INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9325160

Contour shape analysis with applications to cytology

Wali, Rahman, Ph.D.

City University of New York, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

CONTOUR SHAPE ANALYSIS WITH APPLICATIONS TO CYTOLOGY

by

RAHMAN WALI

A dissertation submitted to the Graduate Faculty in Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1993

This manuscript has been read and accepted for the graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 28, 1993
Date

Professor Joseph Barba
Chair of Examining Committee

4/29/1993
Date

Dean G. Lowen
Executive Officer

Professor Samir Ahmed

Professor Norman Scheinberg

Professor Sanghamitra Basu

Professor Michael Colef

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

CONTOUR SHAPE ANALYSIS WITH APPLICATIONS TO CYTOLOGY

by

Rahman Wali

Advisor: Professor Joseph Barba

Shape analysis is important in classification where deviation of a contour from a model are measured. In pathology, deviation of nuclear contour from an ellipse is sometime useful in cell classification. We present a method based on Fourier series expansion, to determine the best fit ellipse to a given contour by minimizing the difference in the ellipticity factors and keeping the same area for the contour and the matched ellipse. We also compared our method to the Grid-Search method in terms of speed and accuracy. Finally we applied our method to the diagnosis of cancer cells.

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my mentor, Professor Joseph Barba for his interest, encouragement and guidance during this research. The many technical discussions and his constructive suggestions are most appreciated.

I would like to thank all the members of doctoral committee; Dean Gerald Lowen, Professor Samir Ahmed, Professor Norman Scheinberg, Professor Sanghmitra Basu and Professor Michael Colef for their guidance and support. Special thanks to my colleague Robert Bloom at Visual Cybernetics for discussing and sharing ideas about my research.

Many thanks to my parents whose sacrifice and motivation led me to higher education and to them I dedicate this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

Everyday, more than a million blood smear examinations are performed in hospitals in the United States [Ingram,70]. The blood smear examination is a necessary and routine examination for every patient admitted to a hospital. It is one of the most useful indicators of a person's health because the body's production of various type of blood cells is highly dependent on general stress, injuries, diseases, poisons, ionizing radiations and other noxious stimuli [Wintrobe,74].

In a blood smear examination, a suitably prepared blood smear is observed under a microscope. Three things are done during the examination process: 1) count the number of white blood cells and establish the percentage of occurrence of the six normal types of white cell and possibly some of the abnormal types; 2) make a qualitative appraisal of the red blood cells with respect to their sizes, shapes and colors and the notation of any abnormal forms; 3) evaluate platelet sufficiency and check for the presence of abnormal forms. The task performed manually is tedious, time consuming, inaccurate, inefficient

and expensive. The blood smear examination is a perfect target for automation to increase speed, reliability, and accuracy, to reduce cost, and to relieve the laboratory technicians.

The need for modern computers in cytological image analysis and classifications have been well recognized for two decades. Computers can be used not only for the rapid identification of cells for the purpose of diagnostic, screening and classification, but also to aid the health professional in making difficult diagnostic decisions. In cyto-pathology, hematology and histo-pathology, there exist many instances where the discriminatory power of human vision and human visual assessment are limited in making a diagnostic decision. The computer may provide a quantitative tool which will be able to help solve these problems.

The majority of computer image processing in biomedical research is devoted to automated analysis of biological cells. The main reason for placing emphasis upon cell analysis are: 1) cell analysis is a well-defined problem in the biomedical field for which modern computers may be successfully employed to produce useful results; and 2) cell analysis is a process of fundamental importance in biological and medical studies as exemplified by the analysis of blood smear and tissue

sections in making diagnosis and prognosis.

Cell analysis by computer requires 1)Image acquisition, 2)Image segmentation and 3)Image Representation and description as shown in figure[1.1].

IMAGE ACQUISITION

In cytology, video images from a camera mounted on a microscope are digitized, processed and displayed on a video monitor. The overwhelming majority of all television cameras utilize either photoconductivity tubes or solid state sensors. An image is optically focused onto the camera image plate in order to convert light intensity into an electrical current.

The digitizing process is performed by an analog-to-digtal (A/D) converter. Standard monochrome video images, when digitized at or above the Nyquist rate, are generally presented at 512x512 array of 8 bit samples. A color (RGB) image is represented by three such 512x512 arrays, for each color. Lower resolution results from using digitizing rates below the Nyquist rate. A typical digitization rate in such systems is 5.0 MHz, resulting in a 256x256 array of 8 bit samples. This lower resolution, though adequate for many image processing applications, is not well suited for medical imaging applications.

IMAGE SEGMENTATION

Image segmentation is the division of an image into different regions, each having certain similarities (e.g. texture, color, gray level etc.). Classification of cell require quantitation of features such as cytoplasm and nuclear area, nuclear shape, cytoplasm color, etc. The quantitation of these features depends on the segmentation of the scene as to which part is the nucleus, which part the is cytoplasm and which part is irrelevant, such as background. Image Segmentation is a critical component of an image understanding system because errors in segmentation will propagate and influence the feature measurements and consequently the classification performance of the system.

Image segmentation have been used in biomedical applications in the identification of lung diseases [Kruger,1974], in automated classification of white blood cells [Ingram, 1970]; in military applications [Keng, 1976], image segmentations have been applied in the recognition of highway, bridge, aircraft and runways; in commercial applications in the area of optical character recognition, and in industry such as automated visual inspection. It also has applications in other

areas, for example, in communications such as data compression, by transmitting the boundaries of the regions instead of the whole picture, we can reduce significantly the bandwidth required.

Many image segmentation techniques have been proposed [Rosenfeld, 77]. These techniques can be categorized into three classes:

(1) Characteristic feature thresholding or clustering

(2) Edge detection (3) Region extraction

Almost all segmentation algorithms are either based on the concept of similarity (e.g. characteristic feature clustering algorithms) or discontinuity (e.g. edge detection algorithms). Despite the large amount of research efforts devoted to image segmentation algorithms, very little is known about how to measure segmentation error besides the simple criteria of the percentage of pixels misclassified [Yasnoff, 77]. As a consequence, it is still very difficult to answer the question "how good is a given algorithm?". Thus, it is difficult to compare different image segmentation algorithms. Further compounding the evaluation process, different authors generally use different data and few authors process more than several hundred images. Unless one specifically implements a given

segmentation algorithm and tries it out on one's data, it is very difficult to evaluate from the published results how well it will work for a given set of data. For these reasons, except in very special cases, the author will not comment exactly how well a given algorithm will work although qualitative statements on the advantages and disadvantages of the approach can be made. For more in-depth study on segmentations the reader is referred to [Canny, 1986, Parent, 1989, Fung, 1988, Bouman, 1988].

REPRESENTATION AND DESCRIPTION

After an image has been segmented the regions have to be described for further computer processing. Basically we have two choices for representing a region: 1) we can represent the region based on its external characteristics (i.e., its boundary), or 2) we may choose to represent it in terms of its internal characteristics (i.e., the pixels comprising the region). The next task is to *describe* the region based on the chosen representation. For example, we may elect to represent a region by its boundary, and to describe boundary by features such as its length, the orientation of the straight line joining the extreme points, and the number of concativity in the boundary etc.

Generally, an external representation is chosen when the primary focus is on *shape characteristics (also called* morphological features), while internal representation is selected when one is interested in reflective properties, such as *color* and *texture*. In either case, it is important that the features selected as descriptors be insensitive to variations as *size, translation and rotation*. In this study, we will focus on external characteristics (shape) with application to the description of biological cells. We are given cell contour and will extract the shape features suitable for classification.

Feature extraction can be considered the data phase of computer vision. It forms the link between image higher level processing such as pattern recognition and the raw acquired image. Whatever the general scheme of the vision system may be, it is built upon some basic and fundamental measurements of the digital image. Features provide numerical values that represent the spectral and or spatial relationship of the pixels within a digital image.

Numerous features exist for quantifying images and the number of possible features for a given situation is limited only by ones creativity and the desired complexity of the system. Some of the basic features are: Area; perimeter; width;

length; principal axis moment; central moment; lobing; minimum lobe tip angle; average lobe tip angle; number lobes; minimum distance between lobes; maximum distance between lobes; complexity; elongation. These features represent basic shape descriptors. Each feature represents a quantitative measurement on digital image and is the first step in extracting objective information about the image.

In Cytology, Cell classification is performed by a trained pathologist by visual inspection of specimen through a light microscope, and employing features he/she subjectively learned from years of experience. These include cell and nucleus size, nucleus shape and texture. In the majority of cases, the diagnosis is simple and can be performed quickly, however in some cases the diagnosis is difficult, and different pathologist can reach different conclusions. In these cases quantitative image analysis of the cell and nucleus size, shape and texture can be useful in the diagnosis. At the Mount Sinai Medical Center of New York we are currently developing algorithms to quantify image features which are suitable for implementation in an interactive environment.

SUMMARY

The objective of this research was to develop shape analysis technique capable of resolving variations in contour shape with specific applications to the quantitative description of biological cell and nucleus shape. Our research work is summarized as following:

- Initially we attempted to use fractal for analysis of contours shape. We use Iterative Function System commonly known as IFS to describe contour shape. IFS is description of a curve by the combination of one or more affine transformations. We abandoned the line of research when we come to the conclusion that there is no guarantee that different contours will generate the same number of affine transformations, making it difficult to discriminate between contours by comparing their affine transformations.
- We implemented the Grid Search which is commonly used to match an ellipse to the contour. The ellipse parameters can be used to discriminate between the contours. We applied this method to contours data showing that the grid search method is very slow due to the interdependent of ellipse parameters theirby making the optimization procedure very slow. We also

demonstrated that the method gives erroneous results for certain irregular contours and therefore can not be used in a clinical environment.

- To overcome the drawbacks (i.e. Speed and Accuracy) of the Grid Search method we developed an algorithm based on the Fourier series to match an ellipse to a contour in terms of area and shape. We show that the Fourier method determines the ellipse parameters independently of each other and therefore is approximately thirty time faster than the Grid Search method. We also show that our method always converges to a global minima, even for very irregular contours. We show that the geometric mean should be used for the contour center instead of the conventional centroid. The method is also shown to be applicable to the multi valued contours.

- Shape analysis can be facilitated by calculating contour descriptors such as area, ellipticity and the contour deviation from ellipse by using the match ellipse parameters. We applied our Fourier method to the diagnosis of cancer cells. The samples were obtained from the files of Mount Sinai Medical Center of New York at random. Application incorporating our algorithm include *Borderline* and *Invasive* tumors of Overy, and *cervical condyloma*.

We show that desciptors such as area, ellipticity and the contour deviation from ellipse highly discriminate between the tumors and the results were in agreement with the pathologist findings.

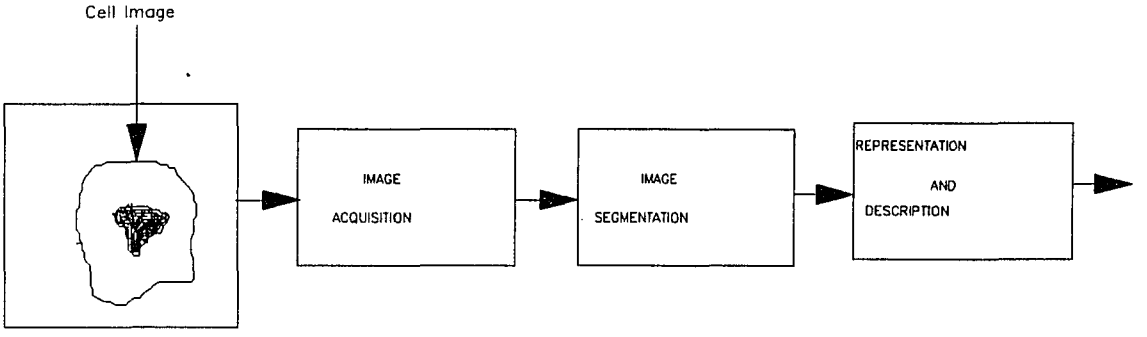


Figure 1.1 Block diagram of cell processing system

CHAPTER 2

FEATURE EXTRACTION SURVEY

Feature Extraction can be considered the data phase of computer vision [Gschwind 1986]. It forms the link between an image higher level processing such as pattern recognition and the raw acquired image. Whatever the general scheme of the vision system may be, it is built upon some basic and fundamental measurements of the digital image. Features provide a numerical value or set of values that represent the spectral and or spatial relationship of the pixels within a digital image. Often it is important for spatially related features to be size and or rotational invariant. This allows object to be randomly oriented or viewed at different distances, thus increasing the system flexibility. A survey of shape features is described below.

Area:

Area is the simplest feature and is generally computed by summing the number of pixels segmented as object pixels [Gonzalez, 1987]. It is not size independent but rotation independent. A very popular use of the area is as a divisor

on other features to create size independence. Freeman (1970) presented a method of measuring the area from the contour chain code. If the 8-connected chain code for the contour is given by $A = a_1, a_2, \ldots, a_n$ then the enclosed area is given by:

$$\hat{A} = \sum_{i=0}^{n} a_{ix}\left(Y_{i-1} + \frac{a_{iy}}{2}\right) \qquad (2.1)$$

Where a_{ix} and a_{iy} are the x and y components of the ith chain element ($a_{ix}, a_{iy} \in \{-1, 0, 1\}$) and y_{i-1} is the ordinate of the point from which the ith chain element departs. Kulpa (1977) estimated area of a polygon with vertices lying on square grids of unit area is given by:

$$\hat{A} = P_i + 0.5P_b - 1 \qquad (2.2)$$

where P_b represent the number of pixels on the boundary and P_i represents the number of pixels in the interior of the polygon.

Perimeter:

Freeman[1969] was the first to employ the chain to find the length of the digitized curve by using the length of each code vector. For an 8-connected chain code with 1:1 aspect

ratio and unit pixel spacing the estimated perimeter length is given by:

$$\hat{L} = n_e + n_o\sqrt{2} \tag{2.3}$$

where n_e and n_o represents the total number of even and odd code vectors respectively. The above yields an overestimated value for the contour parameter. Kulpa (1977) showed that for polygons whose sides have uniformly distributed slopes that the true length of a straight line is overestimated by a factor of $\frac{8(\sqrt{2}-1)}{\pi} \approx 1.055$. Consequently, Kulpa's perimeter estimator is given by:

$$\hat{L} = \frac{(n_e + n_o\sqrt{2})\pi}{8(\sqrt{2}-1)} \tag{2.4}$$

and results in maximal errors of +2.5% and -5.3% for the squares and circle simulated.

Principal Axis Moment:

The Principal axis of a contour is the minimum inertia line which passes through the centroid. *The Principal axis moment* is determined by summing the moment of all pixels of the contour about the object principal axis and can be calculated by:

$$PMI = \sum_y \sum_x \{(x-\bar{x})\sin\theta - (y-\bar{y})\cos\theta\}^2 \qquad (2.5)$$

This is the moment of inertia about the line of slope θ through $(\bar{x},\bar{y})$ [Rosenfeld and Kak, 1982]. Where $\bar{x}$ and $\bar{y}$ are the coordinate of the center. To find the slope θ of the principal axis take the origin at the centroid and determined the moment of inertia of the contour about the line $y = x\tan\theta$ as

$$\sum_y \sum_x \{x\sin\theta - y\cos\theta\}^2$$

$$= \bar{m}_{20}\sin^2\theta - 2\bar{m}_{11}\sin\theta\cos\theta + \bar{m}_{02}\cos^2\theta \qquad (2.5a)$$

Differentiating with respect to θ and equating to zero gives

$$\tan^2\theta + \frac{\bar{m}_{20} - \bar{m}_{02}}{\bar{m}_{11}}\tan\theta - 1 = 0 \qquad (2.6)$$

where

$$\bar{m}_{20} = \sum_y \sum_x (x-\bar{x})^2$$

$$\bar{m}_{02} = \sum_y \sum_x (y-\bar{y})^2$$

$$\bar{m}_{11} = \sum_y \sum_x (x-\bar{x})(y-\bar{y})$$

Central Moment:

Central moment has the same properties as the *principal*

axis moment with the exception that the moment is calculated about the object's centroid instead of the principal axis [Rosenfeld and Kak, 1982]. The central moment carries units of pixels squared and is calculated as follows:

$$CMI = \sum_{y} \sum_{x} \{(x-\bar{x})^2 (y-\bar{y})^2\} \qquad (2.7)$$

Width:

The width can be measured by the sum of the distances from the object's principal axis to the pixel furthest from each side of that axis [Guyer, 1986], as shown in figure[2.1] This will yield a rotationally independent value. Thus

$$Width = a + b \qquad (2.8)$$

where "a" and "b" are the distances from the object principal axis to the pixel furthest from each side of that axis. This feature is rotationally independent and can be size independent and dimensionless if divided by a length measurement, such as perimeter.

Length:

Length is very similar to width; however the axis perpendicular to the principal axis is used instead of the principal axis. The furthest points on each side of perpendicular axis

are used to determined length [Guyer, 1986].

Lobing:

Lobing is a measure of the variability of the object's perimeter from the centroid with respect to perimeter pixels [Guyer, 1988]. A circle for example, would have a small or zero value for lobing while an object long and narrow or deeply lobed would have a high value or degree of lobing. This feature is size dependent and rotationally independent.

(a) Minimum Lobe Tip Angle:

The first step in determining this feature is to identify 'lobes' on the object under consideration. A lobe is defined as a local maxima where a point on the perimeter has maximum distance from the centroid under some criteria. One criteria require the following conditions: 1) that in tracing the border there is no pixel further away from the centroid with in 10 pixels following the last maxima; 2) the distance from the centroid is greater than the distance of previous minima by a specified threshold and 3) the distance from the centroid must exceed the distance of the closest pixel in the 10 pixel neighborhood following the candidate local maxima pixel. The purpose of this criteria is to put constraints on what can

be considered a lobe or local maxima. Local minimas are found in much the same way as the local maximas except that the closet point to the centroid within a neighborhood of ten pixels before and ten pixels after the candidate local minima. This criteria ensure that very small bumps in the perimeter and maximum points with very slight surrounding slope change are not considered as lobes. Once the lobes have been determined the angles at their tips are calculated. The angle at the tip is formed by the angle between the two lines representing the principal axis of the ten pixels before and the axis of the ten pixels following the local maxima (figure[2.2]). The minimum tip angle is the minimum angle formed over the entire perimeter. This feature is rotationally independent and size invarient.

(b) Average Lobe Tip Angle

Average lobe tip angle can be determined simultaneously with the above feature and is simply the average of all located lobe tip angles.

(c) Number of Lobes:

Number of lobes is closely related to the above features and is simply the count of all local maximas located.

(d) Maximum and Minimum distance between Lobes:

The distance between lobes is a measure of the number of

pixel moves between consecutive local maxima. Because the perimeter is 8-connected, the measurement consider horizontal, vertical, and diagonal pixel connections, and the pixel aspect ratio which causes each of the three types of connections to represent a different distance. Considering the 8-connected relationships in such a manner, instead of simply counting pixels, create rotation invarience in the measurement. Both maximum and minimum distance between lobes are size dependent and are the maximum and minimum over the entire perimeter.

(e) Maximum and Minimum Lobe Length:

Lobe length is computed as the difference between the distance from the local maxima to the centroid and the distance of one of the adjacent local minimas to the centroid. Figure[2.3] shows how this feature would be calculated with lobe length equaling either "a" minus "b" or "a" minus "c" depending on whether the maximum or minimum length is being considered. Minimum lobe length would be the minimum over the entire perimeter of all lobes lengths calculated. Maximum lobe length would thus follow as the longest length calculated. Both features are rotationally independent but size dependent.

(f) Minimum Lobe Size:

Minimum lobe size is a measurement around a lobe or local

maxima. It is the pixel measurement from one local minima to the next local minima along the object's perimeter figure[2.4]. Minimum lobe size is an indicator of the presence or absence of relatively small lobes on an object. This feature is rotationally independent and size dependent. Lobe size can be used in relation to lobe length to determine the general shape of an individual lobe.

Elongatedness:

Elongatedness is another common shape descriptor and is defined as area divided by width; the determination of area and width were described above. The ratio is dimensionless and size independent and the feature is rotationally independent.

Roundness Factor(P2A):

The area A and the contour length or circumference P can be used to calculate the roundness factor RF in the following manner [Gray, 1971]:

$$RF = \frac{P^2}{4\pi A} \tag{2.9}$$

This feature produces a minimal value of 1.00 for a perfect circle. For all other shapes RF is greater than 1. The dis-

advantage of this shape factor is that it can yield similar values for contours that are significantly different.

Ellipticity Factor:

By determining the axial ratio ("b/a", where "b" is the minor and "a" is the major half axis of the fitted ellipse) of a given contour, an ellipticity factor EL for this contour can be defined as [Gschwind, 1986]:

$$EL = \frac{RF_c}{RF_{el}} \qquad (2.10)$$

Where RF_c is the roundness factor of the contour, and RF_{el} is the roundness factor of the corresponding (fitted) ellipse. This yields a value of 1.00 for any elliptical contour. Thus, the definition of an ellipticity factor is similar to the roundness factor but describes the deviation from an ellipse.

Convex Hull

The convex hull of a given contour can be defined as the polygon that remains after selectively eliminating all points that lie on concave segments of the contour (figure[2.5]). Sklansky[1972] proposed an algorithm for determining the convex hull of a polygon. His algorithm starts with a point on the

convex hull and tests each vertex for the concavity of the interior angle at the vertex, removing that vertex whenever concavity is ascertained. The concavity tests terminate when the starting point is encountered.

Although this algorithm will perform well for simple polygons it may fail on self-intersecting polygons. Indeed self-intersection may be created by the removal of vertices with concave angles (figure[2.6]).

Graham[1972] overcomes this problem by reordering all points of the polygon in order of increasing polar angle, with origin at a point interior to the polygon. A concavity test similar to that of Sklansky is then performed, with concave points deleted when discovered. The Graham algorithm is as below.

Step 1: Convert all the points on the polygon to (r,θ) polar coordinates about any point z which is interior to the convex hull of the polygon.

Step 2: Order the point by increasing polar angle.

Step 3: If $\theta_i = \theta_{i+1}$, delete the point closest to z. Additionally, any point with an amplitude $r_i = 0$ can also be deleted.

Step 4: The remaining points of the polygon can now be

considered the vertices of a star-shaped polygon each sequence of three consecutive vertices $(i,j,k,) \in S$ (where S represents all points on the polygon) is checked to see if it is a concave vertex. The vertex at j is deleted from S whenever it is concave, resulting in a new sequence of points which in turn must be checked for concavity.

Measurement of convex area:

This method consists of determining the *convex hull* of a contour, calculating its area, and subtracting the area of the original from it. Once the convex hull has been determined, its area can be calculated in the same manner as for the original contour, and concativity factor CF can be defined as [Bykat, 1978]:

$$CF = 100 \frac{(area\ of\ convex\ hull\ -\ area\ of\ contour)}{area\ of\ contour} \quad (2.11)$$

Bending Energy:

Bending energy is the integrated sum of the squared curvature and is a measure of the boundary variation of the contour [Young, 74]. Physically, the bending energy is the energy required to deform a straight rod and the average bending energy E is given by

$$E = \int_0^l |K(p)|^2 dp \qquad (2.12)$$

Where

K(p) = curvature at the point p,

p = position on the contour, and

l = length of the contour.

A circle yields the minimum bending energy for a given contour length. The average bending energy E_{MIN} of a circle of radius r is given by:

$$E_{min} = \left(\frac{1}{r}\right)^2 \qquad (2.12a)$$

To obtain a size invarient measure for the bending energy of a contour, the value should be normalized by dividing it by the bending energy calculated for a circle of equal contour length. The normalized bending energy *BEND* measures the excess energy compared to a circle of the same contour length.

FOURIER DESCRIPTORS

Many researchers have assumed shape to be two dimensional (expressed by the variable "sphericity" and "roundness") although no data exist concerning the sufficiency of two dimensions [Ehrlich, 1980]. Instead of trying to describe shape

with one or two variables, the Fourier technique allows the shape to be multi-dimentional.

In time series analysis a continuous periodic function S(t)=S(t+T) is often approximated over some interval of time, t, by a finite Fourier Series with fundamental period, T. The same may be done for a function $R(\Theta)=R(\Theta+2\Pi)$ which describes, in polar coordinates, the two-dimensional shape of a contour (figure[2.7]). The approximation

$$R(\theta)=R_0+\sum_{n=1}^{N-1} R_n\cos(n\theta-\phi_n) \qquad (2.13)$$

can be made to fit the cell contour shape as closely as required by including sufficient number of harmonic terms, N, in the summation. The function $R(\theta)$ is assume to be continuous and single valued for any angle, θ. Here R_0 is the mean radius from the origin to the edge of the cell and R_n, Φ_n are the amplitude and phase respectively, of the nth harmonic term. For computation with x,y coordinates it is convenient to represent the Fourier Series in the form

$$R(\theta) = R_0 + \sum_{n=1}^{N-1} A_n \cos(n\theta) + \sum_{n=1}^{N-1} B_n \sin(n\theta) \qquad (2.14)$$

Where

$$A_n = R_n \cos(\phi_n) \qquad (2.14a)$$

$$B_n = R_n \sin(\phi_n) \qquad (2.14b)$$

$$R_n = \sqrt{A_n^2 + B_n^2} \qquad (2.14c)$$

$$\phi_n = \tan^{-1}\left(\frac{B_n}{A_n}\right) \qquad (2.14d)$$

and

$$R_0 = \frac{1}{2\pi} \int_0^{2\pi} R(\theta) d\theta \qquad (2.14e)$$

$$A_n = \frac{1}{\pi} \int_0^{2\pi} R(\theta) \cos(n\theta) d\theta \qquad (2.14f)$$

$$B_n = \frac{1}{\pi} \int_0^{2\pi} R(\theta) \sin(n\theta) d\theta \qquad (2.14g)$$

The Fourier Descriptors are used to test for similarities and differences among families of shapes. The ability to perform such an analysis is based on the assumption that the Fourier expansion homologously describes the external geometry on a shape to shape basis. Homology in this sense, means that since

a Fourier series consist of an ordered collection of terms, two objects of the same shape should generate identical series, term for term. Given this homology, one can then validly compare the amplitude spectra for two or more shapes term by term because the value of each term of equivalent order represent the "same thing" in each series. One way of ensuring homology is to minimize the first harmonic term of the series. This term represent the centering error and is an indication of the difference between the center of gravity of discrete set of points used to determined the Fourier Series and the origin of the polar coordinate system used to define those discrete points.

In addition two conditions are inherently assumed with Fourier series representation and are imposed upon the set of discrete points used to calculate the series: (1) the external edge points must occur at equal angles about the polar origin, and (2) all radii from the polar origin to these edge points must intersect the edge exactly once (single value). Therefore, any algorithm that ensures homology between shapes must also satisfy these two conditions as well as minimizing the centering error term of the series. In the case of relatively simple, smooth shapes these conditions can be satisfied, however, for

shapes with highly protuberant and invaginated boundaries where it may be difficult or impossible to generate a valid series. The extent to which a finite Fourier series can be used at all for a given set of shapes, or the level of algorithmic difficulty in satisfying the above criteria, can only be assessed in the course of investigations in which large number of shapes are analyzed.

Equation[2.13] can be used to model the projected cell outline to any degree of precision as long as the number of harmonics used to describe the outline is sufficiently high. Each harmonic term in the above equation represents a simple cosine wave. The amplitude of each harmonic term is a relative measure of how much each wave contribute to the contour. Each of these harmonic amplitudes when squared and divided by the mean radius of the contour represents an independent dimensionless measure of the contribution of that harmonic to the total variance of the perimeter about a circle of mean radius. It is these harmonic amplitudes that are used as shape descriptors for a given contour. The Fourier series contains two variables, the harmonic amplitude and harmonic phase angle, which together are needed for the series to converge precisely to the empiric shape. The amplitude represents the relative

contribution of each term (figure[2.8]) and the phase defines the angular relationship of one term to another. Even harmonic terms (2,4,6,...) are symmetrical with respect to the origin whereas odd terms (3,5,7,...) are asymmetric about the origin.

Ehrlich [1980] has shown that sufficient information is contained in the first twenty harmonics for close reproduction of the original shape; in this case, shape might be considered to be twenty-dimensional. The first two harmonics commonly account for less than 50 percent of the variation in the contour. At least six harmonics are commonly necessary to account for 90 percent or more of the shape variation, meaning that a linear combination of six or more independent variable is necessary to measure shape so precisely that the shape can be effectively regenerated. Thus, while the 20-term Fourier series can be considered to "overdescribe" contour shape, sphericity and roundness may be considered to be insufficient.

Examples of contour regeneration by different harmonics are shown in figure[2.9-12].

FRACTAL SHAPE ANALYSIS

An initial attempt was made to characterize the given contour shape by Fractal Shape analysis techniques, which are discussed in the following sections.

THE FRACTAL DIMENSION

Fractals are class of mathematical functions which have been used to characterize geometrical properties of lines, planes and solids; they have been used to construct synthetic images of a variety of natural shapes and scenes such as biological cells, trees and mountains. The technical definition of a fractal is a set for which the Hausdorff-Besicovich is strictly greater than the topological dimension. Thus a defining characteristic is that a fractal has fractional dimension. The world "fractal" was coined by Mandlebrot (1977) for classes of curves and surfaces which continue to exhibit "detail" as the scale of the measurement decreases. A classic example is that of a coastline, which, when viewed at increasing magnification, continues to reveal feature of progressively smaller size.

A "real world" approach to the meaning of fractal dimension may be made by considering the practical problem of determining the length of an irregular curve. Operationally, one might set a pair of dividers to a given step length, s, and "step" along the curve counting the number of steps, N. Thus the length of the curve at the scale of s, L(s), would have the value Ns. One would expect that by taking successively shorter steps, L(s) would approach a limiting value as the steps become infinitesimally small. This value may be defined as the length of the curve. For the Euclidean curves the limit exists and is the basis for the expression for the length of a curve as derived in calculus textbooks. However, if a curve exhibits detail at all scale of measurement the limit may not exist. As the step size is decreased, the smaller details which were previously stepped over add their contributions to the length of the curve. Thus the value of L(s) continues to increase with decreasing step length. The functional relationship between step size and length is:

$$L(s) = s^{1-D} \tag{2.15a}$$

Or

$$D = 1 - \frac{\text{Log}\, L(s)}{\text{Log}\, s} = 1 - m \tag{2.15b}$$

where L(s) = Total length

s = step size

D = Fractal dimension and

m = the slope of the line in the log(L(s)-log(s) plot also known as Richardson plot.

The Fractal dimension method used by Frisch et al[1987] to analyze the shape of irregular grain particle contours as following:

Step 1.

The contour was unrolled by the tangent-angle function (Zahn and Roskie 1972). This procedure would convert a perfectly smooth circle into a straight line by removing its constant curvature.

Step 2.

The contour's perimeter was determined as a function of step length.

Step 3.

The slope (m) of the plot (Richardson plot) of the step length verses perimeter was calculated.

Step 4.

The fractal dimension (D) was then determined from the equation D=1-m. This fractal dimension was used as the representative shape parameter for that particular contour.

ITERATED FUNCTION SYSTEMS (IFS)

Barnsley proposed IFS as model for fractal. The idea here is to represent a fractal by a set of numbers which can be thought of as a basis for the fractal along with a set of operations to be performed on these numbers so that the fractal can be regenerated. The theory can be described as follow.

An affine transformation can be described as the combination of rotations, scaling and translations of the coordinate axis in n-dimensional space. The general form of an affine transformation is

$$W\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \qquad (2.16)$$

Where coefficients a,b,c,d,e and f are real numbers. This transformation can also be written in the form

$$W\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r_1\cos\theta_1 & -r_2\sin\theta_2 \\ r_1\sin\theta_1 & r_2\cos\theta_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \qquad (2.17)$$

where r_1 = scaling factor on x

r_2 = scaling factor on y

θ_1 = angle of rotation on x

θ_2 = angle of rotation on y

e = translation on x and

f = translation on y.

This transformation will move a point (x,y) to (ax+by+e, cx+dy+f).

IFS theory is an extension of classical geometry and is the collection of affine transformations. It uses affine transformations to express relation between parts of an image. An example of an IFS is Sierpinski triangle which requires three affine transformations to generate,

$$w_1\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}.5 & 0\\0 & .5\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}0\\0\end{pmatrix} \qquad (2.18a)$$

$$w_2\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}.5 & 0\\0 & .5\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}1\\0\end{pmatrix} \qquad (2.18b)$$

$$w_3\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}.5 & 0\\0 & .5\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}.25\\.5\end{pmatrix} \qquad (2.18c)$$

Each transformation must also have an associated probability, p_i, determining its importance relative to other transformations. In the present case we might have p_1, p_2 and p_3. Notice that the probabilities must add up to 1. That is, $p_1 + p_2 + p_3 = 1$. An IFS can contain any number of affine transformations. The IFS codes for the Sierpinski triangle are

W	a	b	c	d	e	f	p
1	.5	0	0	.5	0	0	.33
2	.5	0	0	.5	1	0	.33
3	.5	0	0	.5	.25	.5	.34

Once we have the IFS code for an image, we can throw away the original image and keep the code, achieving compression ratio of 10,000 to 1. The original image can be generated using the IFS code by Random iterated algorithm. The following code summarizes the method:

(i) Initialize: x=0, y=0.

(ii) For n=1 to 2500, do steps (ii)-(iv)

(iii) Choose k to be one of the numbers 1,2,...m, with probability p_K. Where m is the number of transformations.

(iv) Apply the transformation W_K to the point (x,y) to obtain $(\tilde{x},\tilde{y})$.

(v) Set (x,y) equal to the new point: $x=\tilde{x}$, $y=\tilde{y}$.

(vi) if n>10, plot (x,y).

(vii) Loop.

Applying this procedure to the transformation of the above table produces the figure shown in figure[21], a fractal known

as the Sierpinski triangle. Increasing the number of iterations n adds points to the image. In step(iv) the first 10 points are not plotted because this will give the randomly dancing point time to settle down on the image.

Attractor of an IFS:

The Attractor of an IFS is the union of the affine transformations of an image and is denoted by G:

$$G = \bigcup_{n=1}^{N} W_n(G) \tag{2.19}$$

In the case of Sierpinsky triangle

$$G = W_1(G) \cup W_2(G) \cup W_3(G) \tag{2.20}$$

Every affine transformation generate part of an image (i.e. sub image) and the union of the sub images is the Attractor of the whole image.

FRACTAL INTERPOLATION

Fractal interpolation function provide a new mean of fitting experimental data, that is the graph of the fractal interpolation function can be made close to the original data [Barnsley, 1988].

Given a data set { (x_n,y_n): n = 0,1,2,....N } be given and $x_0<x_1<x_2<...x_N$ we explain how to construct an IFS in R^2 (i.e. in 2D) such that the attractor, denoted by G, is the graph of a continuous function $f:(x_0,x_N)\to R$ interpolates the data. We consider an IFS of the form $\{R^2; W_n; n=1,2,..N\}$ where the maps are affine transformation with the special structure

$$W_n\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}a_n & 0\\c_n & d_n\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}e_n\\f_n\end{pmatrix} \qquad (2.21a)$$

The transformations are constrained by

$$W_n\begin{pmatrix}x_0\\y_0\end{pmatrix}=\begin{pmatrix}x_{n-1}\\y_{n-1}\end{pmatrix} \text{ and } W_n\begin{pmatrix}x_N\\y_N\end{pmatrix}=\begin{pmatrix}x_n\\y_n\end{pmatrix} \text{ for n = 1,2,....N.}$$

The transformation W_n is specified by the five real number a_n, c_n, d_n, e_n and f_n which must obey the following four linear equations:

$$a_nx_0 + e_n = x_{n-1} \qquad (2.22a)$$

$$a_nx_N + e_n = x_n \qquad (2.22b)$$

$$c_nx_0 + d_ny_0 + f_n = y_{n-1} \qquad (2.22c)$$

$$c_nx_N + d_ny_N + f_n = y_n \qquad (2.22d)$$

Let d_n be any real number, solving the above equations for a_n, c_n, e_n and f_n in terms of the data and d_n we get:

$$a_n = \frac{x_n - x_{n-1}}{x_N - x_0} \tag{2.23a}$$

$$e_n = \frac{x_N x_{n-1} - x_0 x_n}{x_N - x_0} \tag{2.23b}$$

$$c_n = \frac{(y_n - y_{n-1}) - d_n(y_N - y_0)}{x_n} - x_0 \tag{2.23c}$$

$$f_n = \frac{(x_N y_{n-1} - x_0 y_n) - d_n(x_N y_0 - x_0 y_N)}{x_N - x_0} \tag{2.23d}$$

We choose the free parameter d_n (which is also called vertical scaling factor) for the following reason: by adjusting the scaling factor of each map we can force the curve to lie in a prescribed range. The transformation W_n maps lines parallel to the y-axis into lines parallel to the y-axis. Let L denote the segment parallel to the y-axis, then $W_n(L)$ is also a line segment parallel to the y-axis. The ratio of $W_n(L)$ to the length of L is $|d_n|$ i.e. $|d_n| = \frac{W_n(L)}{L}$ and $0 \le d_n \le 1$ for n = 1,2,...N. With d_n = 0 for n=1,2,...N one can recover the piece wise linear interpolation function.

We tried the fractal interpolation to describe the contour curves in terms of IFS. We then tried to compare the IFS's of two different curves in order to discriminate. We reached to

the following conclusion:

There is no guaranty that the number of windows contain in an IFS of one curve will be equal to the number of windows in the other curve IFS. Therefore it is not possible to discriminate between the two curves by comparing the IFS's.

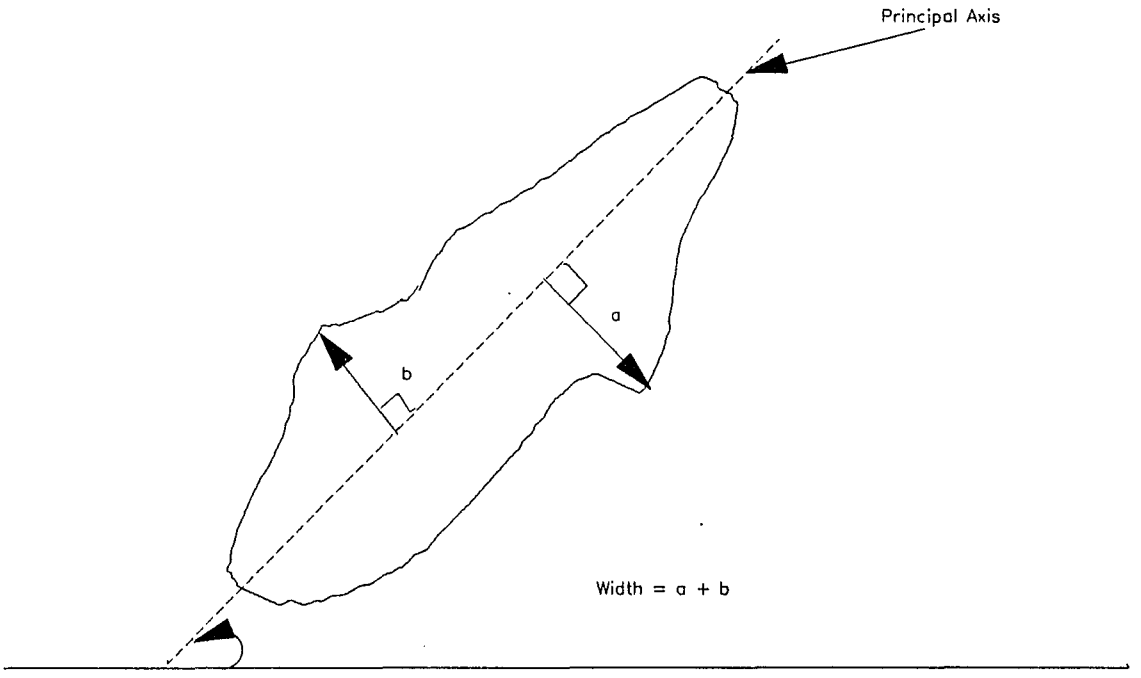


Figure 2.1 Determination of the width feature.

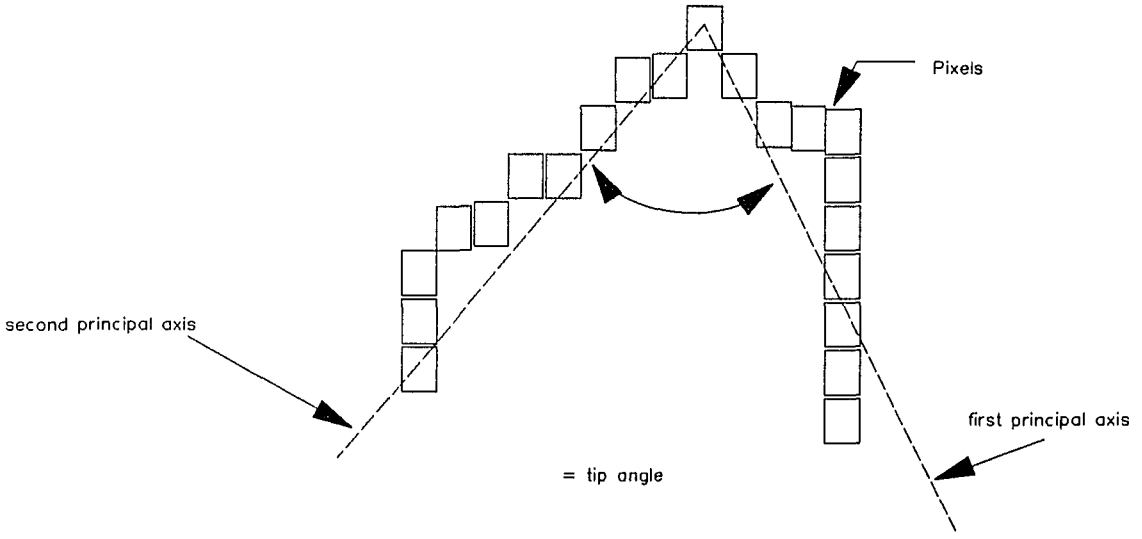
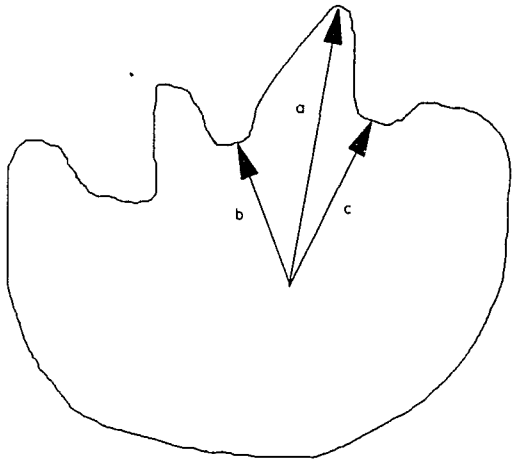


Figure 2.2 Determination of tip angle feature.



lobe length = a−b (for maximum length)

or

lobe length = a−c (for minimum length)

Figure 2.3 Determination of length of lobe feature.

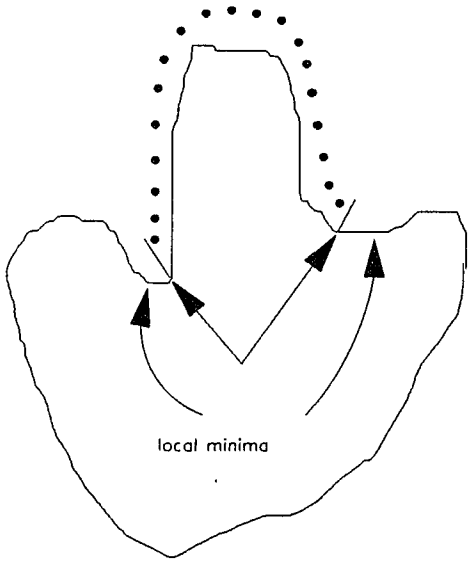


Figure 2.4 Determination of size of lobe feature.

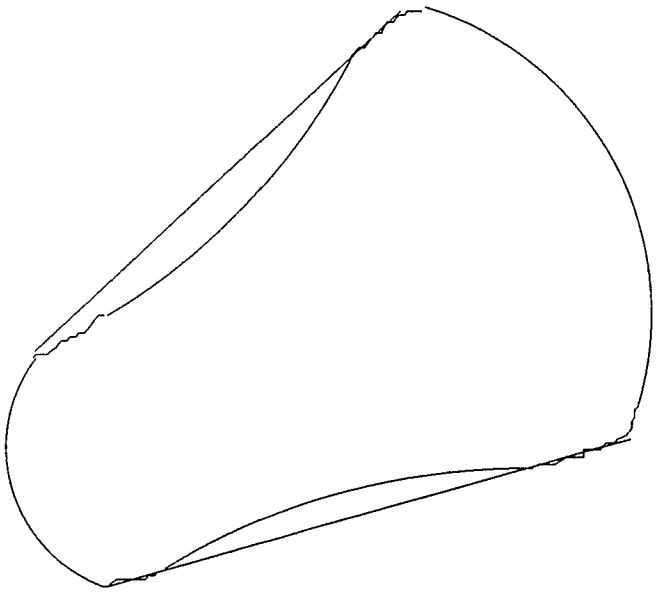


Figure 2.5 Determination of the convex hull. The convex hull combines the countourpoints remaining after having eliminated all points lying on concave contour segments

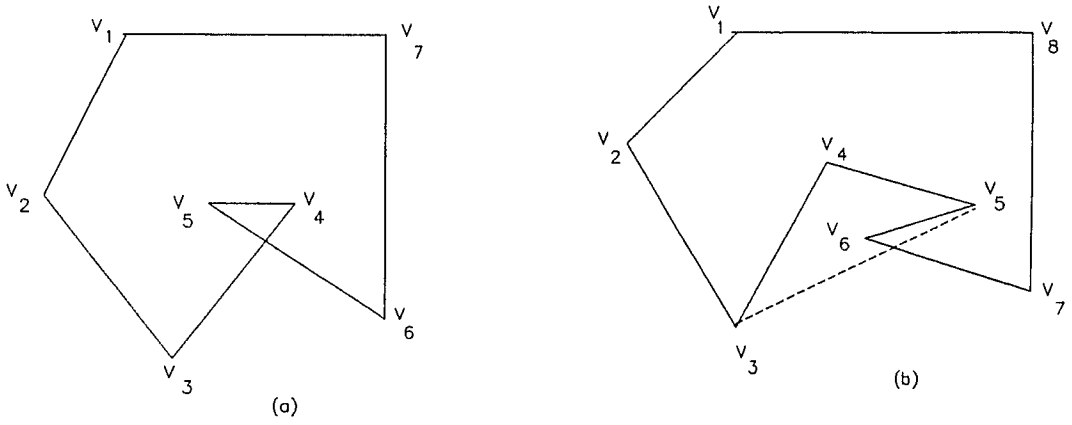


Figure 2.6

a: A self-intersecting polygon with convex interior angles.

b: A non self-intersecting polygon with a loop induced by removing vertex V_4.

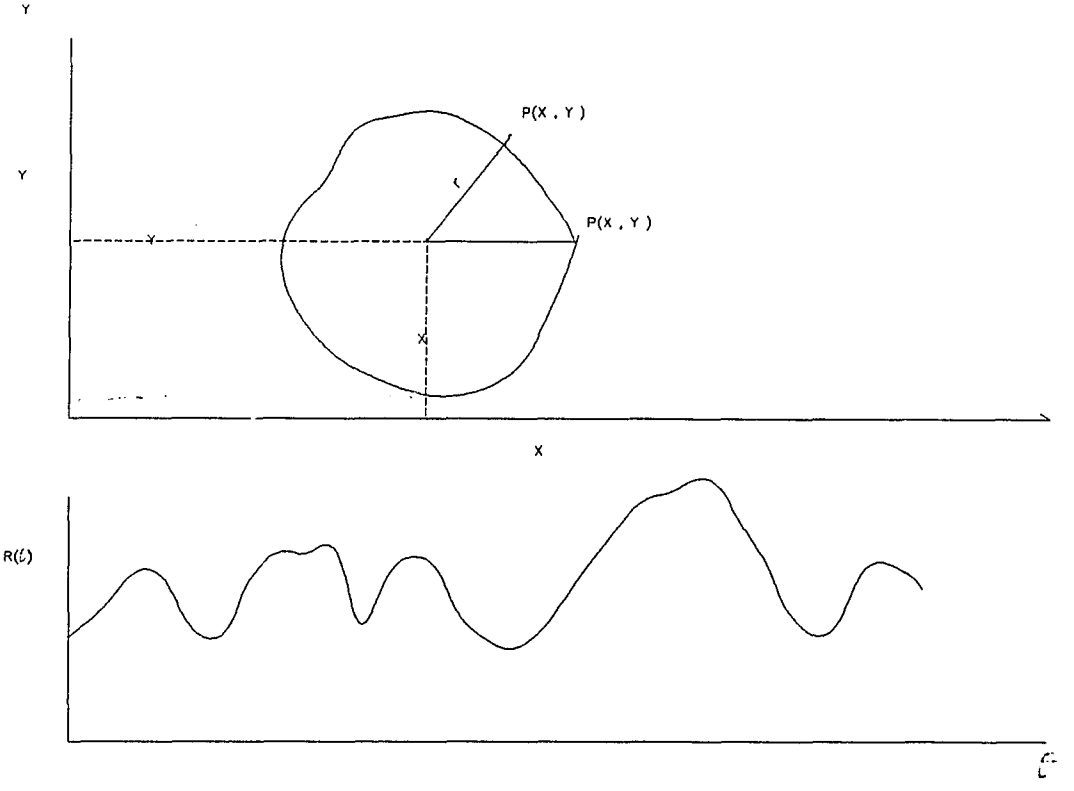


Figure 2.7 Two dimensional Cell image shown in (a) X–Y coordinate and (b) polar coordinate form.

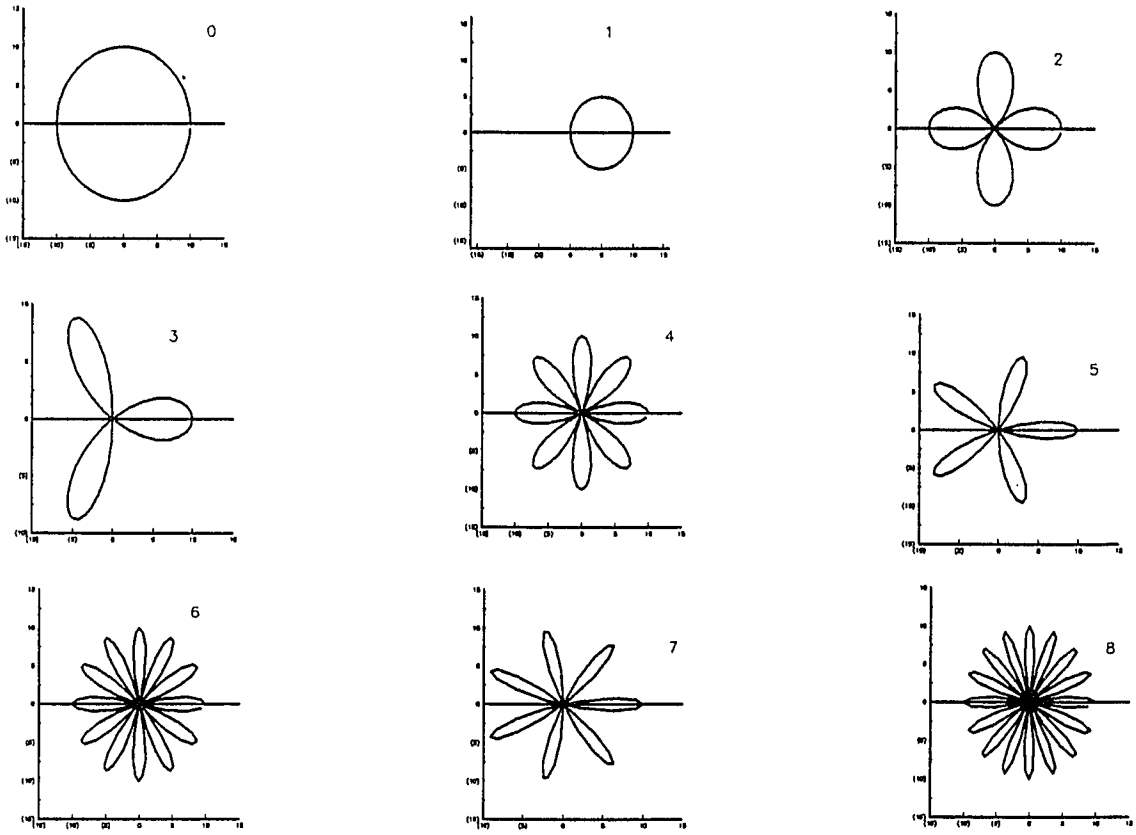


Figure 2.8 Plots of circular harmonics. Amplitude = 10.0; Phase Angle = 0

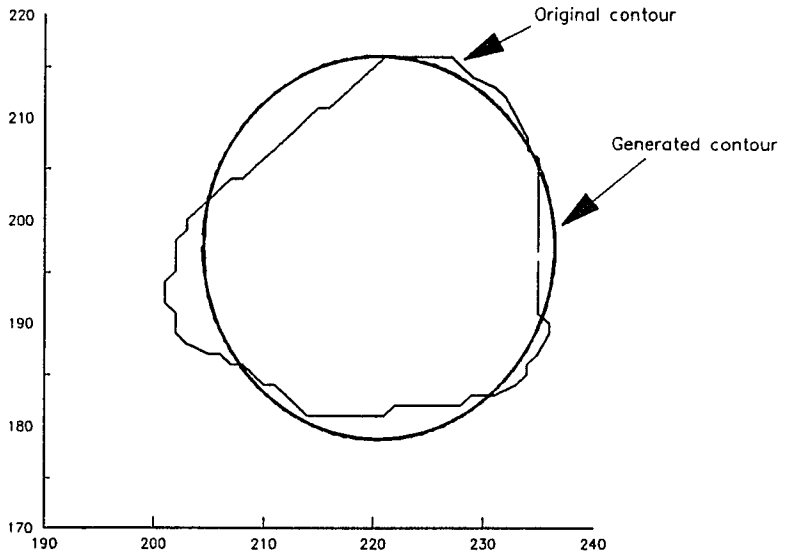


Figure 2.9 The contour generated by the D.C. harmonic and the original contour.

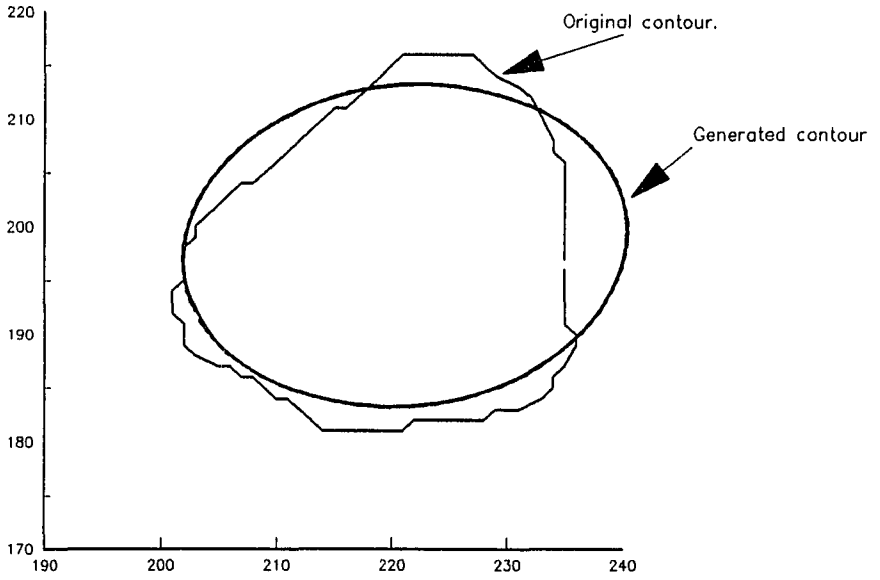


figure 2.10 The contour generated by the D.C. and second harmonics and the original contour.

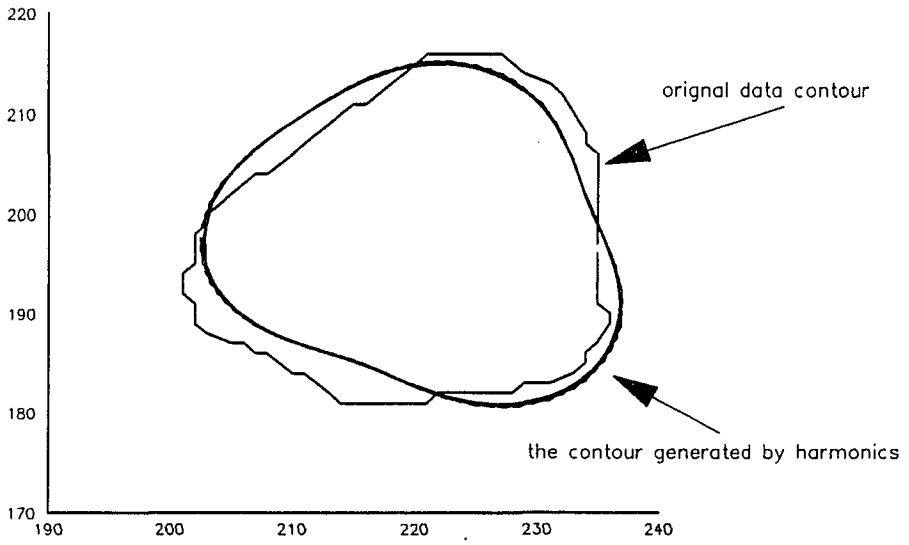


Figure 2.11 First Tree harmonics and the original contour.

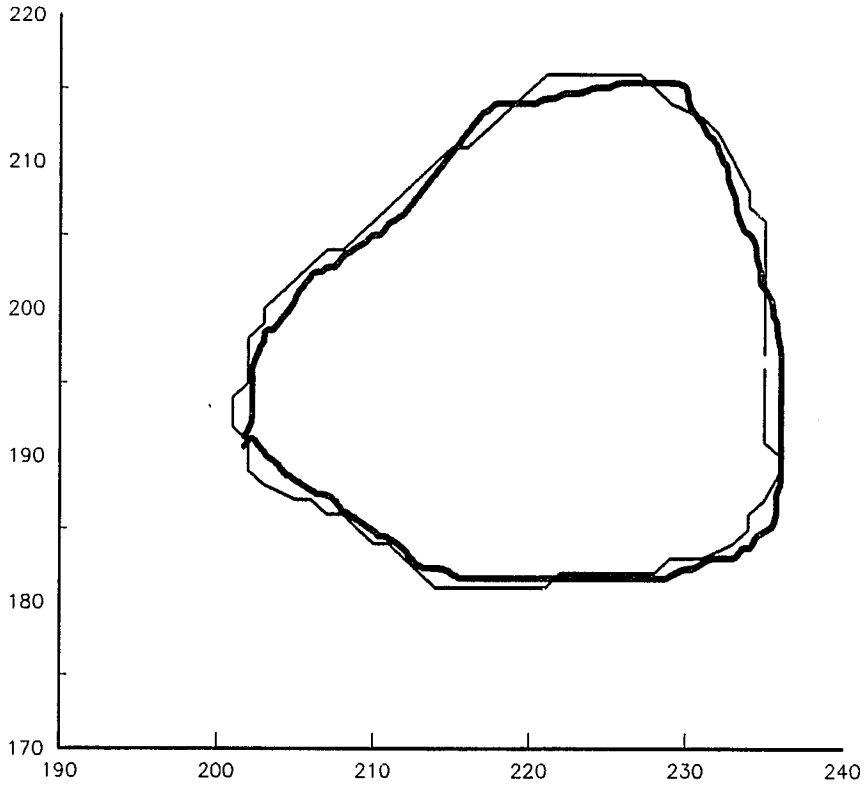


Figure 2.12 The contour generated by the first 20 harmonics and the original data contour.

CHAPTER 3

GRID-SEARCH METHOD FOR CELL SHAPE ANALYSIS

In this section we present the Grid search method for determining the best fit ellipse to the contour data of cells, which will be compared with our Fourier method in the later chapter. This method was proposed by Bevington(1969). The characteristic of the contour shape can be extracted by computing the parameters of the ellipse and measuring the difference of area of the contour from the ellipse. We assume that the cell/nucleus contour has been extracted and is available to us in the form of x,y coordinates.

The Grid Search method consist of iteratively fitting an ellipse to the contour until an ellipse with minimal difference area is obtained. This requires that five parameters be optimized: the x,y coordinate of the center; the area; the long and short axis of the ellipse; and the angle of the long axis of the ellipse with respect to the x-axis. The *grid-search method* optimizes these parameters by minimizing a criterion function, which measures the goodness of fit (in this case, the difference area), separately for each parameter describing

the fit ellipse.

The method of least squares is built on the hypothesis that the optimum description of a set of data is one which minimizes the weighted sum of squares of deviations of the data y_i from the fitting function $y(x_i)$. The sum is characterized by the variance of the fit X^2 (also called the criterion function), which is an estimate of the variance of the data σ^2. For a function with n coefficient fit to N data point [Bevington, 1965]:

$$X^2 \equiv \sum \left(\frac{1}{\sigma_i^2} \{ y_i - y(x_i) \}^2 \right) \qquad (3.1)$$

If the variation of X^2 with each parameter a_j (j=1..n) is fairly independent of how well optimized the other parameters are, then the optimum value can be determined most simply by minimizing X^2 with respect to each parameter separately. This method is called the grid search method. The procedure for the grid search is as follows:

1. One parameter a_j is incremented by a quantity Δa_j, where the magnitude of this quantity is specified and the sign is chosen such that X^2 decreases.

2. The parameter a_j is repeatedly incremented by the same amount Δa_j until X^2 starts to increase.

3. Assuming the variation of X^2 near the minimum can be described in terms of a parabolic function of the parameter a_j, we can use the value of X^2 for the last three values of a_j to determine the minimum of the parabola in figure[3.1].

$$a_j(3) = a_j(2) + \Delta a_j = a_j(1) + 2\Delta a_j \qquad (3.1a)$$

$$X^2(3) > X^2(2) \leq X^2(1) \qquad (3.1b)$$

4. The minimum of the parabola is given by

$$a_j(\min) = a_j(3) - \Delta a_j \left(\frac{X^2(3) - X^2(2)}{X^2(3) - 2X^2(2) + X^2(1)} + \frac{1}{2} \right) \qquad (3.2)$$

5. X^2 is minimized for each parameter in turn.

6. The above procedure is repeated until the last iteration yields a negligibly small decrease in X^2.

Computer Implementations of the Grid Search Method:

We assume that the cell/nucleus contour has been extracted and is available to us in the form of x,y coordinates. The initial parameters for the ellipse were estimated by the following equations:

$$x_c = \frac{(x_{max} - x_{min})}{2} \qquad (3.3)$$

$$y_c = \frac{(y_{max} - y_{min})}{2} \tag{3.4}$$

Using the initial center coordinates, the contour points were converted into polar coordinates and the rest of the parameters were approximated as following:

θ = the orientation of the largest radius on the given contour

a = the largest radius one the given contour and

b = the smallest radius on the given contour.

The criterion function which we used for determining the best fit ellipse (in terms of area) for a given contour is the following:

$$X^2 \equiv \frac{A_c + A_e - 2A_I}{A_c} \tag{3.5}$$

Where

A_C = Area of the given contour

A_e = Area of the matched ellipse and

A_I = Area of the intersection between the ellipse and the contour.

The area of the contour was calculated by the following equation:

$$A_c \approx \left| \frac{1}{2} \sum_{i=1}^{n} (x_i - x_{i+1})(y_i + y_{i+1}) \right| \qquad (3.6)$$

and the area of the ellipse was calculated by:

$$A_e = \Pi a b$$

where a and b are the long and short axis of the ellipse respectively.

Calculation of the area of Intersection (A_I):

The contour point was converted from X-Y plane to polar coordinates (i.e.$R_c(\theta)$) by approximating the initial center X_C and Y_C and the polar coordinates for the ellipse (i.e.$R_e(\theta)$) were calculated by the following:

The equation for the ellipse is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad (3.6a)$$

where

$$x = R_e \cos\theta \qquad (3.6b)$$

and

$$y = R_e \sin\theta \qquad (3.6c)$$

therefore

$$R_e = \sqrt{\frac{a^2 b^2}{a^2 \sin^2\theta + b^2 \cos^2\theta}} \qquad (3.7)$$

and

$$R_I(\theta) = Minimum\{R_c(\theta), R_e(\theta)\} \qquad (3.8)$$

Once $R_I(\theta)$ is determined the area of the intersection A_I was calculated by converting $R_I(\theta)$ into x-y plane and using equation [3.6] .

The other parameters which are needed as input to the program after the first pass are Δa_j , which were estimated as:

$$\Delta a_j = \frac{a_j}{50} \qquad (3.9)$$

where j = 1,2..n

The main disadvantage of this method is that if the variation of the criterion function with the different parameters depends on each other, the criterion function may converge very slowly towards the minimum.

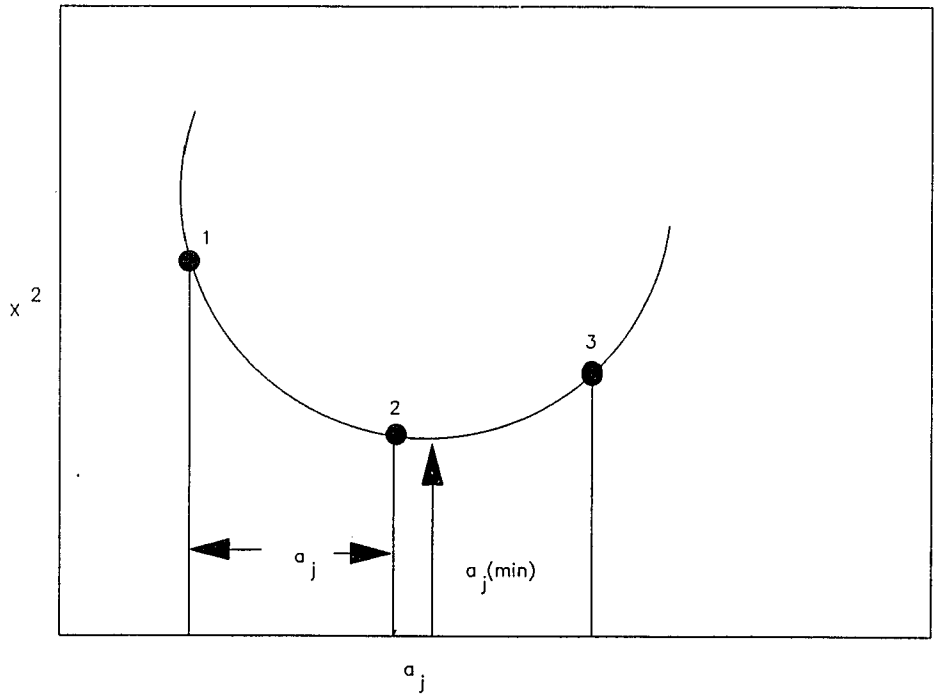


Figure 3.1 Parabolic interpolation to find position a_j(min) for minimum X^2 using the values of X^2 for a_j(1), a_j(2), and a_j(3).

CHAPTER 4

FOURIER METHOD FOR CELL SHAPE ANALYSIS

In this chapter we present our Fourier Method for cell shape analysis, which will be compared for efficiency and accuracy with the Grid search method in the next chapter. We will also describe the applications of our method for the diagnosis of cells in the later chapter.

The Fourier Series consist of an ordered collection of terms, two object of the same shape should generate identical series, term by term. One way of ensuring homology is to minimize the first harmonic of the series which represents the centering error on the $R(\alpha)$ plane. This term is an indication of the difference between the center of gravity of discrete set of points and the origin of the polar coordinate system used to define those discrete points. By minimizing the first term of the series in the $R(\alpha)$ plane, we locate the exact center of the relatively smooth contour this will fail if there are very high perturbations on the contour. In our case we used the geometric center, which means that if a line through the center cuts the contour into two parts, the area of the

two should be the same; or the summation of the area of the triangle elements in every direction should be equal to zero. This summation represents the first harmonic of the Fourier Series on the $R^2(\alpha)$ plane, therefore the first term of the $R^2(\alpha)$ is minimized. This technique will locate the exact center even though there are high perturbation on the contour.

In addition to minimizing the centering error, two additional conditions are imposed upon the set of discrete points used to compute the series: (1) the external edge points must occur at equal angular increment about the polar origin, and (2) all radii from the polar origin to these points must intersect the edge exactly once (single value). These two conditions are inherently assumed with any closed form Fourier Series.

The Fourier Series expansion can be made to fit the cell contour shape as closely as required by including sufficient number of harmonic terms, provided the centering error is minimized and the function $R(\alpha)$ is both continues and single valued for any angle α. Our task is to match an ellipse to the cell contour. Therefore, we compare the expansion of the Fourier Series of the cell contour to the Fourier Series of an ellipse. It is easier to assume the given contour to have a perfect

elliptical shape (i.e zero orientation and centered at the origin) by translating and rotating it and then comparing it to the Fourier expansion of the perfect ellipse. The Fourier series of the perfect ellipse contains only d.c. and second harmonics. Therefore, by computing the d.c. and second harmonics of the Fourier expansion of the contour and compare it to the parametric equation of the perfect ellipse, the initial long axis "a" and short axis "b" for the matched ellipse is determined. In our algorithm we takes the initial values of "a" and "b" and use an iterating procedure to generate new sets of "a's" and "b's" until the difference of the area between the ellipse and the contour is minimized.

Parametric Equation of an ellipse

An ellipse can be represented by the following parametric equations:

$$X(t) = a\cos(t)\cos(\theta) - b\sin(t)\sin(\theta) + X_c \qquad (4.1)$$

$$Y(t) = a\cos(t)\sin(\theta) + b\sin(t)\cos(\theta) + Y_c \qquad (4.2)$$

where

a = long axis of the ellipse

b = short axis of the ellipse

θ = the orientation of the ellipse (i.e. the angle of the long axis to the x-axis)

X_c, Y_c = the coordinates of the center of the ellipse and

t = a parameter $0 \le t \le 2\pi$

The goal is to find a, b, θ, X_c, and Y_c from the contour data of the cell so that the match ellipse can be fully determined. Representation of an ellipse with basic parameters is shown in figure[4.1].

Fourier Series

Any periodic function can be expended in terms of Fourier series. Since the original contour is given by ordered pair(x,y), we must first convert it to $R-\alpha$ and $R^2-\alpha$ plane by assuming an initial center (x_c, y_c) by the following equations:

$$R^2 = (x_i - x_c)^2 + (y_i - y_c)^2$$

$$\alpha_i = \tan^{-1}\left(\frac{(y_i - y_c)}{(x_i - x_c)}\right) \tag{4.3}$$

$$\Delta\alpha_i = \alpha_{i+1} - \alpha i \tag{4.4}$$

$R^2(\alpha)$ is a periodic function of period 2π. Therefore, it can be expanded in terms of its Fourier series:

$$R^2(\alpha) = A_0 + \sum_{m=1}^{n-1} A_m \sin\left(\frac{2\pi m\alpha}{T_0}\right) + B_m \cos\left(\frac{2\pi m\alpha}{T_0}\right) \qquad (4.4a)$$

where

$T_0 = 2\pi$ is the period of the function $R^2(\alpha)$.

Therefore

$$R^2(\alpha) = A_0 + \sum_{m=1}^{n-1} A_m \sin(m\alpha) + B_m \cos(m\alpha) \qquad (4.5a)$$

Utilizing the orthogonality properties of sine and cosine, the Fourier coefficient may be evaluated as

$$A_0 = \frac{1}{2\pi}\int_0^{2\pi} R(\alpha)d\alpha \qquad (4.5b)$$

$$A_m = \frac{1}{\pi}\int_0^{2\pi} R^2(\alpha)\cos(m\alpha)d\alpha \qquad (4.5c)$$

$$B_m = \frac{1}{\pi}\int_0^{2\pi} R^2(\alpha)\sin(m\alpha)d\alpha \qquad (4.5d)$$

Equation (4.5a) can also be written as

$$R^2(\alpha) = S_0 + \sum_{m=1}^{n-1} S_m \cos(m\alpha - \phi_m) \qquad (4.5e)$$

where

$S_m = \sqrt{A_m^2 + B_m^2}$ are called Fourier amplitudes, and

$\phi_m = \tan^{-1}\left(\frac{B_m}{A_m}\right)$ are the phases of the Fourier series.

Centroid

Two techniques were analyzed to find the centroid of the contour; the first one suggested by Full and Ehrlich [1982] says that minimizing the first harmonic term of the Fourier expansion of the contour in $R(\alpha)$ plane will determined the exact center of the contour. This term represent the centering error and it is an indication of a difference between the center of gravity of a discrete set of points used to determine the Fourier series and the origin of the polar coordinate system used to define those discrete points. To minimize the distance between this center of gravity and the polar origin the first term of the series must also be minimized. This technique will work for a relatively smooth contour, but as shown in figure[4.2]; will fail for a contour with a spike on it or highly irregular shape. In order to overcome this problem we used the geometric center, which is explained below.

The Geometric center

The geometric center which is defined as " The summation of the area of the triangle elements in every direction should

be zero." It is can also be called the center of mass of the area.

That is

$$\lim_{\Delta \to 0} \sum_{\alpha=0}^{2\pi} \Delta E(\alpha) = 0 \tag{4.6a}$$

where

$$\Delta E(\alpha) = \frac{1}{2} R(\alpha) R(\overrightarrow{\alpha}) \Delta\alpha \tag{4.6b}$$

therefore

$$\lim_{\Delta E \to 0} \sum_{\alpha=0}^{2\pi} \Delta E = \frac{1}{2} \int_0^{2\pi} R^2(\alpha) e^{j\alpha} d\alpha = 0 \tag{4.6c}$$

$$= \frac{1}{2} \int_0^{2\pi} R^2(\alpha) \cos(\alpha) d\alpha + \frac{j}{2} \int_0^{2\pi} R^2(\alpha) \sin(\alpha) d\alpha = 0 \tag{4.6d}$$

In order for the above equality to be satisfied, we must have:

$$\int_0^{2\pi} R^2(\alpha) \cos(\alpha) d\alpha = 0 \tag{4.7}$$

and

$$\int_0^{2\pi} R^2(\alpha) \sin(\alpha) d\alpha = 0 \tag{4.8}$$

The integrals represent the first harmonic of the Fourier series in the $R^2-\alpha$ plane.

where

$$R^2=(x_i-x_c)^2+(y_i-y_c)^2 \tag{4.8a}$$

$$\alpha_i=\tan^{-1}\left(\frac{(y_i-y_c)}{(x_i-x_c)}\right) \tag{4.8b}$$

$$\Delta\alpha_i=\alpha_{i+1}-\alpha i \tag{4.8c}$$

The $R^2(\alpha)$ is a periodic function of period 2π so it can be expanded in terms of Fourier series. The first harmonic of this series represents the error in the center. An iterating procedure is used to locate the center until the first harmonic of the series is minimized.

Normalizing the contour points

In order to have a valid Fourier series expansion the external edge points must occur at equal angles about the polar origin. The M evenly spaced points (normalized) $P(R_j,\alpha_j), j=1,...M$, are obtain through linear interpolation using

$$R_j = R_i + \frac{(R_{i+1} - R_i)(\alpha_j - \alpha_i)}{\alpha_{i+1} - \alpha_i} \qquad (4.10)$$

where

$P(R_i, \alpha_i), i = 1, \ldots N$, are the original contour points.

$$\alpha_{i+1} > \alpha_j \geq \alpha_i \text{ and}$$

$$\alpha_{j+1} = \alpha_j + \frac{2\pi}{M}$$

The equivalent cartesian points are

$$x_j = R_j \cos\alpha_j + x_c \qquad (2.10a)$$

$$y_j = R_j \sin\alpha_j + y_c \qquad (2.10b)$$

Where x_c, y_c is the centroid of the contour.

Elimination of the multi-values

Once the contour points are normalized at equal intervals, the next step is to eliminate the multi-valued edge point on the cell contour. As shown in figure[4.3], the function R^2 in the α direction should be single valued in order to have a valid Fourier expansion.

A way to circumventing the multi-value problem is to represent the contour by plotting the incremental area enclosed by the contour as a function of angle. This representation is

periodic with period 2π and can be represented by its Fourier series. Assuming that two connective contour points separated by an angle $\Delta\alpha$ are close so their radius from the centroid can be approximated by R then the incremental area between the centroid and these points is given by $\frac{1}{2}R^2\Delta\alpha$. Thus a single value of R^2 in α direction can be achieved by the following equation.

$$R^2(\alpha_i)\Delta\alpha_i = R_k^2(\alpha_i)\Delta\alpha_k + R_m^2(\alpha_i)\Delta\alpha_m + R_n^2(\alpha_i)\Delta\alpha_n + \ldots \quad (2.10c)$$

where $1 <= i,k,m,n <= N$ and N is the number of points on the contour.

$$\Delta\alpha_i = \alpha_{i+1} - \alpha_i \quad (2.10d)$$

and

$$|\Delta\alpha| = |\Delta\alpha_i| = \frac{2\pi}{N} \quad i = 1,2....N \quad (2.10e)$$

The magnitude of all $\Delta\alpha$ are the same because contour points were normalized at equal intervals. However, $\Delta\alpha$ can be negative or positive. In the example of figure[4.3] $\Delta\alpha_2$ is negative while $\Delta\alpha_1$ and $\Delta\alpha_3$ are positive.

Determination of the Orientation of the contour

We assume that the contour is an elliptical shape, therefore the angle to the long axis to the x-axis (i.e. Orientation θ) can be found with the help of the second harmonic of the Fourier series as shown in figure[4.4].

$$\theta = \frac{1}{2}\sin^{-1}\frac{A_2}{s_2} \qquad (4.11)$$

Where

θ = the Orientation of the contour,

s_2 = Amplitude of the second harmonic of the Fourier series and

A_2 = sine term of the second harmonic.

Matching

So far we have determined the center coordinates (x_c, y_c), and the orientation θ. In order to match an ellipse with this contour with the same orientations and center we need to determine two additional parameters; the long axis "a" and the short axis "b". Since the Fourier coefficients are invarient under rotation, translation and scaling we can rotate the original contour by angle θ, and translate to bring the center to the origin to align the contour with a perfect ellipse. The perfect ellipse is one which has a zero orientations and is centered at the origin (Fig.[4.5])

That is

$$\theta = 0$$

$$(x_c, y_c) = (0,0)$$

The equation for the perfect ellipse is

$$x = a\cos(t) \tag{4.12a}$$

$$y = b\sin(t) \quad 0 \le t \le 2\pi \tag{4.12b}$$

Using the above two equations we get

$$R_o^2 = x^2 + y^2 = a^2\cos^2(t) + b^2\sin^2(t) \tag{4.12c}$$

or

$$R_e^2 = \frac{a^2+b^2}{2} + \frac{a^2-b^2}{2}\cos(2t) \qquad (4.12d)$$

comparing equations (4.5a) and (4.12), we conclude that only the d.c. and the second harmonic terms are present in the Fourier expansion of the perfect ellipse.

That is

$$s_0 = \frac{a^2+b^2}{2} \qquad (4.13)$$

$$s_2 = \frac{a^2-b^2}{2} \qquad (4.14)$$

therefore

$$a = \sqrt{s_0+s_2} \qquad (4.15)$$

$$b = \sqrt{s_0-s_2} \qquad (4.16)$$

We use the d.c., and second harmonic terms of the Fourier expansion of the original contour, and determined the initial "a" and "b" for the match ellipse. Using this initial "a" and "b", an ellipse can be generated, that transforms $R^2(\alpha)$ into the $R^2(t)$ plane. This can be done with the help of the following equations:

$$x = a\cos(t) \qquad (4.16a)$$

$$y = b\sin(t) \tag{4.16b}$$

or

$$\tan(t) = \frac{a}{b}\left(\frac{y}{x}\right) = \frac{a}{b}\tan(\alpha) \tag{4.16c}$$

therefore

$$t_i = \tan^{-1}\left(\frac{a}{b}\tan\alpha_i\right) \tag{4.17}$$

where i = 1,....N are the number of points on the contour.

Using the initial values of "a" and "b" we iterate to generate new sets of "a's" and "b's" such that

$$\pi ab = area\ \ of\ \ contour$$

$$\left|\frac{a_c}{b_c}\right| - \left|\frac{a_e}{b_e}\right| \leq \delta_0 \tag{4.18}$$

where

πab is the area of the ellipse

$\frac{a_c}{b_c}$ = the ratio of the long axis to the short axis of the given contour

$\frac{a_e}{b_e}$ = the ratio of the long axis to the short axis of the matched ellipse, and

δ_0 = minimum threshold value.

Algorithm

-Original contour given by ordered pairs (x_i, y_i).

-Convert the original contour into polar coordinates (R_i, α_i).

-Determine the centroid of the original contour.

-Expand the signature of R^2 *vs* α into a Fourier series and use the second harmonic coefficient to determined the orientation of the long axis. The original contour is rotated to aline the orientation with the horizontal axis.

-Normalize the contour points at equal intervals through linear interpolation and eliminate the multi-value edge points.

-Use the DC and second harmonic coefficient to determined the initial long and short axis of the ellipse.

-A matching criteria based on the minimization of the elliptical ratio between the two contours and maintain equal area for

both the contours is used to determine the best long and short axis.

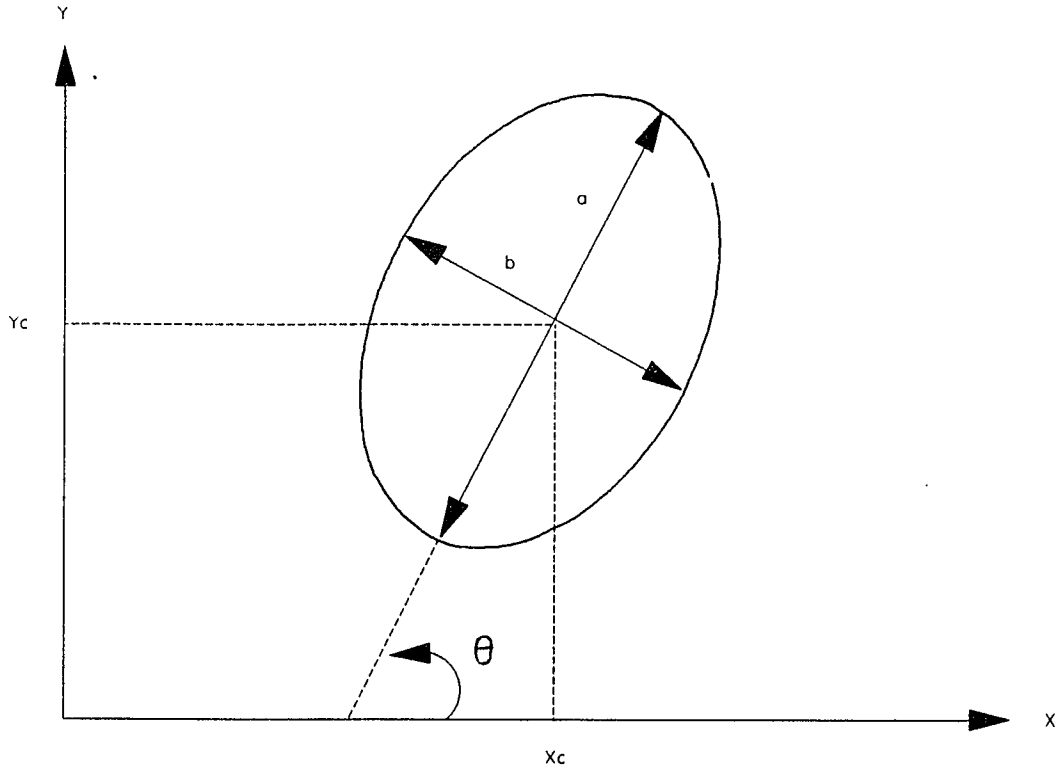


Figure 4.1 Representation of an ellipse with basic parameters (i.e. long axis "a", short axis "b", orientations "θ" and the center coordinates Xc,Yc).

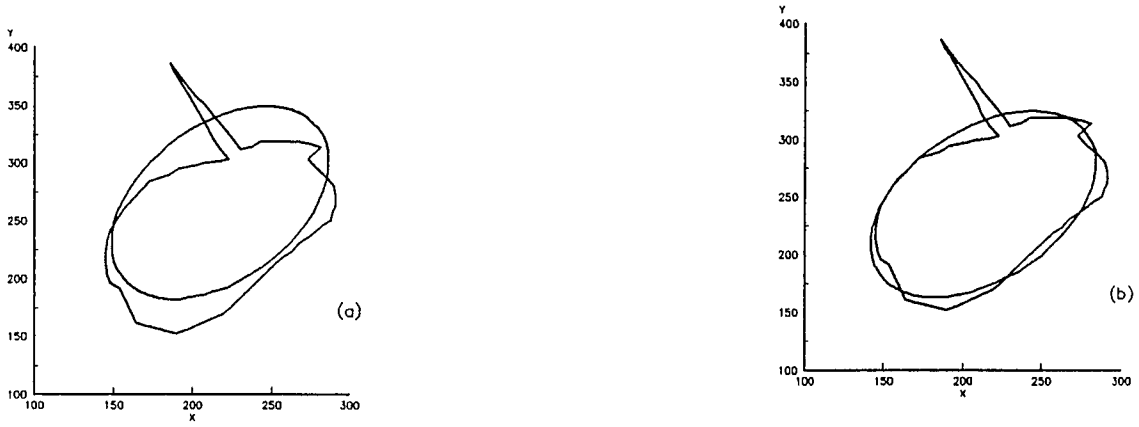


Figure 4.2 Effect of error in centroid location for relatively irregular shape the center was found

(a) by eliminating the first harmonic of the Fourier series and

(b) by geometric centering method.

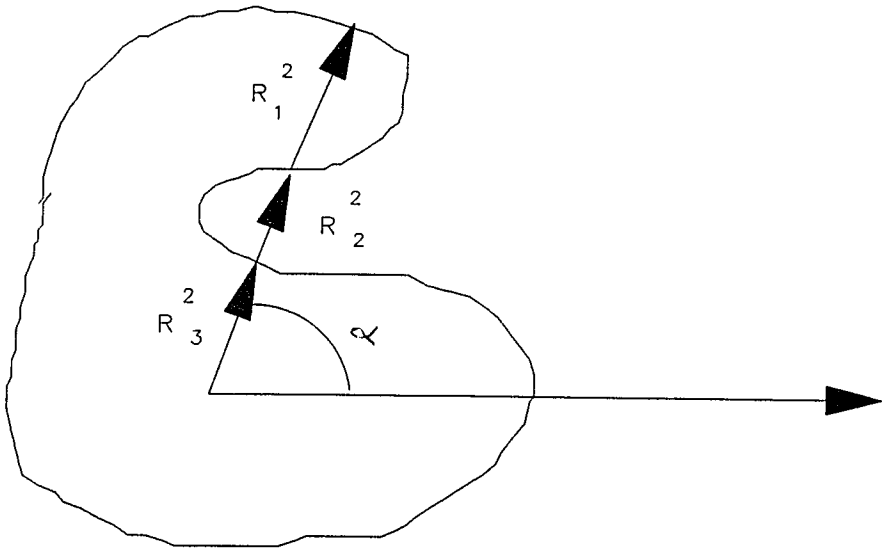


Figure 4.3 Representation of a multi–value contour

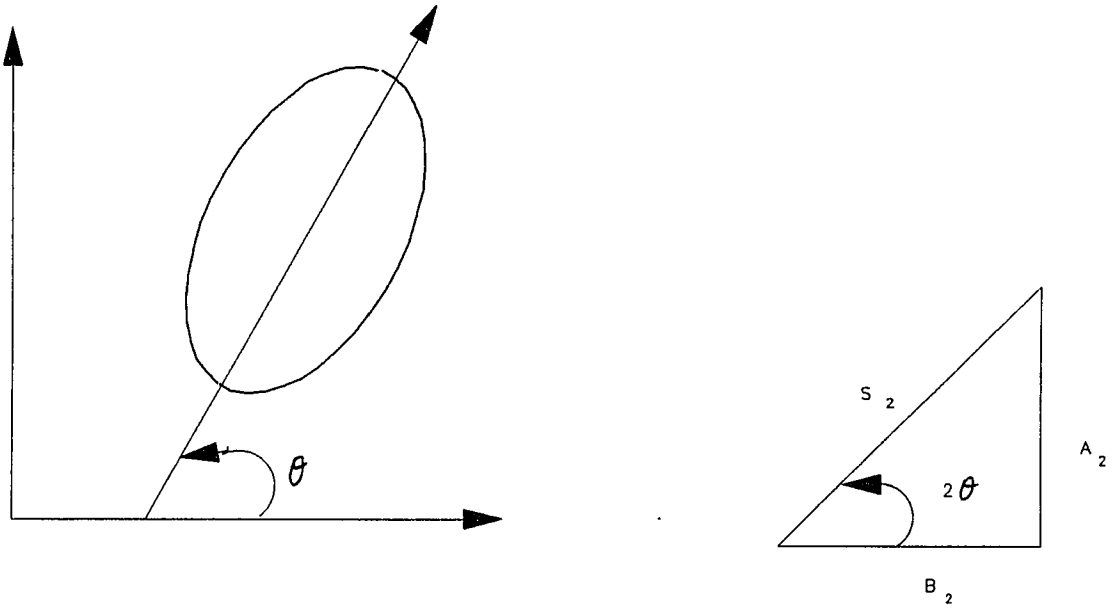


Figure 4.4 Relationship between the orientation of the ellipse and the Second harmonic.

Where s_2 is the second harmonic of the fourier series of the ellipse.

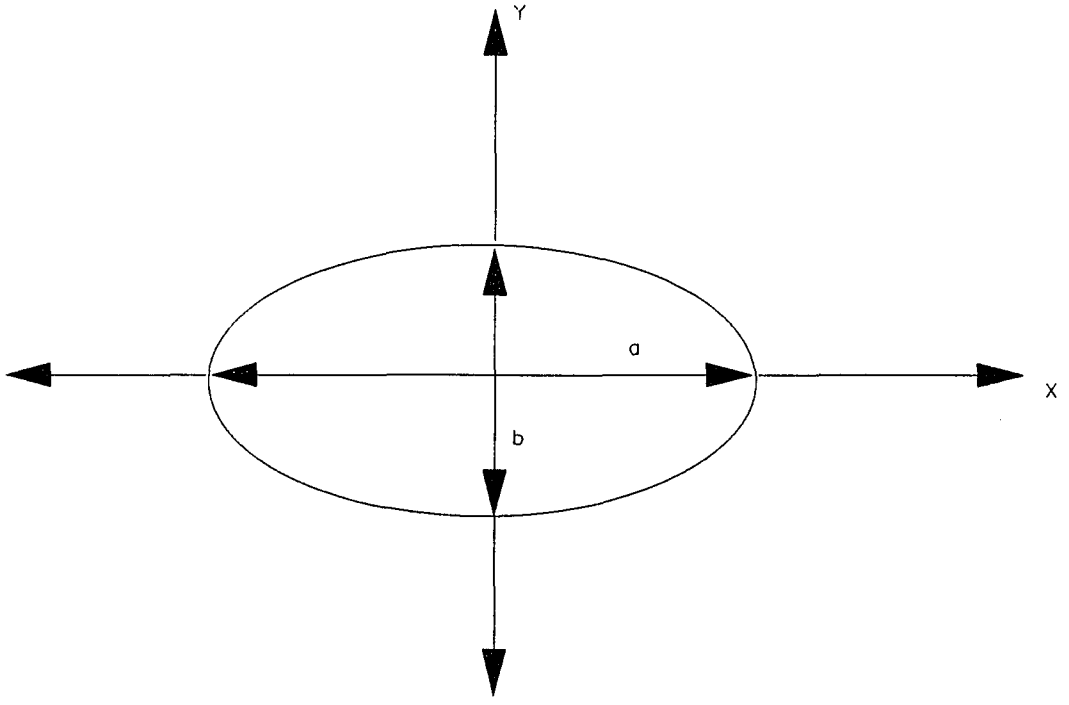


Figure 4.5 Representation of a perfect elllipse with zero orientations and centered at the origin.

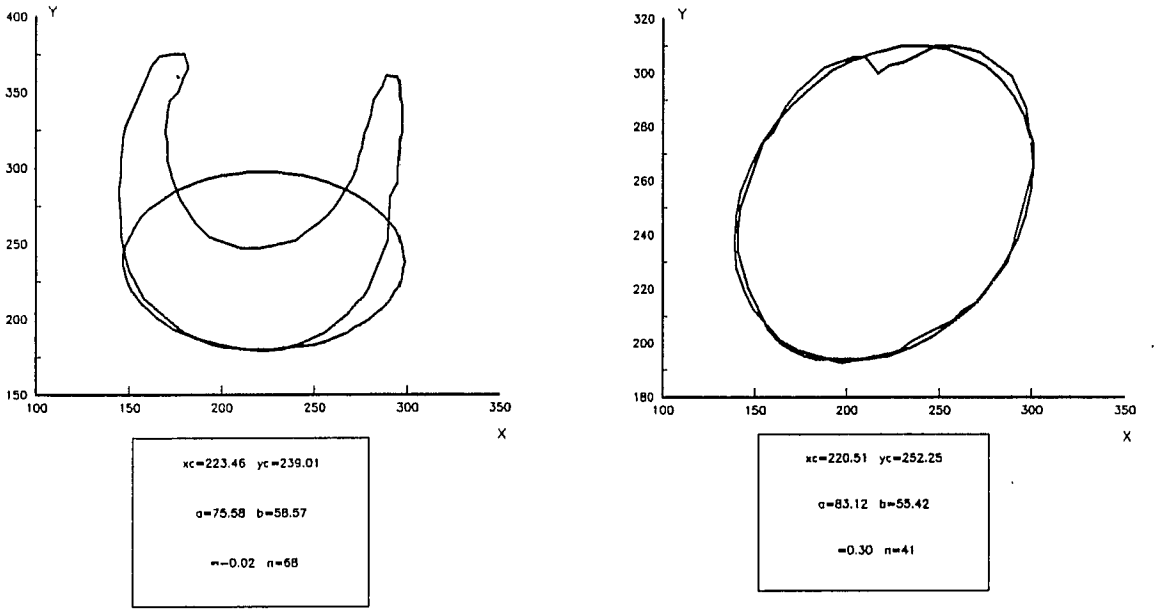


Figure 4.6 Representation of the matched ellipse with two different contours.

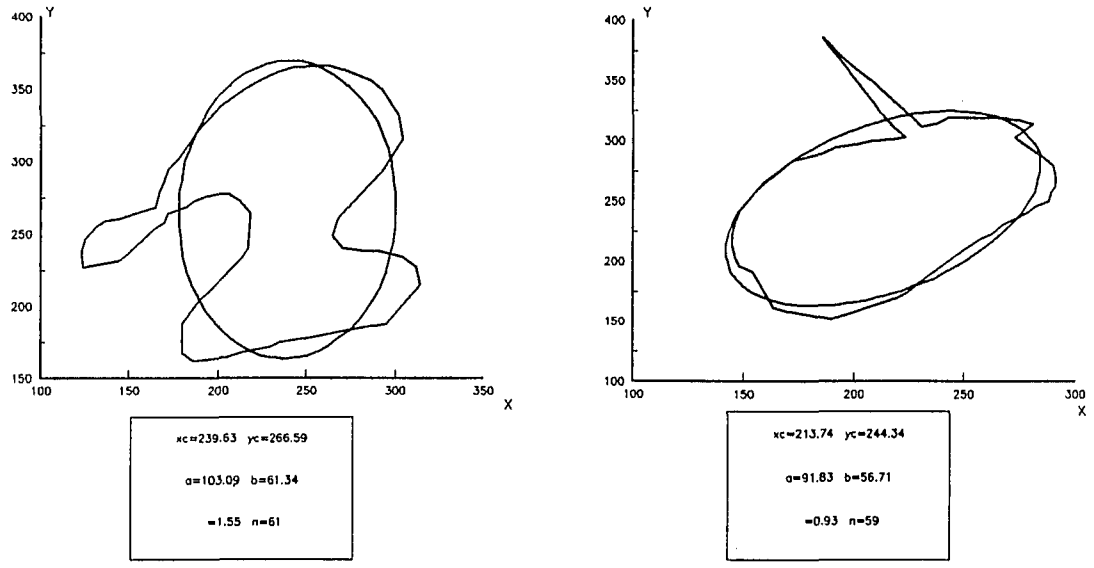


Figure 4.7 Representation of the matched ellipse with two different contours.

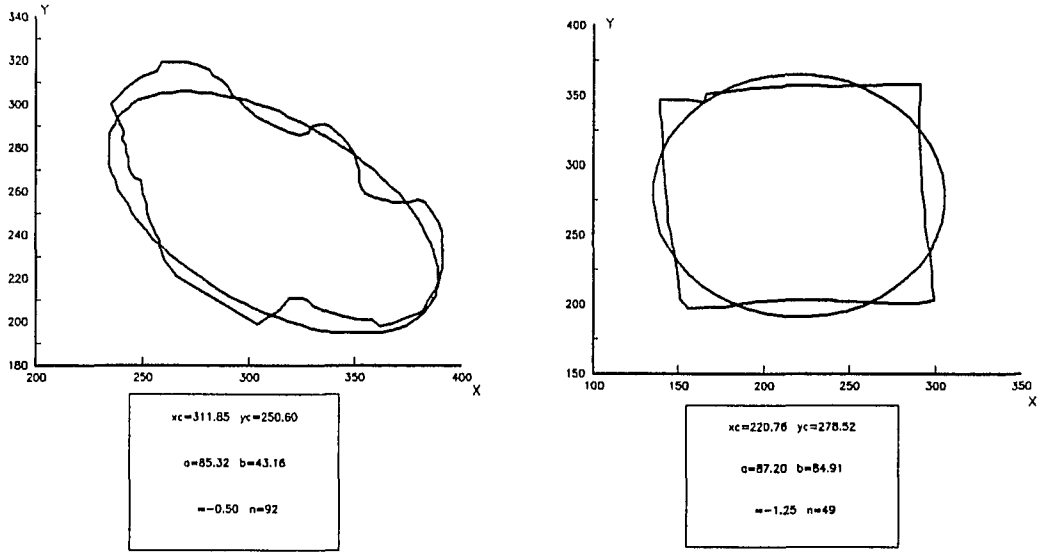


Figure 4.8 Representation of the match ellipse with two different contours.

CHAPTER 5

COMPARISON BETWEEN FOURIER AND GRID SEARCH METHOD

In this chapter we will compare our Fourier method against the Grid Search method. Both methods were implemented in C programing language on IBM 386 PC and were ported to a SUN work station on the Unix platform. The algorithms were tested for overall computional speed and accuracy of individual ellipse parameters.

Computational Time:

The main disadvantage of the Grid Search method is the computational speed; because the variation of the criterion function with different parameters depends on each other, the criterion function converge very slowly toward the minimum. The Fourier method is very efficient in terms of computational speed; because the parameters for the matched ellipse are determined separately.

Table[5.1] shows computational time taken on Sun work station for different contours by both algorithms

Contours	Computational time in seconds		Fourier to Grid Search Ratio
	Fourier method	Grid Search method	
T1	0.2	5.3	1:26
T2	0.2	7.5	1:37
T3	0.2	7.0	1:35
T4	0.1	6.1	1:61
T5	0.3	7.4	1:25
T6	0.2	6.4	1:21
T7	0.3	6.4	1:21
T8	0.3	6.8	1:23
T9	0.2	7.4	1:37
T10	0.3	7.4	1:25

Table[5.1] Comparison of computational time between Grid-Search and Fourier methods.

As observed from table[5.1], the Fourier method is faster than the Grid Search method. Cosidering the above 10 contour files, the average computational time for the Fourier method to Grid Search method was 1:31. It would take on the average 30 minutes for a given contour, to run the Grid Search method on PC plate form (i.e. IBM 386 with 20 MHZ) making it unusable in an interactive clonical environment.

Accuracy:

Cosidering the Grid Search method as the optimum best fit ellipse matching technique in terms of area, our Fourier method was tested against it. Table[5.2] shows the resulted ellipse parameters for 10 contours. The error was calculated separately for each parameters, and turned out to be under 10 percent for xc, yc,a and b. The error in θ in some cases is greater than 10%, that's because when the contour is very circular instead of elliptical, where orientation does not make any difference.

For an irregular contour, as shown in figure[5.5], the Grid Search method may converge to an uncorrect local minima while the Fourier method always conveges to the global minima.

CONTOURS	ELLIPSE PARAMETERS CALCULATED BY FOURIER METHOD					ELLIPSE PARAMETERS CALCULATED BY GRID SEARCH METHOD					ERROR				
	XC	YC	θ	A	B	XC	YC	θ	A	B	%XC	%YC	%θ	%A	%B
T1	221	252	.301	83.1	55.4	220.9	252.6	.32	82.8	55	.04	.23	5.9	.36	.72
T2	223.9	242.4	6.27	75.6	58.1	220	235.9	6.08	68.7	54.4	4.13	2.7	3.23	9.95	6.8
T3	215.3	248.8	.931	91.6	56.8	213.8	241.3	.933	91	56.4	.7	3.13	.11	.59	.58
T4	234.5	233	.665	168.4	20.4	230	232.4	.658	161.4	19.1	1.98	.3	.03	4.14	6.9
T5	236.1	265.7	1.55	102.5	61.38	244	269	1.45	99.7	61.9	3.24	1.22	7.22	2.8	.99
T6	219	177.5	5.04	87.2	84.9	219	278.6	6.28	82.2	84.1	.009	.39	19.6	6.17	.93
T7	311.8	252.7	2.63	85.8	42.55	310.9	254.3	2.6	86.98	41.6	.28	.6	.74	1.3	2.23
T8	294.2	183.9	1.44	77	29.4	294.1	180.7	1.36	77.5	27.7	.02	1.72	5.8	.62	6.1
T9	241.8	276.3	5.24	83.65	20.12	250.4	262	5.23	82.6	18.9	3.41	2.03	.19	1.26	6.38
T10	278.8	188	.267	88.2	62	278.4	191.4	.28	85.8	66.2	.14	1.76	4.28	2.8	6.38

TABLE[5.2] COMPARISON OF ERROR IN FIT BETWEEN GRID-SEARCH AND FOURIER METHODS.

The correlation of each parameter for both method are shown in graphs[5.1-5]

Contours	Area error by Fourier Method in %	Area error by Grid-Search Method in %
T1	2.28	1.5
T2	19.28	19.6
T3	7.5	5.5
T4	10.7	8.6
T5	7.7	9.4
T6	8.9	7.9
T7	10.3	9.6
T8	17.6	17.0
T9	9.0	8.1
T10	11.5	10.5

Table[5.3] Comparison of error in fit between Grid-Search and Fourier methods.

Table[5.3] shows the area error between the contour and the matched ellipse for different contours. The error were estimated by the following criterion function:

$$X^2 = \frac{(A_c + A_e - 2A_I)100}{2A_c}$$

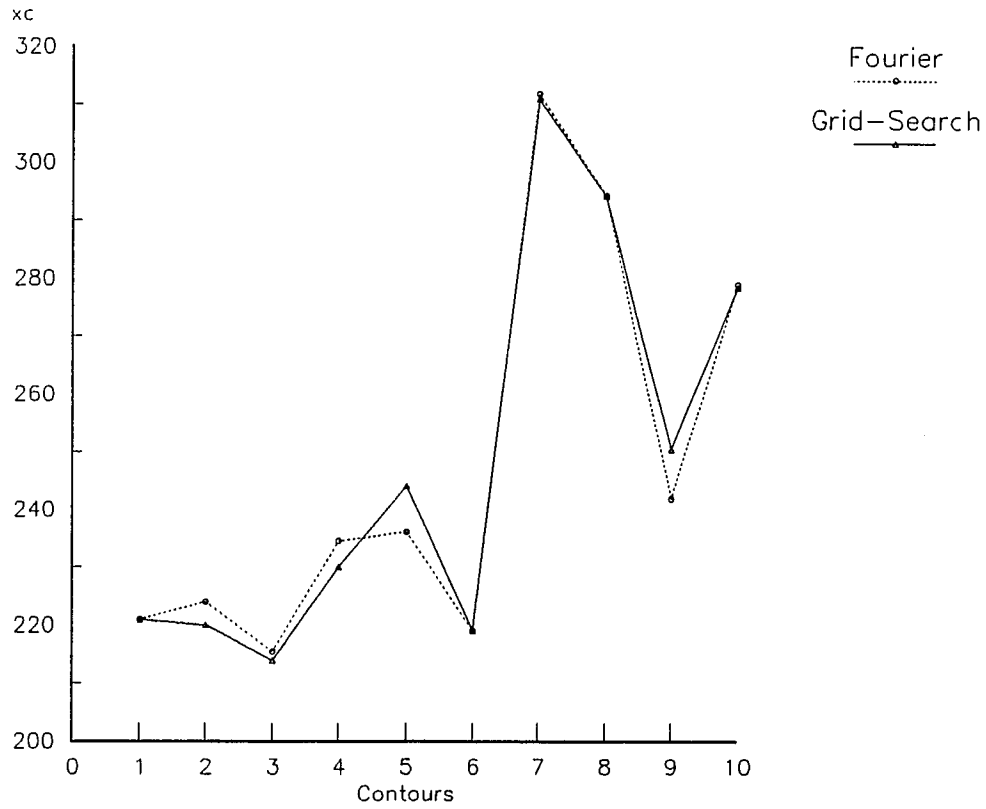
Where

A_c = Area of the given contour

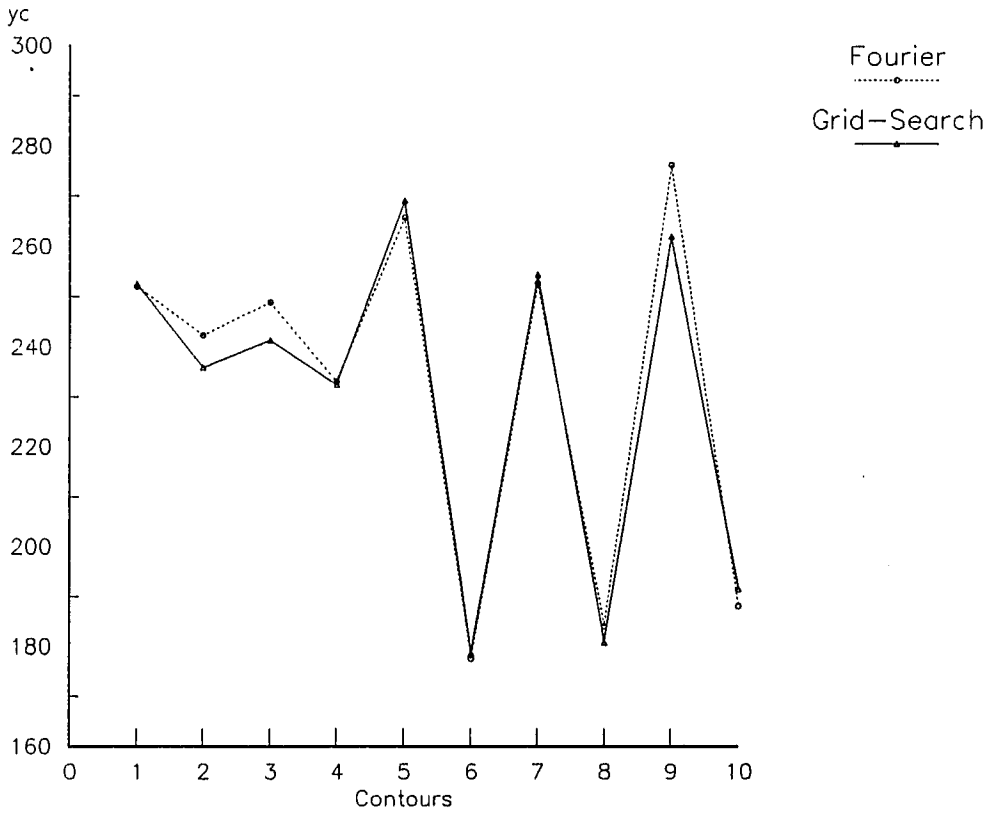
A_e = Area of the matched ellipse and

A_I = Area of the intersection between the ellipse and the contour as shown by the shaded region in figure[5.1].

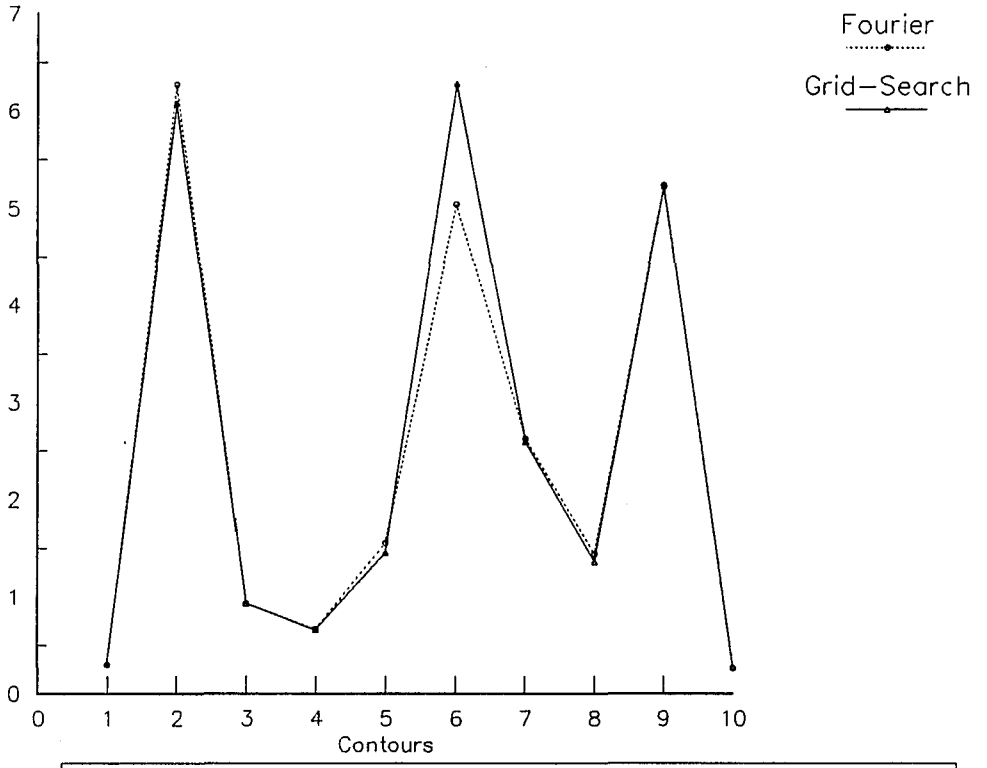
It is obvious from graph[5.6] that area errors for the 10 given contour are correlated for both algorithms. The error is minimum for elliptical contours and varies with the shape of the contour.



Graph 5.1 The correlation graph of parameter xc between Grid–Search and Fourier methods for 10 different Contours.

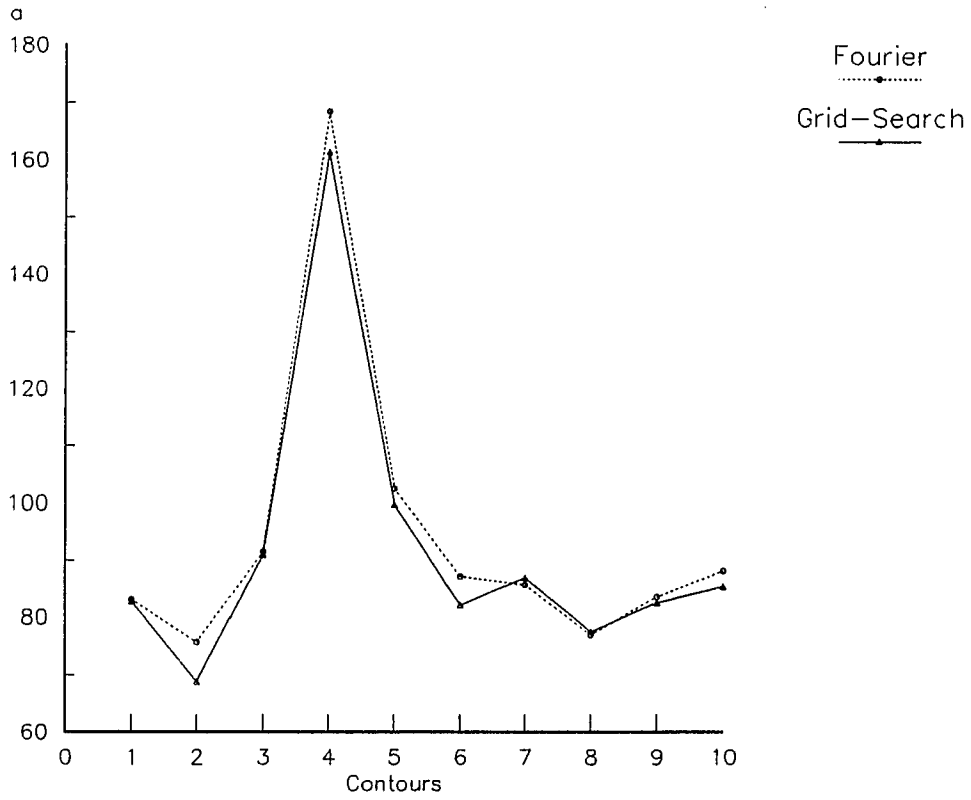


Graph 5.2 The correlation graph of parameter yc between Grid–Search and Fourier methods for 10 different Contours.

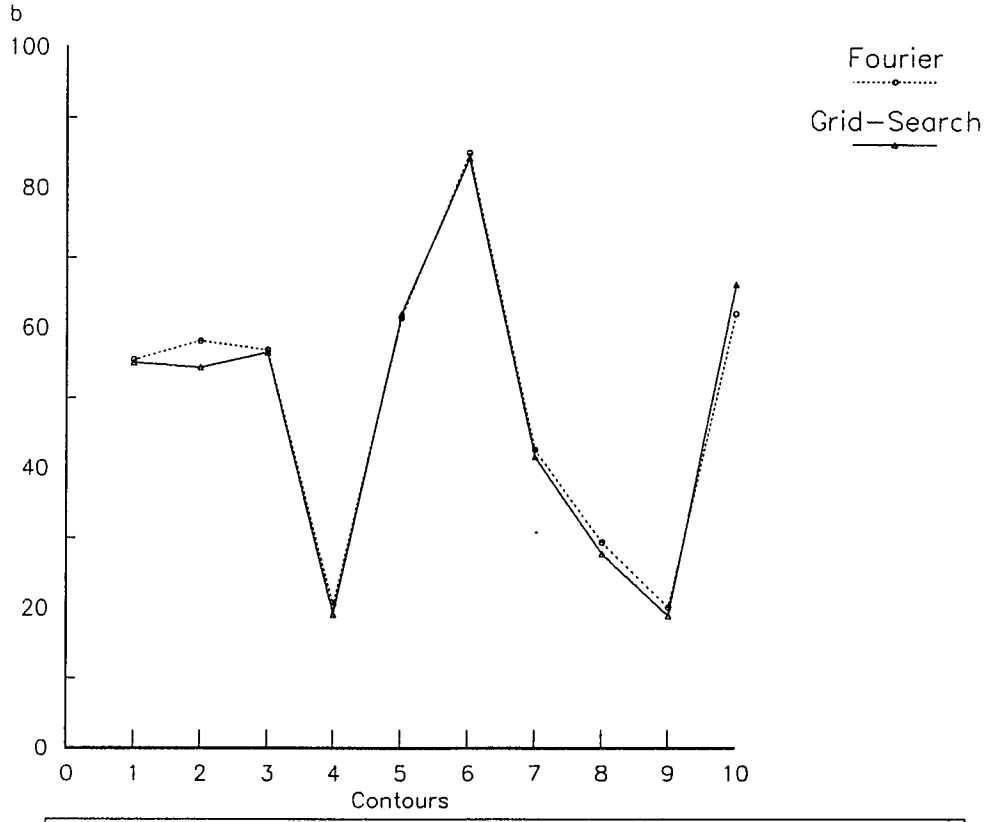


Graph 5.3

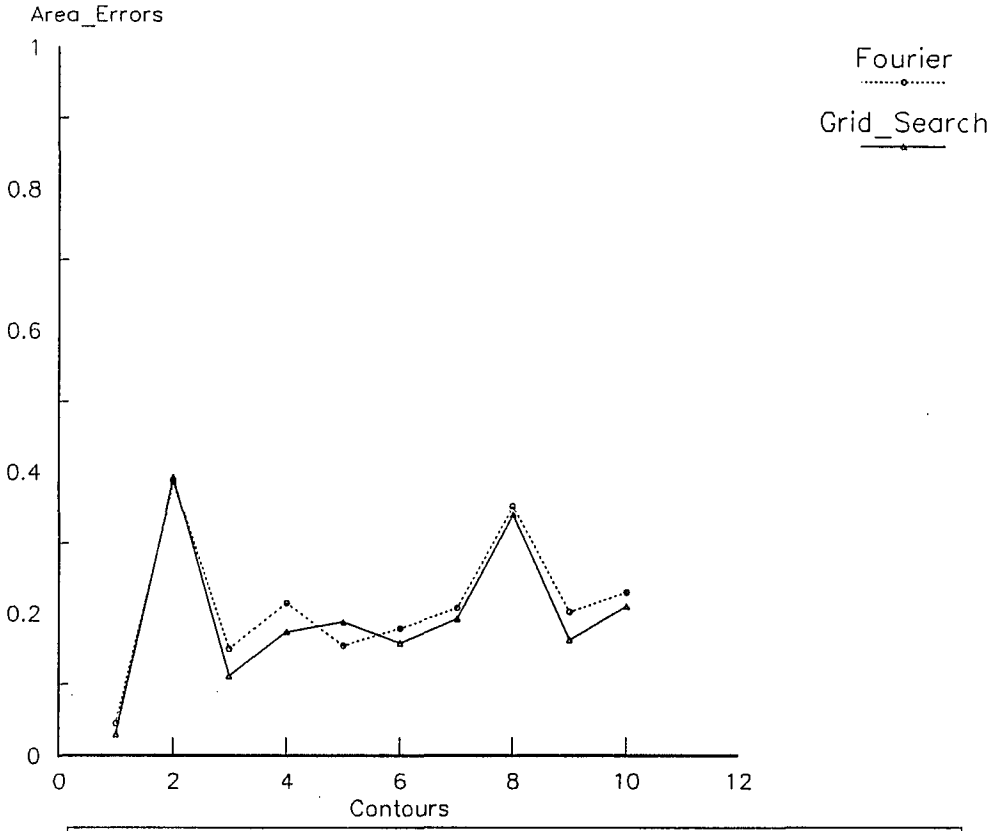
The correlation graph of parameter between Grid–Search and Fourier methods for 10 different Contours.



Graph 5.4 The correlation graph of parameter a between Grid–Search and Fourier methods for 10 different Contours.



Graph 5.5 The correlation graph of parameter b between Grid–Search and Fourier methods for 10 different Contours.



Graph 5.6 The area of non overlap regions (i.e area error) between the contour and matched ellipse w.r.t. the contour area for 10 different contours.

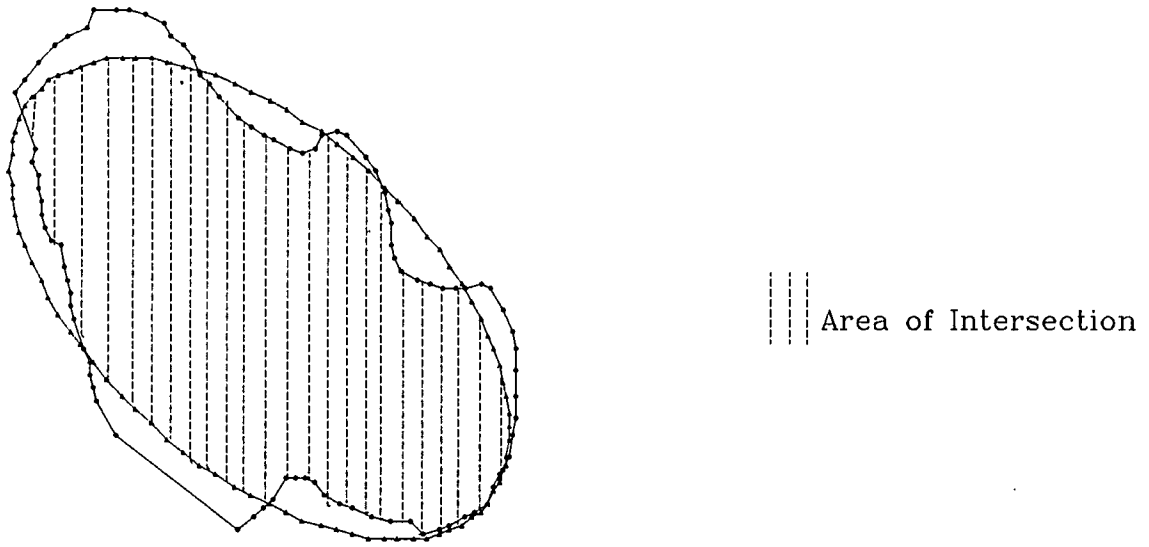


Figure 5.1 Area of intersection between the contour and ellipse

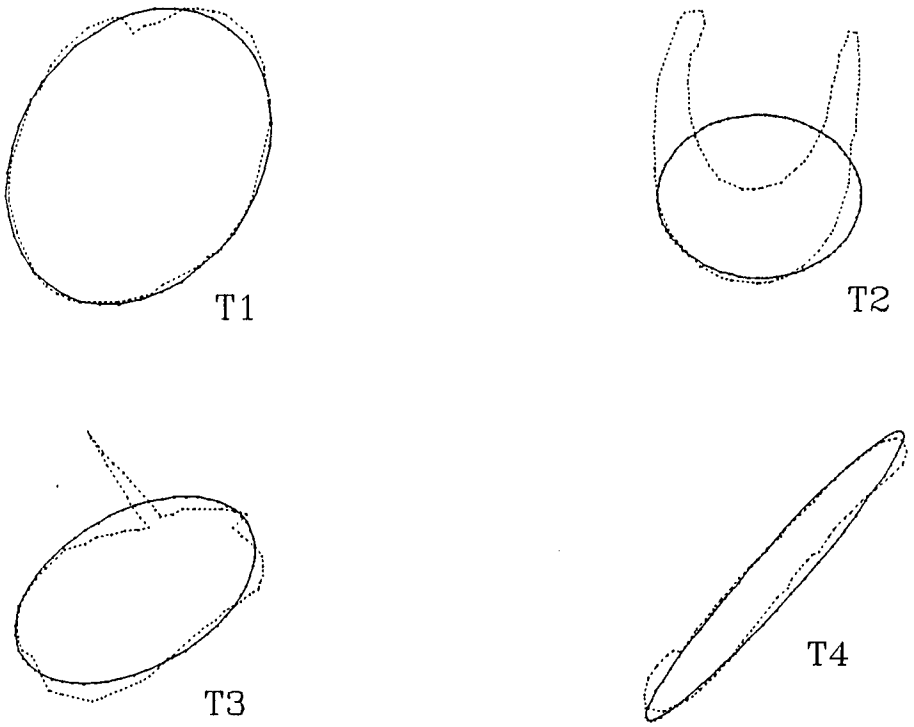


Figure 5.2 Representation of match ellipse with four different contours.

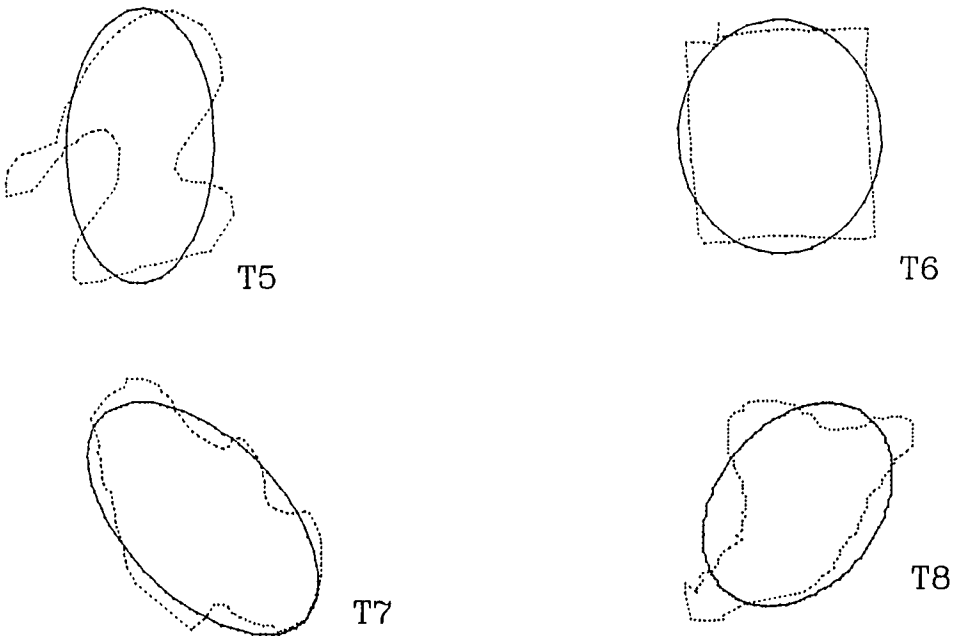


Figure 5.3 Representation of match ellipse with four different contours.

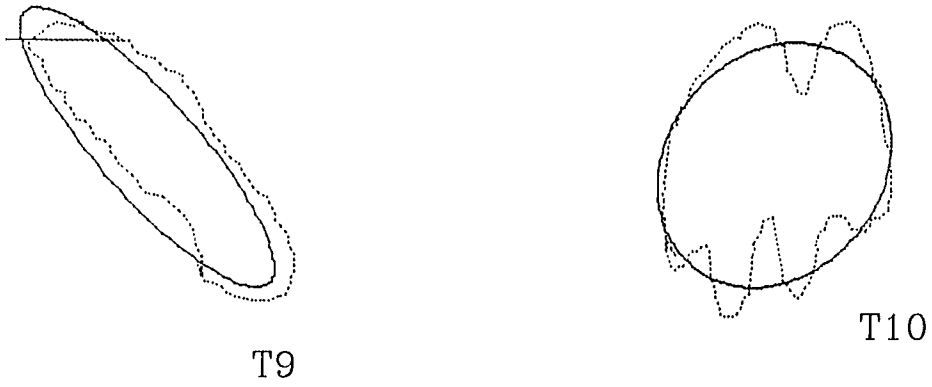
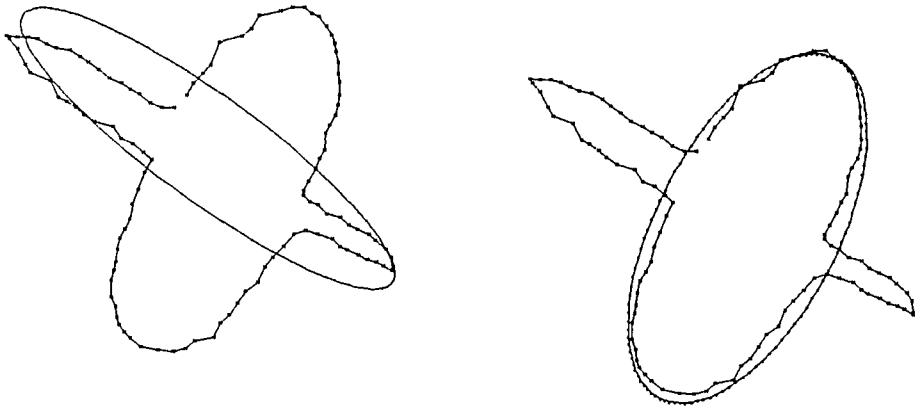


Figure 5.4 Representation of match ellipse with two different contours.



Grid Search

Fourier

Figure 5.5
For some irregular contours the Grid Search Method converges to local minima while Fourier method always converges to global minima.

CHAPTER 6

APPLICATIONS OF FOURIER METHOD TO THE DIAGNOSIS OF CANCER CELLS

In this chapter we will present some of the applications of our method for the diagnosis of cancer cells.

I) Diagnosis of Borderline and Invasive Serous Cystadenocarcinomas of the Ovary by Elliptical Analysis of the Nuclei

Whether borderline serous tumors (BST) of the Overy can be differentiated from invasive serous cystadenocarcinomas (ISC) by elliptical analysis of nuclei of the neoplastic epithelium was examined. Elliptical parameter were extracted from nuclei tracings and were evaluated in three borderline and three invasive tumors. After the extraction of these parameters for each sample of nuclei, they were statistically analyzed to determined the mean and standard deviation of these findings. Statistical analysis showed the internal consistancy of the two groups in terms of physical descriptors and the

discriminating values of the parameters. The results indicated that elliptical analysis of nuclei, combined with statistical analysis, could be used to distinguished between these tumors.

Borderline serous tumors (BST), or serous cystadenocarcinomas of low malignent potential, account for approximately 15% of malignent serous tumors of the Overy. Histologically they are characterized by a typical serous epithelium that does not invade the adjacent ovarian stroma. The prognosis of BST is substantially better than that of invasive tumors (ISC), and the treatment is different for these two groups of tumors. Therefore, accurate classification of ovarian serous tumors, currently based on demonstrating the presence or absence of invasion, has important implications. Thorough tumor sampling is required, and interpretive problems, such as presence of epithelial invaginations that closely simulate invasion, may be confronted. A study of interobserver variability in the interpretation of epithelial ovarian tumors has shown a high level of discordance between observers in the diagnosis of border line versus invasive tumors.

Tissues and Staining Technique:

Three cases of BST and three cases of ISC from the files of the department of Pathology of The Mount Sinai Hospital were selected at random. The samples were prepared by staining 5μm-thick paraffin embedded sections with a simple one-step silver staining method. The areas selected for study were considered histologically typical of each tumor type.

Image Analysis:

The image analysis system was built around an expansion box containing an Intel 310/17 multibus micro computer (Intel, Santa Clara, CA) based on the 16-bit 80286 microprocessor and an 80287 mathematics coprocessor. Additional image processing boards from Imaging Technology (Woburn, MA), including the analog processor AP-512 and frame buffers FP-512 were added. The image was generated by a tube color camera (Sony, Park Ridge, NJ), installed atop a light microscope (Nikon, Garden City, NY). The image was previewed on a high resolution color monitor, but the interactive peripheral was a touch-sensitive screen, placed on a 13-inch video screen. The morphometric procedures were interactive. The cells of interest were identified on the screen, and the contours of their nuclear

profile were traced manually. Randomlly 100 cell nuclei were traced from the selected regions of each sample and the x-y coordinates of the contours were stored in the output file.

Our best fit ellipse technique were then applied and ellipses were matched to 100 contours of each samples. The descriptors such as the long axis "a" and short axis "b" of the ellipse, the elipticity (i.e. a/b), the product of a and b (which is also proportional to the area of the contour) and the deviation of the ellipse from the contour interms of area were calculated for each contour. All the descriptors for 100 contours of each sample were the then statistically analyzed and the mean variance and standard deviations were determined. These descriptors were then evaluated for each case and identified the descriptors for correct classification. Each case was assigned one of the diagnoses under consideration (BST or ISC), based upon the homogeneity of the variables. The classification was done without prior knowledge of the pathologic diagnosis.

Cases	MER	SER	MEA	SEA	MDEC	SDEC	DIAGNOSIS
1	1.43	0.14	332.7	33.44	9.01	0.91	BST
2	1.39	0.14	328.7	33.04	10.05	1.01	BST
3	1.76	0.18	384.9	38.7	10.08	1.01	BST
4	2.32	.23	554.2	55.7	11.78	1.18	ISC
5	2.57	.26	591.4	59.4	13.32	1.34	ISC
6	2.03	0.20	498.5	50.11	11.46	1.15	ISC

Table [6.1] Comparing Borderline Serous Cystadenocarcinomas and Invasive Serous Cystadenocarcinomas case by case

BST: borderline serous Cystadenocarcinomas; ISC: invasive serous Cystadenocarcinomas; MER: mean elliptical ratio; SER: standard deviation of elliptical ratio; MEA: mean ellipse area; SEA: standard deviation of ellipse area; MDEC: mean deviation of ellipse from the contour; SDEC: standard deviation of ellipse from the contour.

The mean values obtained from the elliptical ratios, product of elliptical axis and the deviation of contour from ellipse, and their standard deviations are shown in table [6.1] and table[6.2]. All values were higher for ISC than for BST, the

most significant are the ellipticity and the product of elliptical axis (which is also proportional to the area of the contour).

Cases	MER	SER	MEA	SEA	MDEC	SDEC
BST (3 cases)	1.52	0.15	348.8	35.06	9.7	0.98
ISC (3 cases)	2.3	.23	548.03	55.07	12.12	1.2

Table [6.2] Summary of parameters Comparing Borderline Serous Cystadenocarcinomas and Invasive Serous Cystadenocarcinomas

Table [6.2] summarizes the discriminating parameters for the two group of tumors. Figure [6.1-3] depicts the distribution of each parameter.

Discrimination between BST and ISC classically depends on the presence or absence of invasion of the neoplastic epithelium into the ovarian storma. This distinction is important because

of the biologic differences between the two groups of tumors, reflected in prognostic and treatment differences. The 10 year survival rate for BST is approximately 95%, and no therapy after operative excision is indicated. However, the 5 years survival rate for ISC depends on the treatment (chemotherapy or radiation therapy) is indicated in addition to surgery.

A recent study, based on Feulgen-stained paraffin-embedded sections, suggested that the cytokinnetic properties of borderline tumors are intermediate between those of benign and malignant ones, it was concluded that quantitative morphometric evaluation can provide objective and reproducible data in the diagnosis of borderline malignancy. Our result show that elliptical parameter of nucleus can be used to discriminate between ISC and BST independently of the classic criterion of stormal invasion. In particular, objective evaluation of the ellipticity of nucleus and the size (product of the long and short axis of the matched ellipse) of nucleus provided a sufficient basis for such distinction. The clear-cut quantifiable differences between BST and ISC the we found may provide the basis for developing a comprehensive data base to be used for automated morphometry-based diagnostic classification of these tumors. Morphometric analysis thus become a useful adjunct

in classifying cases with ambiguous histology.

II) Diagnosis of Cervical Condyloma of Human Papilloma Virus (HPV) by Elliptcal Analysis of Cervical Lesion's Nuclie

Human papilloma virus (HPV) infection of the uterine cervix is associated with lesions including warts or Condyloma and dysplasias. Infected superficial cervical epithelial cells can show a viral cytopathic effect termed koilocytosis. Such cells, or koilocytes, show the following characteristic features.

i) a para-nuclear cytoplasmic halo; or clearing;

ii) nuclear enlargement

iii)contour irregularity, or wrinkling of the nuclear membrane

iv) nuclear hyperchromasia, or darkening of the nuclear chromatin.

Cervical condyloma is the disease of young women, with a large population impact and risk factors characteristic of a sexually transmitted disease. More than 65% of patients with genital Condyloma are aged 15-29 years, the highest risk group being in the 20-24 year range in both male and females.

Nine cases of the Cervical lesions from the files of the

department of Pathology of The Mount Sinai Hospital were selected at random. Average of 50 cell nuclei contours were extracted using the imaging system discussed earlier for each case. Our Fourier method were then applied to extract the elliptical parameters from each nucleus contour. The parameters which were discriminating between the normal and Cancerous cells were then statistically analyzed and are tabulated in tables [6.3-4].

Cases	MER	SER	MEA	SEA	MDEC	SDEC	DIAGNOSIS
1	1.44	.21	256.09	36.58	6.13	.90	Normal
2	1.26	.18	287.36	41.05	6.10	.87	Normal
3	1.35	.19	153.18	21.66	5.40	.76	Normal
4	1.21	.17	266.53	37.69	3.84	.54	Normal
5	1.32	.16	293.10	36.64	4.49	.56	Normal
6	2.49	.12	545.07	77.09	10.90	1.54	Condyloma
7	2.07	.09	652.63	92.30	8.43	1.19	Condyloma
8	2.66	.14	404.87	57.84	9.95	1.42	Condyloma
9	2.57	.13	598.42	85.49	11.34	1.62	Condyloma

Table [6.3] Comparing Normal and Condyloma for each case.

MER: mean elliptical ratio; SER: standard deviation of elliptical ratio; MEA: mean ellipse area; SEA: standard deviation of ellipse area; MDEC: mean deviation of ellipse from the contour; SDEC: standard deviation of ellipse from the contour.

Cases	MER	SER	MEA	SEA	MDEC	SDEC
Normal (5 cases)	1.31	0.18	251.25	34.72	5.23	0.73
Condyloma (4 cases)	2.44	0.34	550.25	78.18	10.15	1.44

Table [6.4] Summary of parameters Comparing Normal and Condyloma

MER: mean elliptical ratio; SER: standard deviation of elliptical ratio; MEA: mean ellipse area; SEA: standard deviation of ellipse area; MDEC: mean deviation of ellipse from the contour; SDEC: standard deviation of ellipse from the contour.

The mean values obtained from the elliptical ratios, product of elliptical axis and the deviation of contour from ellipse, and their standard deviations are depicted graphically in figure[6.4-7]. All values were higher for Condyloma than for Normal, the most significant are the ellipticity and the product of elliptical axis (which is also proportional to the area of the contour).

Illustration [6.1] Borderline Serous Tumor (BST) of Ovary

Illustration [6.2] Profile of nucleus and the matched ellipse of (BST)

Illustration [6.3] Invasive Serous Cystadenocarcinomas (ISC) of Ovary

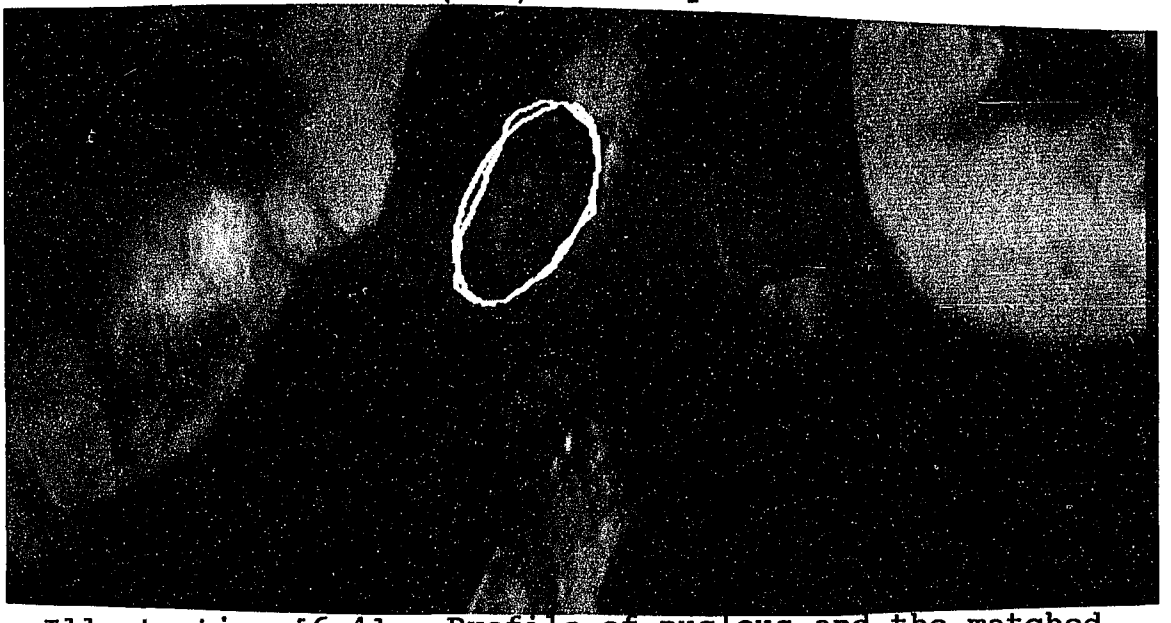


Illustration [6.4] Profile of nucleus and the matched ellipse of (ISC)

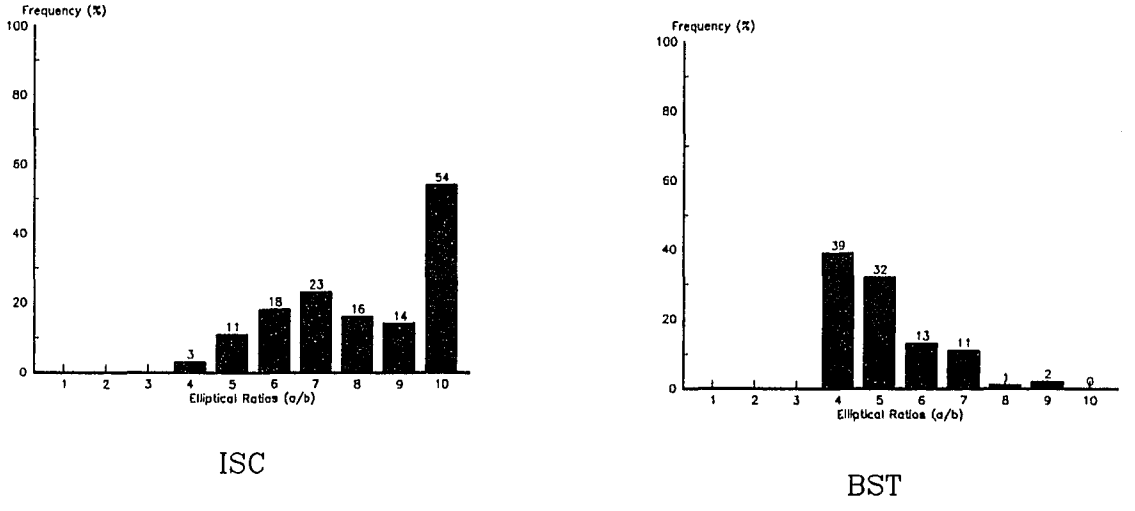


Figure 6.1 The distributions of ellipticity for borderline and invasive tumors

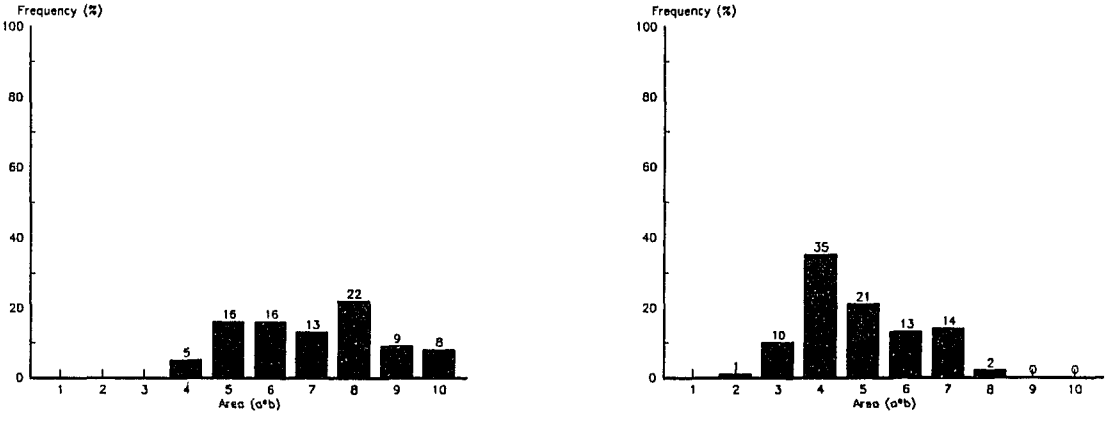


Figure 6.2 The distribution of area for borderline and invasive tumors

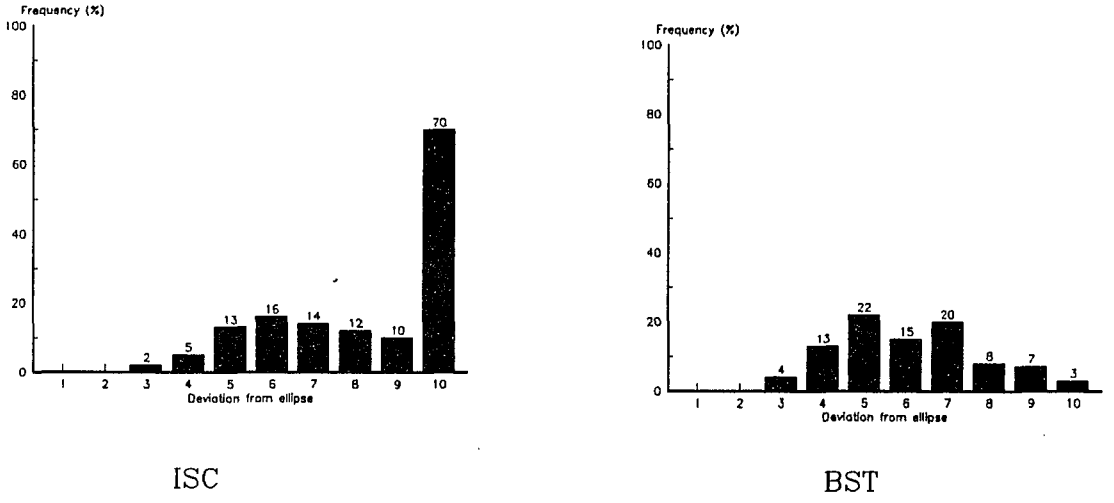


Figure 6.3 The distributions of deviation from ellipse for borderline and invasive tumors

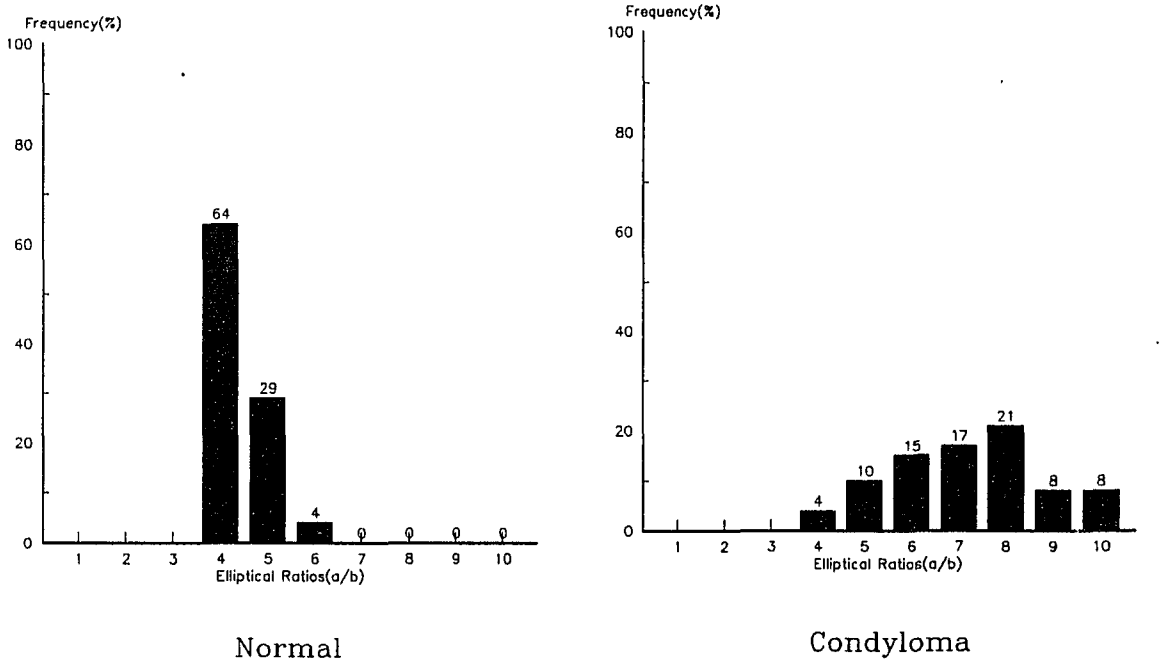


Figure 6.4 The distributions of Elliptical ratio for normal and condyloma cases

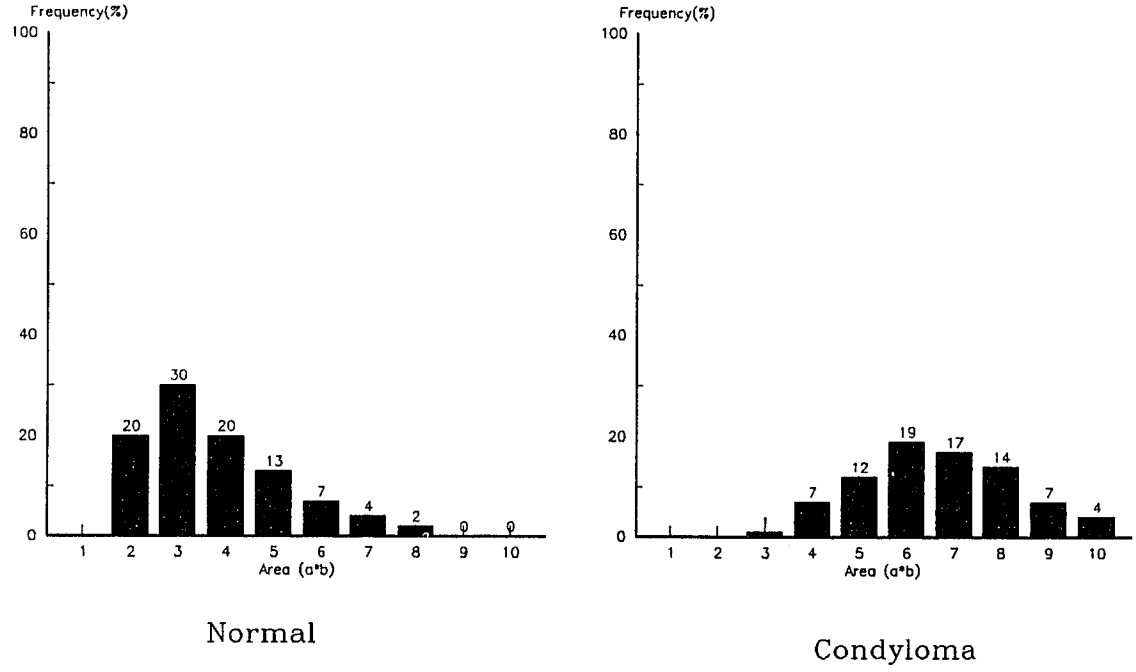


Figure 6.5 The distributions of area(a*b) for normal and condyloma cases

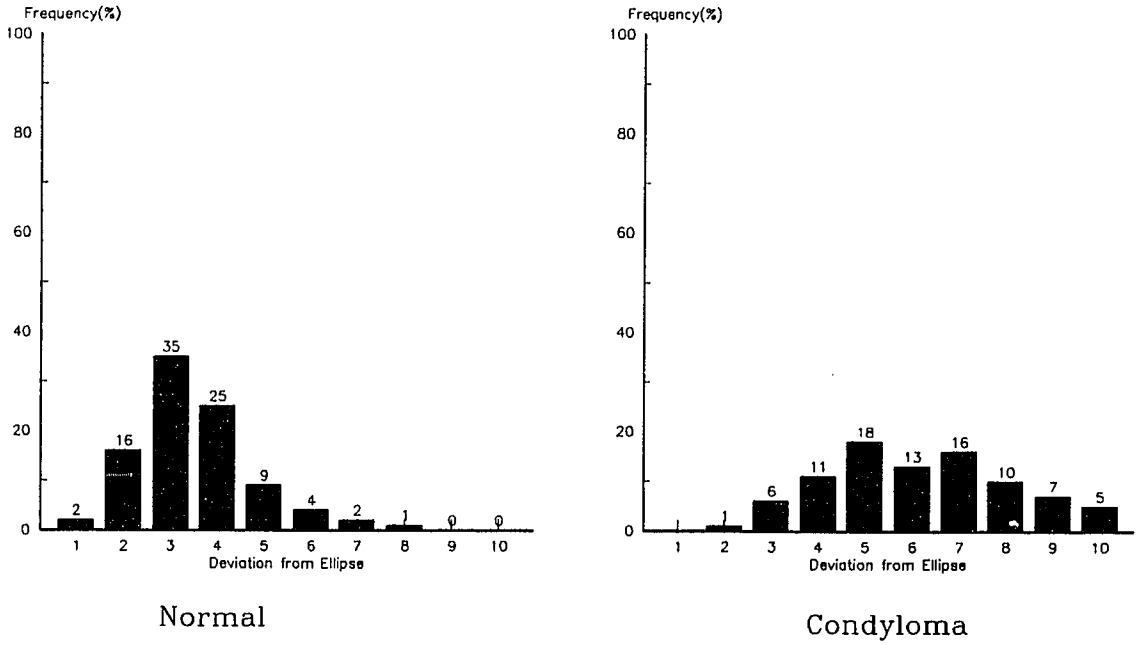
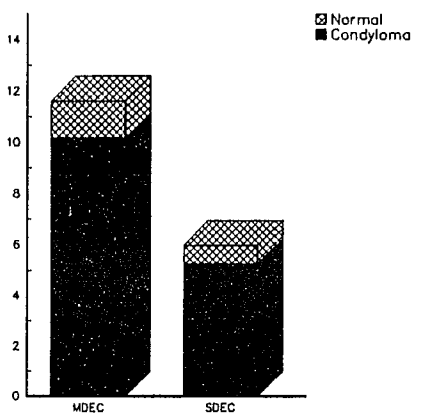
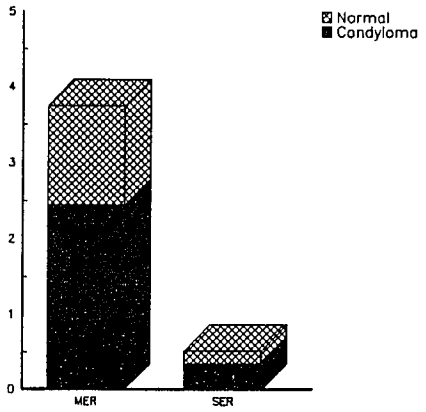


Figure 6.6 The distribution of parameter DEC (deviation of ellipse from contour)





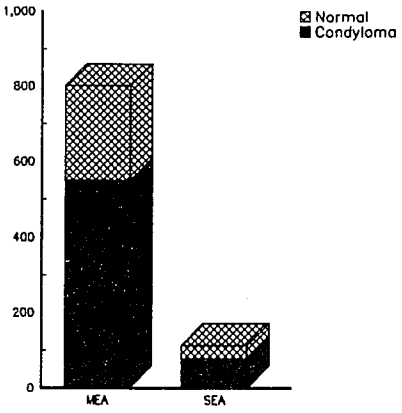


Figure 6.7 Representation of mean and standard deviation of each parameter for normal and condyloma

REFERENCES

[1] Barnsley, M.F., Fractal Everywhere. Academic Press, 1988.

[2] Barnsley, M.F. And Sloan, A.D., "A Better Way to Compress Images". Byte Jan. 1988.

[3] Bevington, PR (1969) Data Reduction and error analysis for the physical sciences. New York: McGraw-Hill, 208

[4] Boon, j.d., Evans, D.A. And Hennigar, H. f., 1984. Spectral information from fourier analysis of digitized quartz grain profiles. Math. Geol., vol.14, P 589-605.

[5] Bouman, C., Et Al., 1988. "Segmentation of texture images using a multiples resolution approach". ICASSP Proceedings, 1988.

[6] Bowie JE, et al,(1977). An analysis technique for biological shape. II. Acta Cytol(Baltimore) 21:455-64

[7] Bowie JE, et al,(1977). An analysis technique for biological shape. III. Acta Cytol(Baltimore) 21:739-746

[8] Bykat, A., 1978. "Convex Hull of a finite set of points in two dimensions" Information processing letters. vol. 7, no. 6, oct. 1978.

[9] Canny, J. "A computational approach to edge detection". IEEE trans. on Patt. Ana. Machine Int. vol. PAMI-8, no. 6, Nov. 1986.

[10] Diamond DA, et al, (1982). Computarized image analysis of Nuclear shape as a prognostic factor for Prostatic Cancer. Prostate 3: 321-32

[11] Frisch, et al, 1987. Shape discrimination of sand sample using fractal dimension. Proc. coastal sediments, 1987 conf., New Orleans, A.S.C.E., p 138-153.

[12] Ehrlich, R., Brown, P. J., 1980. The origin of shape frequency distributions and the relationship between Size and Shape. Journal of Sedimentary Petrology, vol. 50 No. 2, June 1980, P. 475-485.

[13] Freeman, H, Glass JM, On the quantization of line-drawing data, IEEE Trans. Syst. Sci. Cybern. vol. SSC-5, 70-79, 1969.

[14] Freeman, H, Boundary encoding and processing, in Picture Processing and Psychopictorics. Eds Lipkin BS, and Rosenfeld A, 241-263, Academic Press, New York, 1970.

[15] Fung, F.W., Et al., 1988. "Automatic segmentation of biomedical images". IEEE ASSP, Proceding, 1988.

[16] Gonzalez, R.C., Wintz, P., 1987. Digital Image Processing. Addison Wesley.

[17] Graham, R.L. An efficient algorithm for determining the convex hull of a finite set. Information Processing Lett. 1, no 4(1972), 132-133.

[18] Gray, S.B., " Loacl Properties of Binary Images in two dimentions," IEEE Transaction on Computers, vol. C-20, no. 5, May 1971, pp. 551-561.

[19] Gschwind, R. Et al, Evaluation of shape descriptors for the Morphometric analysis of Cell Nuclei. Path. Res. Pract. 181, 213-222(1986).

[20] Guyer, D.E. Et al, 1986. Machine vision and image processing for plant Identification, Transaction of the ASAE, 29(6): 1500-1507.

[21] Guyer, D.E., 1988. Application of machine vision to human shape analysis. Ph.D thesis, Purdu University graduate school, 1988.

[22] Ingram, M., Preston, K. Jr., "Automatic analysis of blood cells", Scientific American, vol. 223, No. 5, pp. 72-82, Nov. 1970.

[23] Keng, J.," A syntax-directed method for land use classification of LANDSAT images," Proceeding of the symposium on current mathematical problems in image science, Monterey, CA, 261-265, 1976.

[24] Kruger, R.P., "Computer diagnosis of pneunoconiosis," IEEE Trans., man and Cybernetics, vol 4, no. 1, 1974.

[25] Kulpa, Z, Area and Perimeter measurement of blobes in discrete binary pictures. Computer graphics and image processing, no. 6, p434-451, 1977.

[26] Levine, Martin D., 1985. Vision in Man and Machine. McGraw-Hill.

[27] Mandelbrot, B. B., 1977. "Form, chances and dimension", W. H. Freeman, San Francisco, 365p.

[28] Parent, et al, 1989. "Trace inference, curvature consistency, and curve detection", IEEE tran. On pat. ana. and machine int., vol. 11, no. 8, Aug. 1989.

[29] Rosenfeld, A., Picture Processing: 1977, Computer Graphic and image Processing, vol. 7, pp. 211-242, 1978.

[30] Rosenfeld, A., And A.C.Kak 1982. Digital Picture Processing, Academic press, New York.

[31] Sklansky, J., Measuring concavity on rectangular mosaic. IEEE Trans. Computers C-21 12 (1972) 1355-1364.

[32] Wintrobe, M. M., Clinical Hematology, Ed. 7, Lea and Febiger, Philadelphia, 1974.

[33] Yasnoff, W.A., Mui, J. K., Bascus, J. W., "Error measures for scene Segmentation," Pattern Recognition, vol. 9, pp. 217-231, 1977.

[34] Young IT, et al,(1974). An analysis technique for biological shape. I. Inform. Control 25: 357-70

[35] Mittal, R.K., Cha, W. "Sensitivity and Specificity of Various Morphological Features of Cervical Condylomas" Arch. Pathol. Lab Med., Vol 114, October 1990.

[36] Anderson, Et al, "Current Views on Cervica intraepithelial neoplasia", Journal of Pathology 1991; 44:969-978