the computer.

Our work achieves metaphase-finding and karyotyping of chromosomes using a less complex but equally reliable pattern recognition algorithm. The complexity of the recognition problem is simplified through the use of two-dimensional detection and data transport, appropriate use of image magnifications, and

efficient use of the classification parameters. With our technique, the process of karyotyping no longer needs a trained technician to implement it.

Hopefully, the results of our research will enable cytogeneticists in our hospitals to do rapid karyotyping and analysis of chromosomes using a standard optical microscope that is fitted with a 2-D CCD camera, image digitizer, and a personal computer, instead of the very expensive workstations that are in the market [e.g. 1, 6, D. Lloyd, et al.].

Background

The production of a karyotype starts with the efficent location of a good metaphase spread in a microscopic slide containing a number of treated cells. A cell is treated to induce its division. A perfect metaphase for pattern recognition is one where the chromosomes formed from protein-DNA complexes in a cell's nucleus (2), are well-spread and do not overlap in space.

Three features may be used to classify a human chromosome [7]: 1) its relative size, 2) the location of its centromere, and 3) characteristics of its band patterns. In a human karyotype, the 46 chromosomes are arranged in decreasing order of size or length. Based on the location of the centromere, three groupings of chromosomes can be generated: a) metacentric, b) submetacentric, and c) acrocentric (see Figure 1). Using Giemsa dye to stain chromosomes, patterns that are unique to every chromosome pair can be realized. Comparison by band patterns alone, is enough to generate all the 23 distinct chromosome pairs. Table 1 enumerates the standard groupings in a human karyotype.

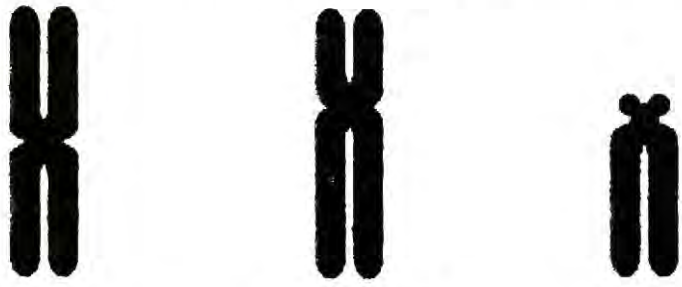


Figure 1. Three types of human chromosomes according to centromere location

To carry out the numerical processing, a 512 x 492 CCD camera (Hamamatsu Photonics C3057, pixel size: 17 x 13 μm) records the microscopic images of the metaphases while a frame storeboard (Tektronix DCS01, 512 by 512 elements, 127 gray scales) digitizes the captured image for storage and numerical manipulation in a personal computer (IBM PC486-compatible). All computer programs were written in Turbo Pascal (Version 6.0).

Images used for metaphase-finding are first background-corrected to cancel the effect of uneven illumination which compromises image contrast [10]. The image signal i(m,n) is a product of the space-dependent illumination function b(m, n) and the space-dependent sensor sensitivity e(m,n):

$$i(m,n) = t(m,n)b(m,n)e(m,n) \tag{1}$$

where the t(m, n) is the true object information function to be recovered. The effect of the product b(m,n)e(m,n) in Eq. (1) can be corrected by noting that a blank image $i_b(m, n)$ which corresponds to imaging without a sample placed in the object space of the microscope can be expressed as:

$$i_b(m,n) = C_1 b(m,n)e(m,n) \tag{2a}$$

where C_1 is an associated gain factor, and (m,n) signifies pixel element location in x-y space. Taking the ratio of i/i_b yields an image function that is proportional to t(m,n):

$$t_m(m,n) = \frac{C_1}{C_2} t(m,n) \tag{2b}$$

where C_2 is a gain factor used to adjust the degree of contrast.

Samples (white blood cells, mitogen: phytohemagglutten, arresting agent: colchicine) used in the experiments were supplied by the Immunology Department, Philippine General Hospital. The chromosomes were fixed on the slide using Carnoy's fixer (methanol:acetic acid, 3:1).

FORMULATION OF ALGORITHM

A. *Metaphase-finding*

Under low-level magnification (4x objective), a good metaphase spread appears as a circular spread (see Figure 2) in image space whose radius is larger than that of an unburst nucleus (see Figure 3). This property simplifies metaphase-finding into an 1-D recognition problem.

To identify a metaphase from an unburst nucleus, 1-D template matching [5, 8] is used. The presence of a pattern in a signal can be ascertained by searching for a location of match between a pattern template p(x) of finite length, and a raw data series t(x). The correlation between p(x) and t(x) is given by

$$C_{p,t}(x) = \sum_x p(x)t(x+u) < \left[\sum_x p^2(x) \sum_x t^2(x+u)\right]^{1/2} \tag{3}$$

In metaphase-finding, the normalized correlation is used to take into account the finite length of the template,

$$\text{normalized correlation} = \frac{C_{p,t}}{\left[\sum_{u}^{u=L} t^2(x+u)\right]^{1/2}} \qquad (4)$$

The normalized correlation becomes maximum when $t(x) = \text{const. } p(x)$. The normalized correlation given Eq. (4) produces a narrower correlation function than $C_{p,t}$.

The reduction of the metaphase-finding into an equivalent I-D problem brings about a significant decrease in the number of multiplications involved for every pixel x, from m^2 to m for a template of length $L = m$ pixels. Unlike the procedure proposed by 4 our procedure minimizes mechanical scanning because fields of view are larger at low magnifications

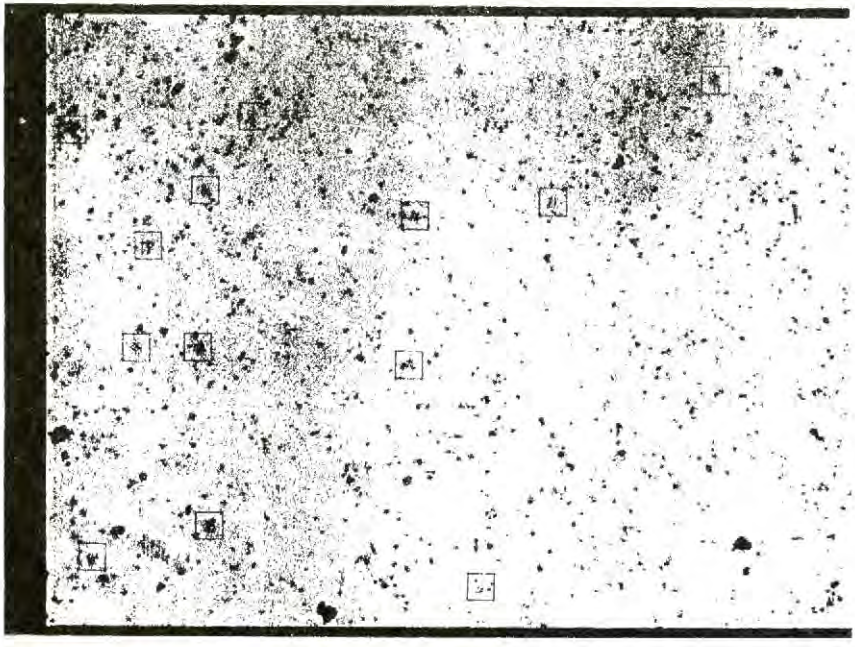


Figure 2. Image of spreads (boxed) under 4x magnification. The sharp dots are unburst nuclei. The bigger spots are accumulated dust particles.

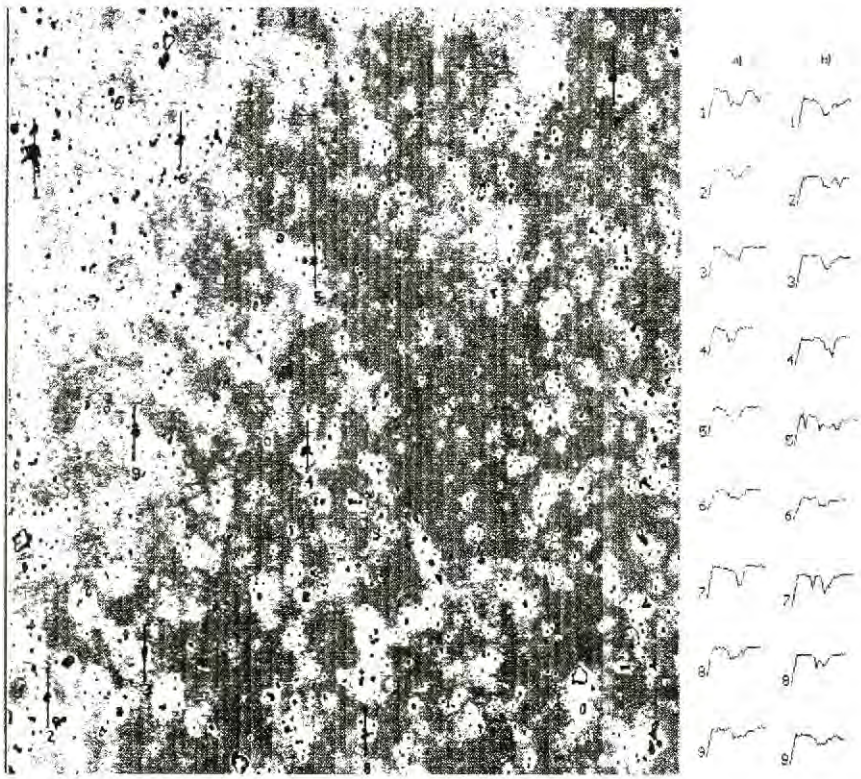


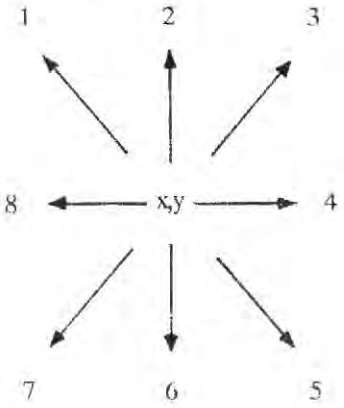
Figure 3. Intensity profile of spreads indicated on the image by numbers and vertical lines. a) vertical cross-section. b) horizontal cross-section.

B. *Karyotyping*

The three classification parameters of relative size, centromeric position, and band pattern signature can be determined only when the 2-D boundary (edge) profile, the major axis-centromeric axis pair of every chromosome in the given metaphase, are known.

In our method, the boundary of each chromosome is traced using Freeman vectors [3]. A Freeman vector is represented by an integer F(i) where i ranges from 1 to 8, such that the angle between a pair of successive vectors (clockwise) is 1/4n (see Figure 4). Thus the 2-D edge profile of chromosome can be represented as a closed-looped string of numbers with values varying from 1 to 8.

As the computer looks for the boundaries, it simultaneously computes the area of the chromosome by pixel counting. For every pixel detected, our algorithm counts the number of neighboring pixels on its left which is used to increment the area variable of the given chromosome.



Freeman vector	Freeman code
(x-1, y+1)	1
(x, y+1)	2
(x+1, y+1)	3
(x+1, y)	4
(x+1, y-1)	5
(x, y-1)	6
(x-1, y-1)	7
(x-1, y)	8

Figure 4. Freeman vectors and codes.

The identification of chromosome arms is realized after following a predefined sequence: a) computation of the Freeman vectors from a given chromosome image, b) computation of the vector gradients as the difference from two adjacent Freeman vectors, and c) computation of the running sum of three gradients. The resulting string is one that has positive values for convexities and negative values for concavities. Chromosome arms are identified by looking for major positive-valued convexities that are delimited by two O's or negative numbers. The major axis spans the longest side of the chromosome, while the minimum distance perpendicular to the major axis defines the centromeric axis which contains the centromere. Figure 5 illustrates how the sequence just discussed is implemented.

PERFORMANCE

A. *Metaphase-finding*

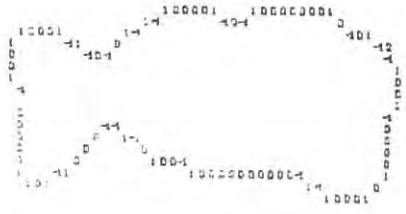
In the experiment, the template (m = 15 pixels) was made from the average of ten good metaphase profiles. Figure 6 shows the correlation results using our algorithm. Note that a correlation peak is obtained in the region containing good metaphase spread.

B. *Karyotyping*

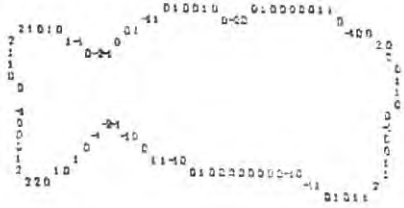
In some preparations, a number of chromosomes overlap in space. A chromosome separation routine was developed to allow the user to separate chromosomes via the mouse before performing boundary tracing.

a) Original image

b) Freeman vectors



c) Vector gradients



d) Running sum of three gradients

e) Localization of chromosome arms

Figure 5. Boundary analysis applied to a chromosome image.

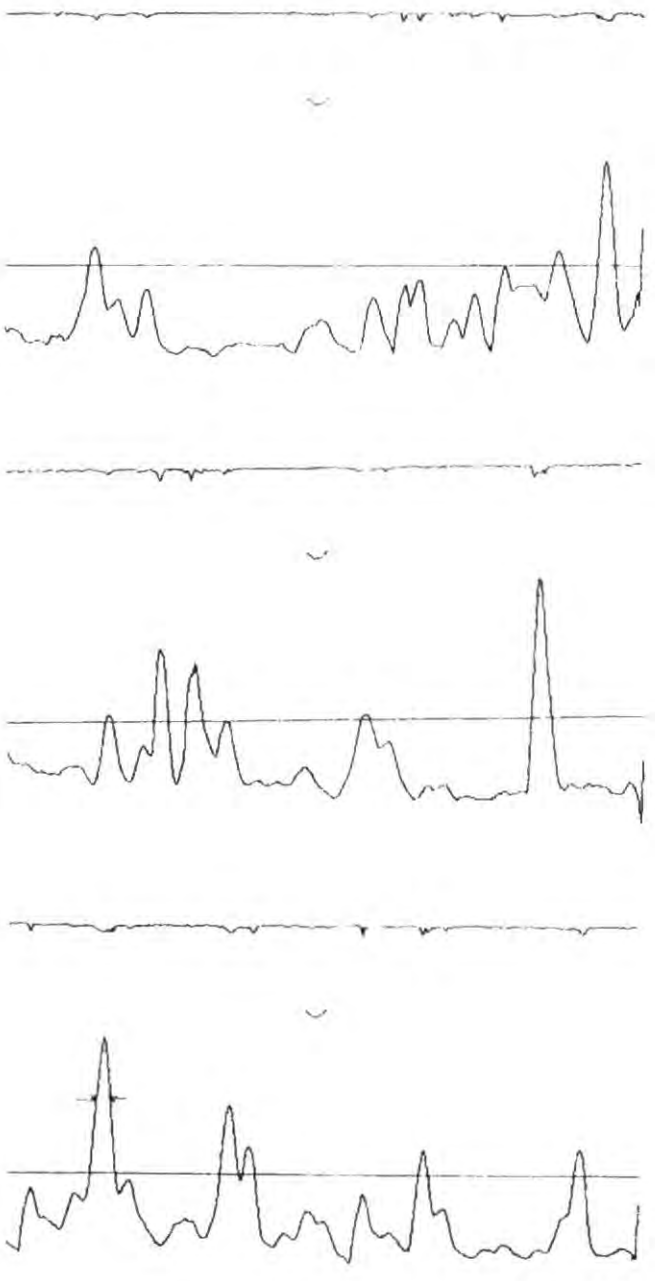


Figure 6. Normalized correlation between image intensity profile and template (shown in the middle).

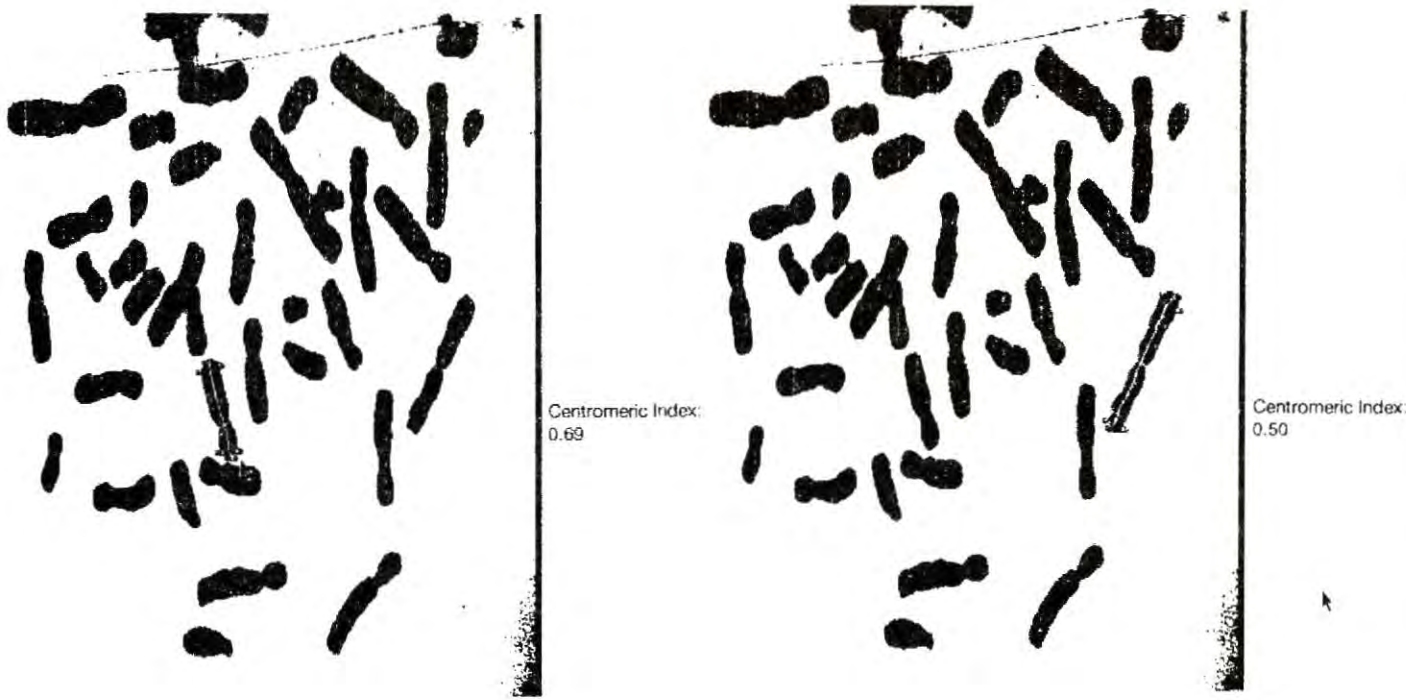


Figure 7. Axis tracing, centomere location and calculation of the centromeric index

Figure 8. Classification of chromosomes according to size: a) original image, and b) reordered image.

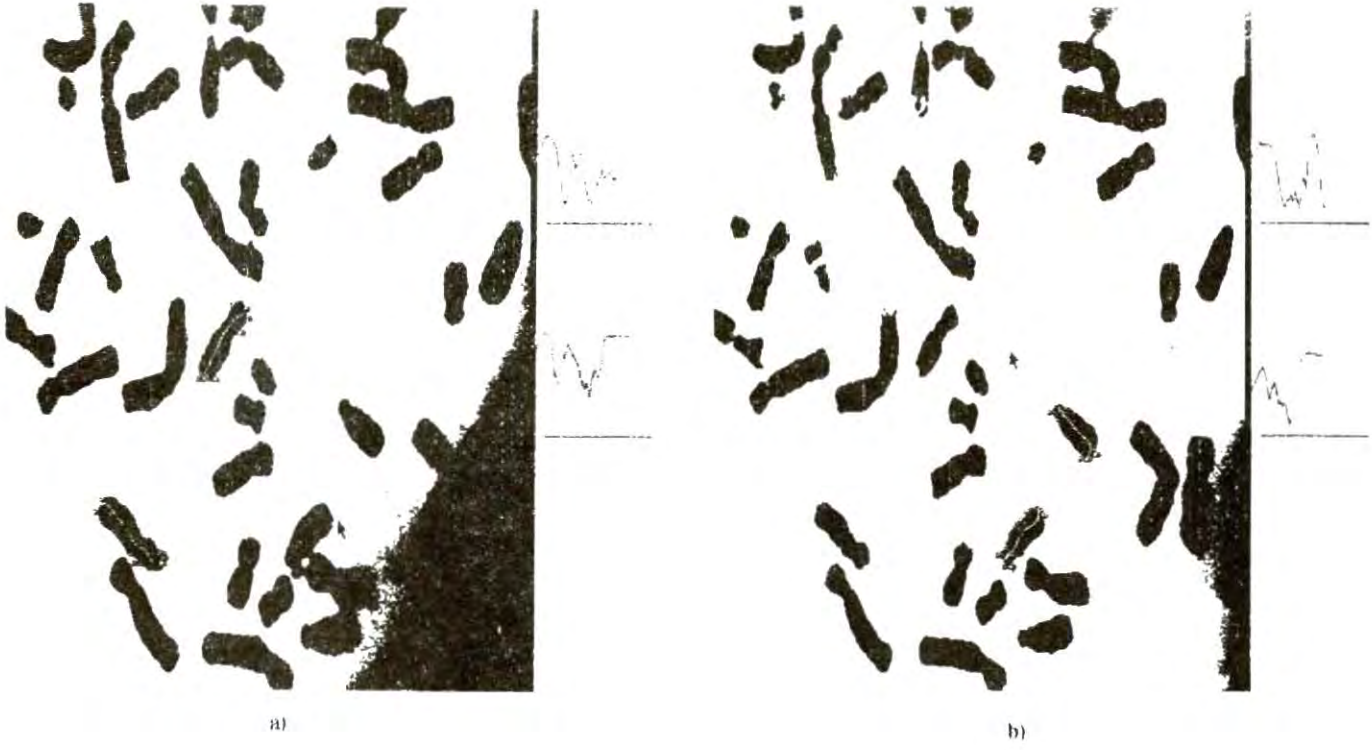


Figure 9. Band profiles of homologous chromosomes. a) Chromosome no. 11. b) Chromosome no. 14. Due to the tracing of the axis, the lower profile in (b) is the reverse of the upper profile.

After the boundary profile of a given chromosome is established using Freeman vectors, the program proceeds to find the chromosome ends, the centromeric line, and the major axis. From these values, the centromeric index is calculated (see Figure 7). The centromeric index is the ratio of the long arm to the entire length of the chromosome.

An ensemble of 76 chromosomes (100x objective magnification) was used to test the success rate of the algorithms discussed. The edge detection algorithm had a 100% success rate. The centromere-finding procedure worked for 61 of the 76 chromosomal images, or an 80% success rate. The common cause for errors was incorrect location of the chromosome arms, accounting for 13 of the 15 centromere location errors. Location errors generally involved the overcounting (lots of convexities caused by diffractions) or undercounting (absence of details) of the number of arms. Figure 8 illustrates how the area parameter and classifies the randomly-arranged of 46 chromosomes in selected metaphase spreads.

Figure 9 shows a (major) axial intensity of a pair of chromosomes. These profiles will be cross-correlated with pre-defined template of band patterns for each chromosome pair. Again the cross-correlation procedure simplifies into an 1-D problem.

Acknowledgment

The authors are grateful to the various scientists of PGH Immunology Laboratory and PNRI Cytogenetics Laboratory, for never falling to provide explanations, microscopic slides, and high-contrast pictures whenever needed.

References

1. **Cytoscan 110,** Medical Research Council of the United Kingdom.
2. **J. Darnell, H. Lodish and D. Baltimore,** Molecular Cell Biology, 2nd Edition (Scientific American Books, New York, 1990), 318.
3. **H. Freeman,** "On the Encoding of Arbitrary Geometric Configurations," IRE Trans. Electron. Comp., 10, 260-268 (1961).
4. **D.K. Green and P. Neurath,** The Design, Operation and evaluation of a High Speed Automatic Metaphase Finder," J. Histochem. Cytochem. 22, 531-535 (1974).
5. **A. Jain,** Fundamenetals of Digital Image Processing, (Prentice-Hall, New York, 1982), 400-402
6. **D. Lloyd, J. Piper, D. Rutovitz and G. Shippey,** "Multiprocessing Interval Processor for Automated Cytogenetics," Appl. Opt. 26, 3356-3366 (1987).
7. **E. Passarge,** The Human Karyotype. Analysis of Chromosomes in Mitosis and Evaluation of Cytogenetic Data (from files of the Philippine General Hospital, Manila, Philippines).
8. **A. Rosenfeld and A. Kak,** Digital Picture Processing, Vol. 2, 2nd Edition (Academic Press, New York, 1982), 37-38.
9. **R. Turpin and J.R. Lejeune,** Human Afflictions and Chromosomal Aberrations (Pergamon Press, New York, 1969).
10. **F. Wahl,** Digital Image Processing (Artech House, New York, 1987), 66-68.