Modeling and Analysis of Shape

with Applications in Computer-Aided Diagnosis of Breast Cancer

Copyright © 2011 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Modeling and Analysis of Shape with Applications in Computer-Aided Diagnosis of Breast Cancer

Denise Guliato and Rangaraj M. Rangayyan

www.morganclaypool.com

ISBN: 9781608450329 paperback
ISBN: 9781608450336 ebook

DOI 10.2200/S00325ED1V01Y201012BME039

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON BIOMEDICAL ENGINEERING

Lecture #39
Series Editor: John D. Enderle, *University of Connecticut*
Series ISSN
Synthesis Lectures on Biomedical Engineering
Print 1930-0328 Electronic 1930-0336

Synthesis Lectures on Biomedical Engineering

Editor
John D. Enderle, ***University of Connecticut***

Lectures in Biomedical Engineering will be comprised of 75- to 150-page publications on advanced and state-of-the-art topics that spans the field of biomedical engineering, from the atom and molecule to large diagnostic equipment. Each lecture covers, for that topic, the fundamental principles in a unified manner, develops underlying concepts needed for sequential material, and progresses to more advanced topics. Computer software and multimedia, when appropriate and available, is included for simulation, computation, visualization and design. The authors selected to write the lectures are leading experts on the subject who have extensive background in theory, application and design.
The series is designed to meet the demands of the 21st century technology and the rapid advancements in the all-encompassing field of biomedical engineering that includes biochemical, biomaterials, biomechanics, bioinstrumentation, physiological modeling, biosignal processing, bioinformatics, biocomplexity, medical and molecular imaging, rehabilitation engineering, biomimetic nano-electrokinetics, biosensors, biotechnology, clinical engineering, biomedical devices, drug discovery and delivery systems, tissue engineering, proteomics, functional genomics, molecular and cellular engineering.

Modeling and Analysis of Shape with Applications in Computer-Aided Diagnosis of Breast Cancer
Denise Guliato and Rangaraj M. Rangayyan
2011

Analysis of Oriented Texture with Applications to the Detection of Architectural Distortion in Mammograms
Fábio J. Ayres, Rangaraj M. Rangayyan, and J. E. Leo Desautels
2010

Fundamentals of Biomedical Transport Processes
Gerald E. Miller
2010

Models of Horizontal Eye Movements, Part II: A 3rd Order Linear Saccade Model
John D. Enderle and Wei Zhou
2010

Models of Horizontal Eye Movements, Part I: Early Models of Saccades and Smooth Pursuit
John D. Enderle
2010

The Graph Theoretical Approach in Brain Functional Networks: Theory and Applications
Fabrizio De Vico Fallani and Fabio Babiloni
2010

Biomedical Technology Assessment: The 3Q Method
Phillip Weinfurt
2010

Strategic Health Technology Incorporation
Binseng Wang
2009

Phonocardiography Signal Processing
Abbas K. Abbas and Rasha Bassam
2009

Introduction to Biomedical Engineering: Biomechanics and Bioelectricity - Part II
Douglas A. Christensen
2009

Introduction to Biomedical Engineering: Biomechanics and Bioelectricity - Part I
Douglas A. Christensen
2009

Landmarking and Segmentation of 3D CT Images
Shantanu Banik, Rangaraj M. Rangayyan, and Graham S. Boag
2009

Basic Feedback Controls in Biomedicine
Charles S. Lessard
2009

Understanding Atrial Fibrillation: The Signal Processing Contribution, Part I
Luca Mainardi, Leif Sörnmo, and Sergio Cerutti
2008

Understanding Atrial Fibrillation: The Signal Processing Contribution, Part II
Luca Mainardi, Leif Sörnmo, and Sergio Cerutti
2008

Introductory Medical Imaging
A. A. Bharath
2008

Lung Sounds: An Advanced Signal Processing Perspective
Leontios J. Hadjileontiadis
2008

An Outline of Informational Genetics
Gérard Battail
2008

Neural Interfacing: Forging the Human-Machine Connection
Thomas D. Coates, Jr.
2008

Quantitative Neurophysiology
Joseph V. Tranquillo
2008

Tremor: From Pathogenesis to Treatment
Giuliana Grimaldi and Mario Manto
2008

Introduction to Continuum Biomechanics
Kyriacos A. Athanasiou and Roman M. Natoli
2008

The Effects of Hypergravity and Microgravity on Biomedical Experiments
Thais Russomano, Gustavo Dalmarco, and Felipe Prehn Falcão
2008

A Biosystems Approach to Industrial Patient Monitoring and Diagnostic Devices
Gail Baura
2008

Multimodal Imaging in Neurology: Special Focus on MRI Applications and MEG
Hans-Peter Müller and Jan Kassubek
2007

Estimation of Cortical Connectivity in Humans: Advanced Signal Processing Techniques
Laura Astolfi and Fabio Babiloni
2007

Brain–Machine Interface Engineering
Justin C. Sanchez and José C. Principe
2007

Introduction to Statistics for Biomedical Engineers
Kristina M. Ropella
2007

Capstone Design Courses: Producing Industry-Ready Biomedical Engineers
Jay R. Goldberg
2007

BioNanotechnology
Elisabeth S. Papazoglou and Aravind Parthasarathy
2007

Bioinstrumentation
John D. Enderle
2006

Fundamentals of Respiratory Sounds and Analysis
Zahra Moussavi
2006

Advanced Probability Theory for Biomedical Engineers
John D. Enderle, David C. Farden, and Daniel J. Krause
2006

Intermediate Probability Theory for Biomedical Engineers
John D. Enderle, David C. Farden, and Daniel J. Krause
2006

Basic Probability Theory for Biomedical Engineers
John D. Enderle, David C. Farden, and Daniel J. Krause
2006

Sensory Organ Replacement and Repair
Gerald E. Miller
2006

Artificial Organs
Gerald E. Miller
2006

Signal Processing of Random Physiological Signals
Charles S. Lessard
2006

Image and Signal Processing for Networked E-Health Applications
Ilias G. Maglogiannis, Kostas Karpouzis, and Manolis Wallace
2006

Modeling and Analysis of Shape

with Applications in Computer-Aided Diagnosis of Breast Cancer

Denise Guliato
Federal University of Uberlândia, Brazil

Rangaraj M. Rangayyan
University of Calgary, Canada

SYNTHESIS LECTURES ON BIOMEDICAL ENGINEERING #39

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

Malignant tumors due to breast cancer and masses due to benign disease appear in mammograms with different shape characteristics: the former usually have rough, spiculated, or microlobulated contours, whereas the latter commonly have smooth, round, oval, or macrolobulated contours. Features that characterize shape roughness and complexity can assist in distinguishing between malignant tumors and benign masses.

In spite of the established importance of shape factors in the analysis of breast tumors and masses, difficulties exist in obtaining accurate and artifact-free boundaries of the related regions from mammograms. Whereas manually drawn contours could contain artifacts related to hand tremor and are subject to intra-observer and inter-observer variations, automatically detected contours could contain noise and inaccuracies due to limitations or errors in the procedures for the detection and segmentation of the related regions. Modeling procedures are desired to eliminate the artifacts in a given contour, while preserving the important and significant details present in the contour.

This book presents polygonal modeling methods that reduce the influence of noise and artifacts while preserving the diagnostically relevant features, in particular the spicules and lobulations in the given contours. In order to facilitate the derivation of features that capture the characteristics of shape roughness of contours of breast masses, methods to derive a signature based on the turning angle function obtained from the polygonal model are described. Methods are also described to derive an index of spiculation, an index characterizing the presence of convex regions, an index characterizing the presence of concave regions, an index of convexity, and a measure of fractal dimension from the turning angle function.

Results of testing the methods with a set of 111 contours of 65 benign masses and 46 malignant tumors are presented and discussed. It is shown that shape modeling and analysis can lead to classification accuracy in discriminating between benign masses and malignant tumors, in terms of the area under the receiver operating characteristic curve, of up to 0.94.

The methods have applications in modeling and analysis of the shape of various types of regions or objects in images, computer vision, computer graphics, and analysis of biomedical images, with particular significance in computer-aided diagnosis of breast cancer.

KEYWORDS

breast cancer, breast masses, breast tumors, compactness, concavity, convexity, Fourier descriptors, fractal dimension, polygonal modeling, shape analysis, shape factors, shape modeling, signature of a contour, spiculation index, turning angle function

To my family, with love.

Denise

Contents

Preface . . . . . **xiii**

Acknowledgments . . . . . **xv**

Symbols and Abbreviations . . . . . **xvii**

1 Analysis of Shape **1**

1.1 The Importance of Shape 1
1.2 Characteristics of Breast Tumors 1
1.3 Representation of Shape 3
1.4 Organization of the Book 4

2 Polygonal Modeling of Contours **9**

2.1 Review of Methods for Polygonal Modeling 9
2.2 Rule-based Polygonal Modeling of Contours 11
2.2.1 Comparative Analysis of Polygonal Models 12
2.3 Polygonal Approximation of Contours based on the Turning Angle Function . . 19
2.3.1 The TAF of a Contour 20
2.3.2 Polygonal Model from the TAF 20
2.3.3 Polygonal Model from the Filtered TAF 27
2.3.4 Illustrations of Application 33
2.4 Remarks 33

3 Shape Factors for Pattern Classification **41**

3.1 Signature Based on the Filtered TAF 41
3.2 Feature Extraction from the STAF 46
3.2.1 Derivation of an Index of Spiculation from the STAF 46
3.2.2 Fractal Dimension from the STAF 46
3.2.3 Index of Convexity 48
3.3 Shape Factors from Contours 49
3.3.1 Compactness 49
3.3.2 Spiculation Index 49

3.3.3 Fractional Concavity 49
3.3.4 Fourier Factor 49
3.3.5 Fractal Analysis 50
3.4 Remarks 51

4 Classification of Breast Masses 53
4.1 Datasets of Contours of Breast Masses 53
4.2 Results of Shape Analysis and Classification 53
4.3 Remarks 63

References 65

Authors' Biographies 73

Index 75

Preface

Shape is an important property of objects that helps in the recognition of several things in day-to-day life. In scientific, engineering, and medical applications, shape not only assists in the identification of natural life forms and artificially created objects, but also helps in distinguishing between normal and abnormal variations or cases of a given entity. In particular, in computer-aided diagnosis of medical conditions, we show in this book that modeling and analysis of shape can facilitate discrimination between benign breast masses and malignant tumors due to cancer as seen on mammograms. Such applications are facilitated by techniques of modeling, characterization, feature extraction, and pattern classification directed to the analysis of shape.

The methods described in this book are mathematical in nature. It is assumed that the reader is proficient in advanced mathematics and familiar with basic notions of data, signal, and image processing. The methods of modeling and analysis are suitable for inclusion in courses for students in the final years of bachelor's programs in electrical engineering, computer engineering, mathematics, physics, computer science, biomedical engineering, and bioinformatics. The techniques should also be useful to researchers in various areas of modeling and analysis, and they could be included in graduate courses on digital image processing, medical imaging, and related topics. The book is copiously illustrated with figures and examples of application to facilitate efficient comprehension of the notions and methods presented.

We wish our readers success in their studies and research.

Denise Guliato and Rangaraj M. Rangayyan
January 2011

Acknowledgments

The research projects that led to the methods and results presented in this book work were supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil; Universidade Federal de Uberlândia, Brazil; the Fundação de Amparo à Pesquisa do Estado de Minas Gerias (FAPEMIG), Brazil; the Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil; Faculdade de Computação, Universidade Federal de Uberlândia, Brazil; the Natural Sciences and Engineering Research Council (NSERC) of Canada; the Canadian Breast Cancer Foundation: Prairies/NWT Chapter; the Alberta Heritage Foundation for Medical Research (AHFMR), Canada; and the Catalyst Program of Research Services, University of Calgary, Canada.

We thank Juliano Daloia de Carvalho for his contributions to related works, hard work, dedication to research, initiatives, enthusiastic support, and camaraderie, as well as for the examples provided from his Master's thesis. We thank Sérgio A. Santiago for his participation in related projects and publications. We thank Fábio José Ayres, University of Calgary, for assistance in related research projects.

Denise Guliato and Rangaraj M. Rangayyan
January 2011

Symbols and Abbreviations

Note: Variables or symbols used within limited contexts are not listed here; they are described within their context. The mathematical symbols listed may stand for other entities or variables in different applications; only the common associations used in this book are listed for ready reference.

abs	absolute value
arctan	inverse tangent, $\tan^{-1}$
arg	argument of
au	arbitrary units
ANN	artificial neural network
AUC	area under the ROC curve
A_z	area under the ROC curve
b	bit
BI-RADS™	Breast Imaging Reporting and Data System
cf	compactness
CAD	computer-aided diagnosis
CBIR	content-based image retrieval
CC	cranio-caudal
C_p	compression rate
CX_{TA}	index of convexity based on the turning angle
$d(n)$	signature based on the Euclidean distance
exp (x)	exponential function, e^x
f_{cc}	fractional concavity
ff	shape factor obtained using Fourier descriptors
FD	fractal dimension
FN	false negative
FNF	false-negative fraction
FP	false positive
FPF	false-positive fraction
$h(A, B)$	Hausdorff distance between A and B
i	index of a series
j	$\sqrt{-1}$
max	maximum
min	minimum

mm	millimeter
mod	modulus or modulo
M	number of samples or pixels
MIAS	Mammographic Image Analysis Society, London, UK
MLO	medio-lateral oblique
MSE	mean-squared error
n	an index
N	number of samples or pixels
NC	number of pixels or points in a contour
NP	number of pixels or points in a polygonal model
ROC	receiver operating characteristics
ROI	region of interest
SI	spiculation index
STAF	signature based on the turning angle function
TA	turning angle
TAF	turning angle function
TN	true negative
TNF	true-negative fraction
TP	true positive
TPF	true-positive fraction
$T_C(s_i)$	turning angle function, value for the segment s_i of the contour C
VR_{TA}	measure of concavity based on the turning angle
XR_{TA}	measure of convexity based on the turning angle
$x(n)$, $y(n)$	x and y coordinates of the nth point on a contour
$(\overline{x}, \overline{y})$	x and y coordinates of the centroid of a contour
$\emptyset$	null set
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
θ	an angle
μ	the mean (average) of a random variable
σ	the standard deviation of a random variable
σ^2	the variance of a random variable
$\overline{x}$	average or normalized version of the variable under the bar
$\overline{x}$	complement of the variable under the bar
$'$, $''$, $'''$	first, second, and third derivatives of the preceding function
$\forall$	for all
$\in$	belongs to or is in (the set)

$\{\ \}$	a set
$\subset$	subset
$\supset$	superset
$\bigcap$	intersection
$\bigcup$	union
$\equiv$	equivalent to
$\mid$	given, conditional upon
$\rightarrow$	maps to
$\Leftarrow$	gets (updated as)
$\Rightarrow$	leads to
$\Leftrightarrow$	transform pair
$[\ ]$	closed interval, including the limits
$(\ )$	open interval, not including the limits
$\lvert\ \rvert$	absolute value or magnitude
$\lvert\ \rvert$	determinant of a matrix
$\lVert\ \rVert$	norm of a vector or matrix
$\lceil x \rceil$	ceiling operator; the smallest integer $\geq x$
$\lfloor x \rfloor$	floor operator; the largest integer $\leq x$

CHAPTER 1

Analysis of Shape

1.1 THE IMPORTANCE OF SHAPE

Shape is an important feature of natural as well as artificial objects that facilitates their recognition and analysis. We identify people, plants, animals, writings, and several objects in our daily lives using specific characteristics of their shapes. For example, we identify different types of flowers and leaves, varieties of tools and implements, alphabets of languages, the letters of an alphabet, and categories of vehicles by their shapes. Indeed, several properties other than shape also play important roles in the recognition of objects, such as color, texture, and three-dimensional (3D) form. In addition, our tactile and olfactory senses are also used in appreciating the nature of several things, objects, and living entities. Regardless, the general caricature or shape of an object is a primary visual feature that plays an important role in its analysis and recognition by a human being or via computer processing [1, 2, 3, 4, 5, 6].

Several human organs have readily identifiable shapes: recognizing and sketching the forms of the human body, face, eyes, nose, mouth, ears, hands, and fingers are activities learned in early childhood. In medical diagnosis, shape plays a vital role in the recognition of anatomical structures as well as the identification of abnormalities caused by disease. In radiology, the parts of the body of interest are identified using several characteristics as visible on X-ray or other types of medical images, with shape playing a major role in the analysis [5]. The shapes of the heart, kidneys, ribs, and several bones are well known and easily recognized. In spite of extensive variations within the normal range for each of the organs of the human body, specialized physicians such as cardiologists and radiologists are capable of identifying small changes due to pathology.

1.2 CHARACTERISTICS OF BREAST TUMORS

Mammography is the best method available for early detection of breast cancer [7]. Large populations of asymptomatic women are participating in mammographic screening programs [8]. With the aim of improving the accuracy and efficiency of screening programs for the detection of early signs of breast cancer, a number of research projects are focusing on the development of methods for computer-aided diagnosis (CAD) to assist radiologists in diagnosing breast cancer [5, 9, 10, 11]. A key requirement in reducing the mortality rate due to breast cancer is to identify and remove malignant tumors at an early stage before they metastasize and spread to neighboring regions.

Evidence of a breast tumor is usually indicated by the presence of a dense mass and/or a change in the texture or distortion in the mammogram. Consequently, the focus during diagnosis is on identifying such abnormal regions, as well as on classifying the type of mass or tumor that caused

the abnormality. A typical benign mass is round and smooth with a well-defined (well-circumscribed) boundary, whereas a typical malignant tumor is spiculated and rough with a blurry (ill-defined or ill-circumscribed) boundary [5, 7, 12]. There could also be some unusual cases of macrolobulated or slightly spiculated benign masses, as well as nearly round, microlobulated, or well-circumscribed malignant tumors; such atypical cases cause difficulties in pattern classification studies [13, 14].

Figure 1.1 shows four regions of mammograms containing masses of different types in gray scale (upper row), as well as their contours drawn by a radiologist (lower row) [15]. The well-circumscribed benign mass has a nearly circular and smooth contour, whereas the macrolobulated benign mass exhibits a few large partitions or lobes in its contour. The microlobulated malignant tumor has small lobules that add some roughness to its contour. The highly spiculated and ill-defined malignant tumor possesses a rough and jagged contour. The examples illustrate different levels of complexity and roughness of contours of breast masses as seen in mammograms, with increasing roughness being associated with increasing levels of suspicion of malignant disease, that is, cancer.

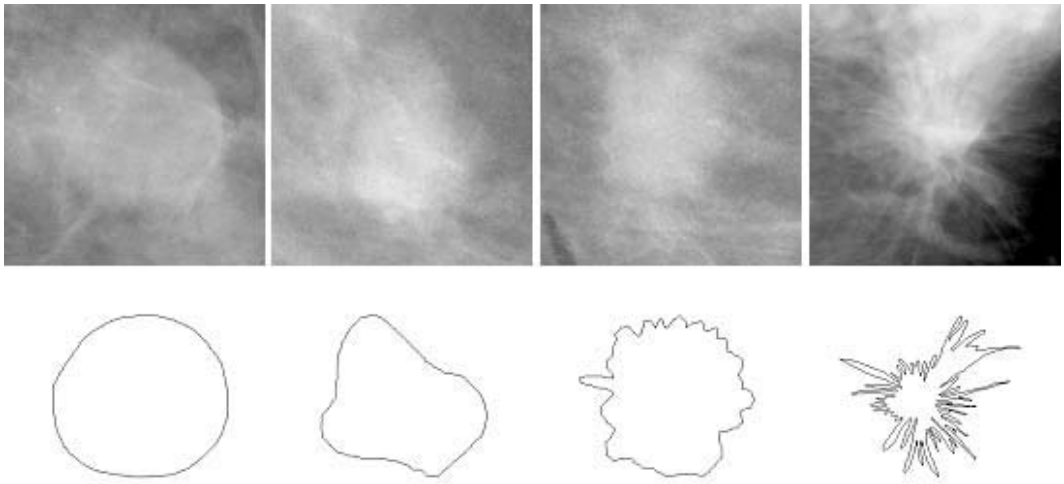


Figure 1.1: Regions of mammograms containing masses of four types and their contours drawn by a radiologist. Left to right: a well-circumscribed benign mass, a macrolobulated benign mass, a microlobulated malignant tumor, and a spiculated and ill-defined malignant tumor. Reproduced with permission from H. Alto, R.M. Rangayyan, and J.E.L. Desautels, "Content-based retrieval and analysis of mammographic masses", Journal of Electronic Imaging, Vol. 14, No. 2, Article 023016, pp 1-17, 2005. ©SPIE and IS&T.

.

1.3 REPRESENTATION OF SHAPE

The most general representation of the shape of an object is in terms of the 3D coordinates of the points on its surface, expressed as $\{x(n), y(n), z(n)\}$, $n = 0, 1, 2, \ldots, N-1$, where N is the number of points on the surface. No information is included regarding the internal properties of the object, such as density or material composition, or on its external characteristics, such as color or texture. In digital image processing, it is common to deal with two-dimensional (2D) representations of objects and regions of interest (ROIs) in images; such an entity could be represented in terms of the 2D coordinates of the points on its boundary or contour, expressed as $\{x(n), y(n)\}$, $n = 0, 1, 2, \ldots, N-1$, where N is the number of points on the contour. Once again, no information is included on the intensity or color of the image on its boundary or within the region contained. The contour or shape of the object may be plotted as a binary drawing or image in a 2D plane.

A 2D contour may be transformed into a one-dimensional (1D) function or signature by computing a certain property for each point on the contour. One of the commonly used signatures is defined as the Euclidean distance from each contour point to the centroid, $(\overline{x}, \overline{y})$, of the contour as a function of the index of the contour point:

$$d(n) = \sqrt{[x(n) - \overline{x}]^2 + [y(n) - \overline{y}]^2}, \tag{1.1}$$

$n = 0, 1, 2, \ldots, N-1$, where

$$\overline{x} = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \tag{1.2}$$

and

$$\overline{y} = \frac{1}{N} \sum_{n=0}^{N-1} y(n). \tag{1.3}$$

A contour may also be expressed using a complex representation of its (x, y) coordinates as $z(n) = x(n) + j\,y(n)$, where $j = \sqrt{-1}$, which facilitates analysis using Fourier descriptors [5]. Another type of signature may be defined as

$$d(n) = |z(n)| = \sqrt{x^2(n) + y^2(n)}, \tag{1.4}$$

$n = 0, 1, 2, \ldots, N-1$.

Pohlman et al. [16] derived the signature of a contour as a function of the radial distance from the centroid to the contour versus the angle of the radial line over the range $[0°, 360°]$; however, this definition could lead to a multivalued function in the case of an irregular or spiculated contour. A signature computed in this manner would also have ranges of undefined values in the case of a contour for which the centroid falls outside the region enclosed by the contour.

A benign breast mass in a mammogram is generally round in shape, being well-circumscribed or macrolobulated, and would have a smooth signature, as shown in Figure 1.2. On the other hand, a malignant tumor is usually rough in shape, being spiculated or microlobulated, and therefore, would have a rough and complex signature, as shown in Figure 1.3. Quantitative measures or features of shape may be derived from either a 2D contour or its 1D signature, depending upon the desired characteristics.

Measures that can quantitatively represent shape roughness and complexity can assist in the classification of malignant tumors and benign masses [13, 14, 17]. On the basis of the shape differences between benign masses and malignant tumors, objective features of shape complexity such as compactness (cf), fractional concavity (f_{cc}), spiculation index (SI), a Fourier-descriptor-based factor (ff), moments, chord-length statistics, fractal dimension (FD), and wavelet transform modulus-maxima have been developed for pattern classification [13, 14, 17, 18, 19, 20, 21, 22].

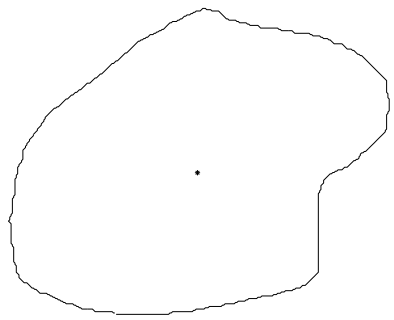
In spite of the established importance of shape factors in the analysis of breast tumors and masses, difficulties exist in obtaining accurate and artifact-free boundaries of the related regions from mammograms. Whereas manually drawn contours could contain artifacts related to hand tremor and are subject to intra-observer and inter-observer variations, automatically detected contours could contain noise and inaccuracies due to limitations or errors in the procedures for the detection and segmentation of the related regions. Modeling procedures are desired to eliminate the artifacts in a given contour, while preserving the important and significant details present in the contour.

1.4 ORGANIZATION OF THE BOOK

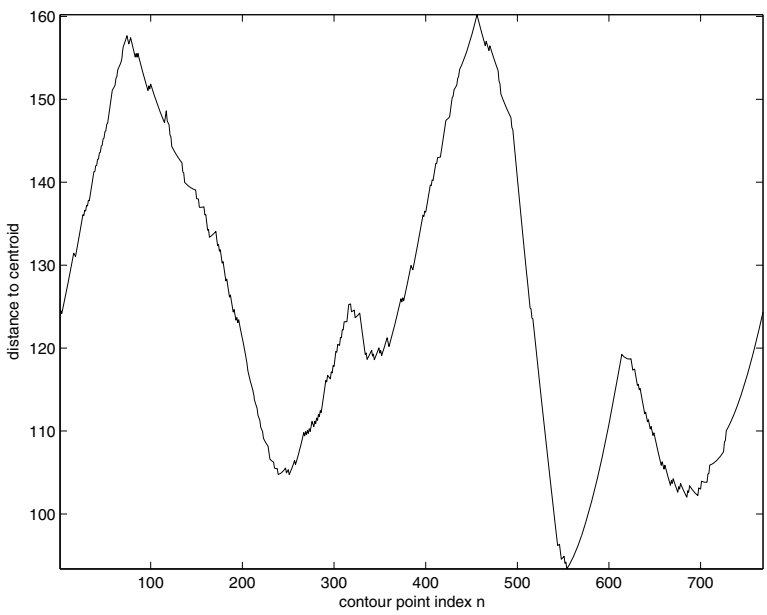
In this book, we present methods for polygonal modeling that reduce the influence of noise and artifacts while preserving the diagnostically relevant features, in particular, the spicules and lobulations in the original contour of a breast mass or tumor [23]. One of the polygonal modeling methods presented is based on straight-line segments, whose end points (or vertices of the polygon) are obtained by an iterative process controlled by conditions related to the lengths of the sides of the polygon as well as its angles. Another method is based on the turning angle function (TAF) [22, 24, 25] of the given contour.

To evaluate the performance of the modeling procedures in terms of the efficiency in the classification of breast masses, we demonstrate the derivation of shape factors that represent the presence of spicules, convex or concave regions, and FD from the models. We compare the results with those provided by SI, f_{cc}, and FD using the methods proposed by Rangayyan et al. [14] and Rangayyan and Nguyen [17], in terms of the area (AUC or A_z) under the receiver operating characteristic (ROC) curve.

The book is organized as follows: Chapter 2 introduces the general concept of polygonal modeling procedures and presents two novel polygonal modeling methods that preserve the relevant features in a given contour. Chapter 3 provides the details of techniques to derive an index of spiculation, FD, and an index of convexity based on the TAF obtained from a polygonal model.

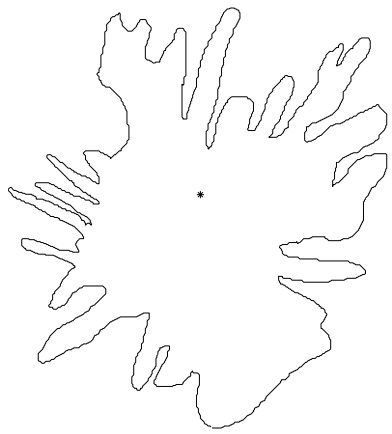


(a)

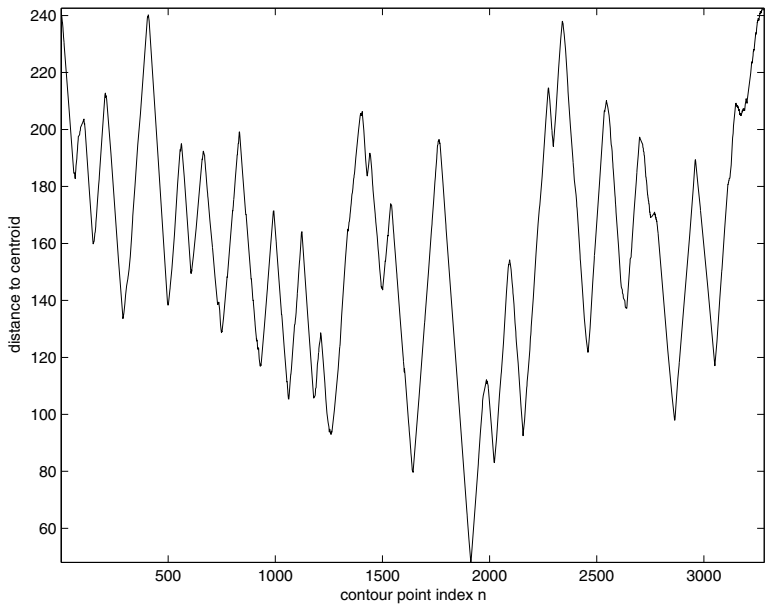


(b)

Figure 1.2: (a) Contour of a benign breast mass; $N = 768$. The '*' mark represents the centroid of the contour. (b) Signature computed as the Euclidean distance from each contour point to the centroid of the contour; $d(n)$ as defined in Equation 1.1. Reproduced with permission from R.M. Rangayyan, "Biomedical Image Analysis", CRC Press, Boca Raton, FL. ©CRC Press. 2005.



(a)



(b)

Figure 1.3: (a) Contour of a malignant breast tumor; $N = 3,281$. The '*' mark represents the centroid of the contour. (b) Signature computed as the Euclidean distance from each contour point to the centroid of the contour; $d(n)$ as defined in Equation 1.1. Reproduced with permission from R.M. Rangayyan, "Biomedical Image Analysis", CRC Press, Boca Raton, FL. ©CRC Press. 2005.

Finally, Chapter 4 gives a description of the dataset used in pattern classification experiments and presents a comparative analysis of the results obtained by the various methods described in the book.

CHAPTER 2

Polygonal Modeling of Contours

2.1 REVIEW OF METHODS FOR POLYGONAL MODELING

The problem of polygonal approximation or polygonal modeling of a contour may be stated as finding the vertices of a polygon along the contour in such a way that the result is a good approximation of the original contour [5, 6]. The available methods for vertex detection and polygonal approximation of a given contour can be divided into two main classes: global methods and local methods.

Typical global modeling methods use, as measures of approximation or stopping criteria, minimization of the mean-squared error (MSE) between the given contour and the model, the minimal polygon perimeter, the maximal internal polygon area, or the minimal area external to the polygon but contained by the given contour [23, 26, 27, 28, 29, 30, 31]. On the other hand, local methods for shape modeling and analysis are based on the idea of coding the object's contour as an ordered sequence of points or high-curvature points, obtained by different techniques [13, 14, 32, 33, 34, 35, 36, 37], or as chain-code histograms [34, 35, 38, 39]. An extensive bibliographic listing on polygonal representation from curves is available online [40].

Ramer [28] proposed a split-based algorithm with the aim of approximating a given contour by a polygonal model, using an iterative procedure. The stopping criterion is based on a predefined error parameter that gives a measure of the maximal error of approximation allowed. The algorithm starts with an initial solution, and proceeds, iteratively, until the error measure is verified for every contour segment approximated by a straight-line segment. Although this is a simple method and provides good results, depending on the contour and on the initial points, the algorithm retains the initial points in the final solution even if they do not represent vertices on the contour. Pavlidis and Horowitz [27] extended the method proposed by Ramer [28] in a split-merge approach: the idea behind this method is to eliminate those points present in the initial solution that do not represent vertices in the polygonal model.

Latecki and Lakämper [30] proposed a discrete curve evolution procedure that is context sensitive, to reduce the influence of noise and to simplify the shape with the aim of image retrieval. At every step of the evolution, a pair of consecutive segments is replaced by the segment resulting from their union. The key property of this evolution is the order of the substitution given by a function of the angle between two adjacent segments and their sizes. The algorithm stops after a number of iterations previously determined by an automatic procedure that takes into account the judgment of the user.

Menut et al. [37] proposed a method to fit each piecewise-continuous part of a given contour with a parabolic model. The parameters of the parabolic segments were used for the classification of

breast masses in mammograms as benign or malignant. Contours of benign masses were typically segmented into a few wide parabolas, and several small and flat sections, due to smooth boundaries and large lobulations. On the other hand, contours of malignant tumors were typically modeled with a large number of narrow parabolas and few flat sections. The method was not extended to provide a reconstructed model of the original contour.

Ventura and Chen [36] presented an algorithm to segment 2D curves in which the number of segments is prespecified to initiate the process, in relation to the complexity of the shape. This may not be a desirable step, depending on the application. Rangayyan et al. [14] proposed a polygonal modeling procedure that eliminates this limitation of the method of Ventura and Chen [36]. The procedure proposed by Rangayyan et al. [14] begins by segmenting the given closed contour into a set of piecewise-continuous curved parts; this is achieved by locating the points of inflection on the contour, based on its first, second, and third derivatives. The algorithm retains the initial points of inflection in the final polygonal model, thereby constraining the fit of the model to the contour provided. In addition, the criteria used do not specifically relate to the notion of preserving the important details of interest, which could vary from one application to another.

Guliato et al. [31] proposed a polygonal modeling method that preserves relevant information for pattern classification. The method is based on merging adjacent segments of the polygonal model being developed, by taking into account the lengths of adjacent segments and the value of the smaller angle between them. Rangayyan et al. [41] proposed a modification to the method proposed by Guliato et al. [31]: in the modified method, the polygonal model is obtained from the TAF of the contour, considering the same rules as in the earlier method to merge adjacent segments. In the modified method, all of the parameters required to derive the polygonal model are explicitly represented through the TAF.

Costa and Sandler [42] proposed a similar approach to merge adjacent segments of a polygonal model based on the angle between them. The work of Costa and Sandler [42] is concerned with the detection of digital bar segments using the Hough transform. To merge adjacent segments, Costa and Sandler used the absolute difference between the angles of their normal and radius parameters, with threshold values. It is worth noting that this approach requires the computation of a parameterized equation for each segment in order to derive the parameters required for the analysis.

Brief descriptions of the methods of Rangayyan et al. [14] and Pavlidis and Horowitz [27] are given below. The method proposed by Guliato et al. [31] is described in Section 2.2 and a modification to the same is described in Section 2.3.

- The polygonal modeling method proposed by Rangayyan et al. [14]:

  Rangayyan et al. [14] proposed a method to derive the polygonal model of a given contour by using the points of inflection as the initial input to an iterative polygonal modeling procedure. The vertices of the initial polygonal model are placed at the points of inflexion. Then, the maximal arc-to-chord distance from each side of the polygonal model to the corresponding segmented curved part of the original contour is computed. If the distance is greater than a predefined threshold, an additional vertex of the polygonal model is placed on the original

contour at the point of maximal distance, thereby increasing the order of the model by one. The procedure is iterated subject to predefined stopping criteria to minimize the error between the perimeter of the original contour and the perimeter of the polygonal model. The maximal arc-to-chord distance permitted in the work of Rangayyan et al. [14] was 0.25 mm or 5 pixels (at the pixel size of 50 μm), and the smallest side of the polygon permitted was 1.0 mm. The method does not require any interaction with the user.

- The polygonal modeling method proposed by Pavlidis and Horowitz [27]:

  This algorithm allows a variable number of segments. After an arbitrary initial choice, segments are split and merged in order to derive the polygonal model that provides the best polygonal approximation to the given contour, under a prespecified error bound, $E_{\max}$, given as input. In the original work, the segment between two points is obtained by minimizing an error measure. However, the resulting segment is not necessarily continuous, although the discontinuity could be resolved, if necessary, with further processing. For the purpose of comparison, the method proposed by Pavlidis and Horowitz [27] was implemented as follows.

  The initial solution is composed of two points: the left-most and the right-most points on the original contour. The approximation error is obtained by computing

$$E = \max(d_i), \tag{2.1}$$

  where d_i is the distance between the point p_i of the given arc segment C, limited by the end-points A and B in the original contour, and the straight segment AB. If E is greater than the given threshold $E_{\max}$, then the curve C is split at the point p_i where E is maximal. The procedure is iterated until the specified stopping conditions are met. Although the method provides good results, the computational cost is high.

2.2 RULE-BASED POLYGONAL MODELING OF CONTOURS

The polygonal modeling procedure proposed by Guliato et al. [31] can be configured according to the needs of the application. The method starts by identifying all of the linear segments of the given contour (some of the segments could be as short as two pixels). Let M, M_i, and N be the number of the points in the given contour, the number of points in the ith linear segment, and the number of the linear segments in the contour, respectively. Then, the original contour is given by $S = \{(x_j, y_j)\}$, $j = 1, 2, \ldots, M$. The contour is partitioned into N linear segments, $S_i = \{(x_{ij}, y_{ij})\}$, $j = 1, 2, \ldots, M_i$, $i = 1, 2, \ldots, N$, with $M = M_1 + M_2 + \ldots + M_N$, and $S_k \cap S_l = \emptyset$ $\forall (k, l), k \neq l$.

The next step is to reduce the influence of noise while maintaining the semantically (or diagnostically) relevant characteristics of the given contour, and attempting to reduce, in each iteration of the algorithm, the number of linear segments in the original contour, as well as to increase the

number of points in each new linear segment. The algorithm to obtain the polygonal model executes the following two rules for every linear segment in each iteration.

Rule 1: if two adjacent segments S_i and S_{i+1} are shorter than a threshold $S_{\min}$, then join S_i and S_{i+1}.

Rule 2: if the length of S_i or S_{i+1} is greater than the threshold $S_{\min}$, then analyze the smaller angle between S_i and S_{i+1}; if the angle is greater than the given threshold $\theta_{\max}$, then join S_i and S_{i+1}, else retain S_i and S_{i+1}.

The threshold $S_{\min}$ depends upon the relevance of a short segment and $\theta_{\max}$ depends upon the relevance of the angle between two adjacent linear segments being analyzed in the application of interest. The angle used is always the smaller of the two angles between the adjacent sides being analyzed, which could be the related internal angle or the external angle of the polygon. The algorithm stops when no two linear segments are joined in an iteration.

Figure 2.1 illustrates the results obtained by the method for a simple test figure with different sets of parameters. It is worth noting that, in both of the cases illustrated, the important and relevant shape-related information has been preserved.

2.2.1 COMPARATIVE ANALYSIS OF POLYGONAL MODELS

The results obtained for a few test patterns by applying the polygonal modeling method described in the preceding paragraphs and the methods of Pavlidis and Horowitz [27] and Rangayyan et al. [14] are compared in the present section. The contours used for the comparative analysis were artificially generated, and are shown in Figure 2.2. To each original contour, noise was added with the length of the segments varying from 5 to 15 pixels and the angles between the segments varying from 155° to 170°; see Figure 2.3.

The results obtained from the three polygonal modeling approaches are shown in Figures 2.4, 2.5, and 2.6.

It should be observed that, in all of the cases illustrated, the polygonal approximations provided by the method of Guliato et al. [31] have removed the noise, while preserving the nature of the original contour (convex or nonconvex shapes). In particular, for the ellipse, the use of a tuned set of parameters should result in a more precise approximation.

To compare objectively the results obtained by the three polygonal modeling approaches, the compression rate, Cp, given by

$$Cp = \frac{NP}{NC}, \tag{2.2}$$

where NP is the number of vertices in the polygonal approximation and NC is the number of points in the original contour, and the Hausdorff distance [43], $h(A, B)$, defined as

$$h(A, B) = \max_{a \in A} \left[\min_{b \in B} \{ \|a - b\| \} \right], \tag{2.3}$$

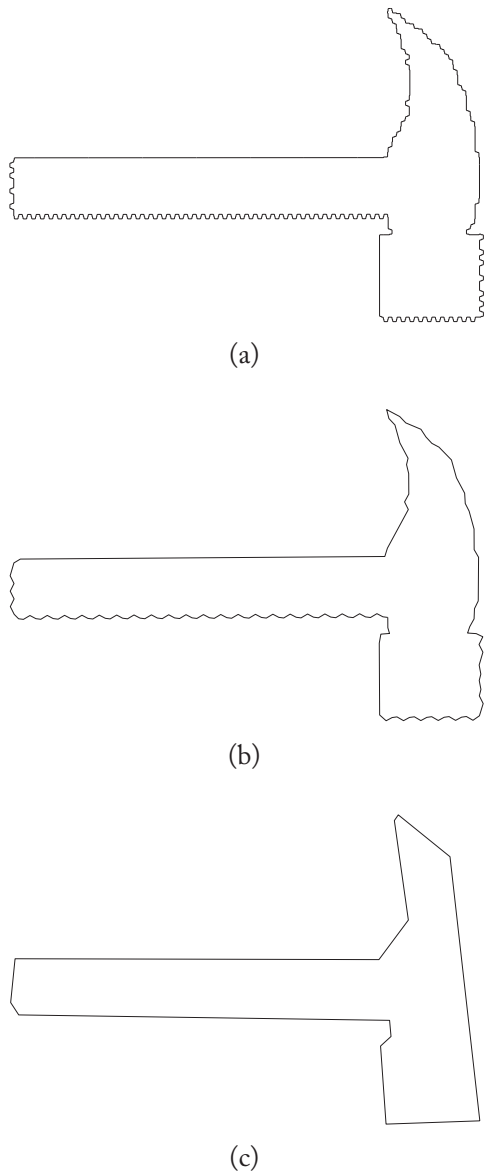


Figure 2.1: (a) Original contour of a hammer; $NC = 3253$. (b) Polygonal approximation with $\theta_{\max} = 160°$ and $S_{\min} = 10$ pixels; $NP = 90$. (c) Polygonal approximation with $\theta_{\max} = 150°$ and $S_{\min} = 20$ pixels; $NP = 13$.

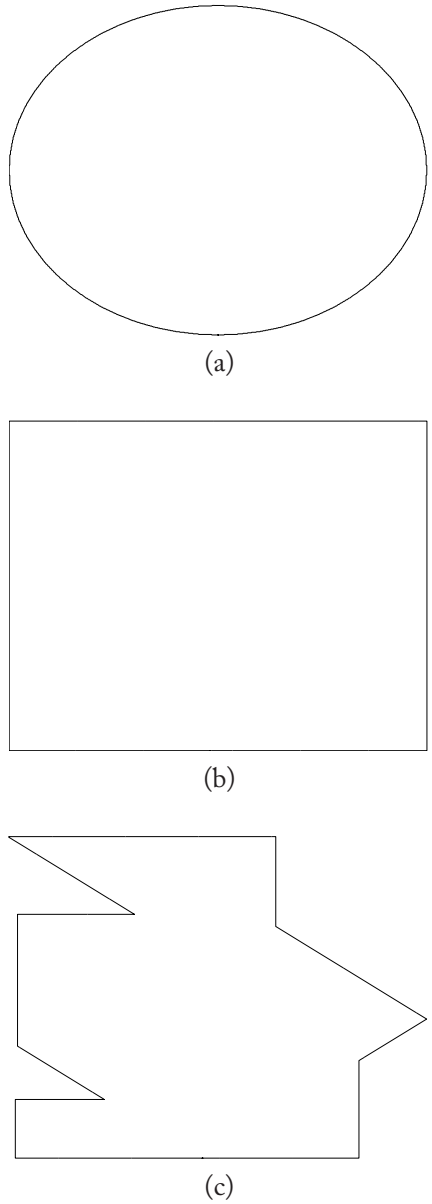


Figure 2.2: (a) Original contour of an ellipse. (b) Original contour of a rectangle. (c) Original contour of a nonconvex shape.

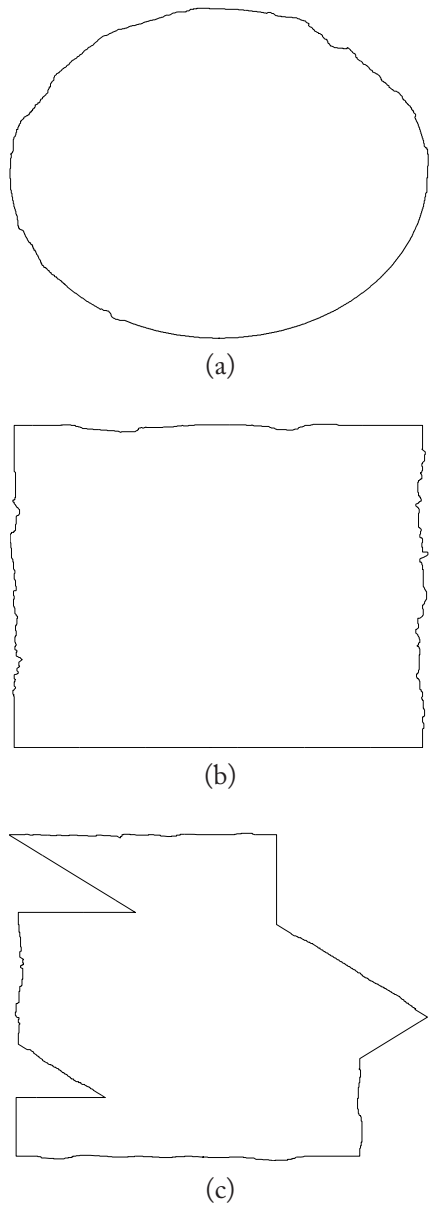


Figure 2.3: Noisy contours obtained from the original contours in Figure 2.2: (a) noisy contour of the ellipse; (b) noisy contour of the rectangle; and (c) noisy contour of the nonconvex shape.

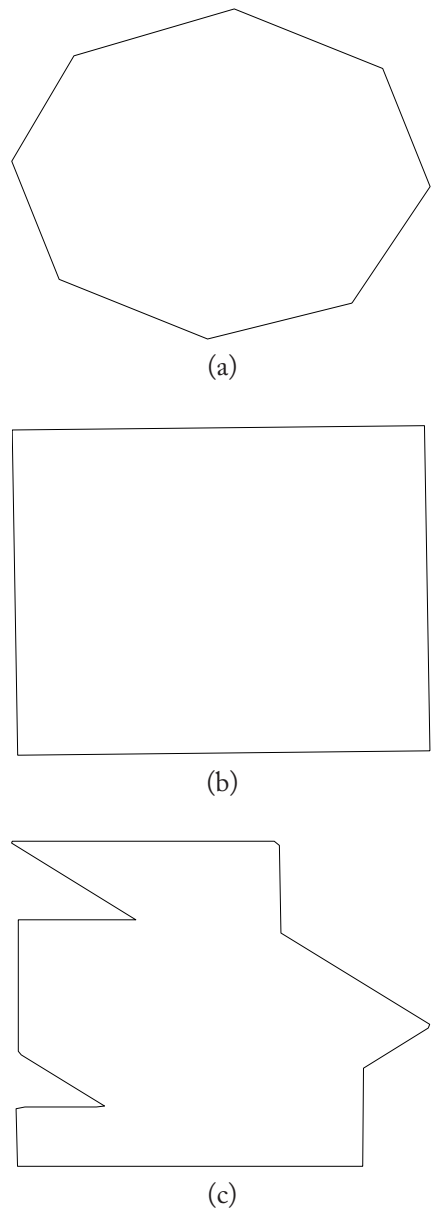


Figure 2.4: Polygonal models obtained for the contours in Figure 2.3 using the method proposed by Guliato et al. [31] with $S_{\min} = 15$ and $\theta_{\max} = 150°$: (a) polygonal model of the ellipse; (b) polygonal model of the rectangle; and (c) polygonal model of the nonconvex shape.

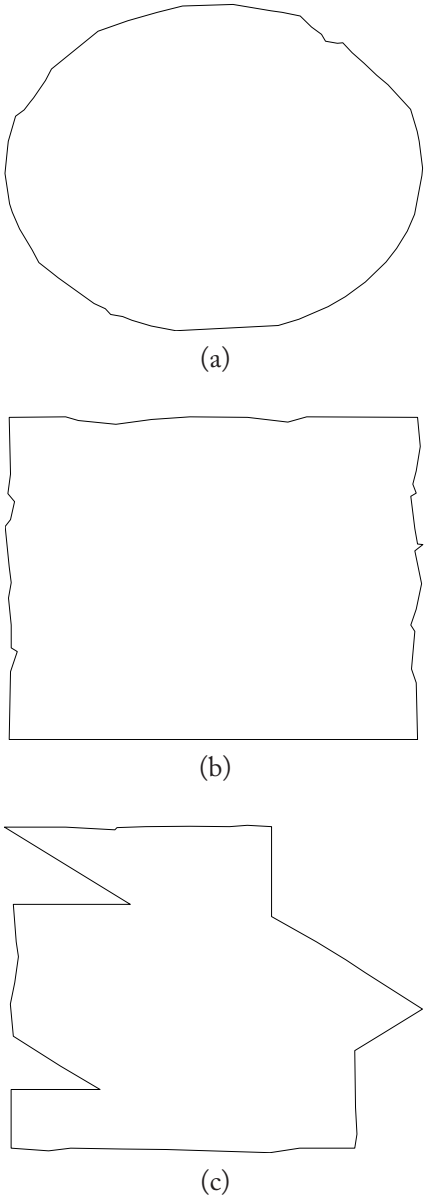


Figure 2.5: Polygonal models obtained for the contours in Figure 2.3 using the method proposed by Rangayyan et al. [14]: (a) polygonal model of the ellipse; (b) polygonal model of the rectangle; and (c) polygonal model of the nonconvex shape.

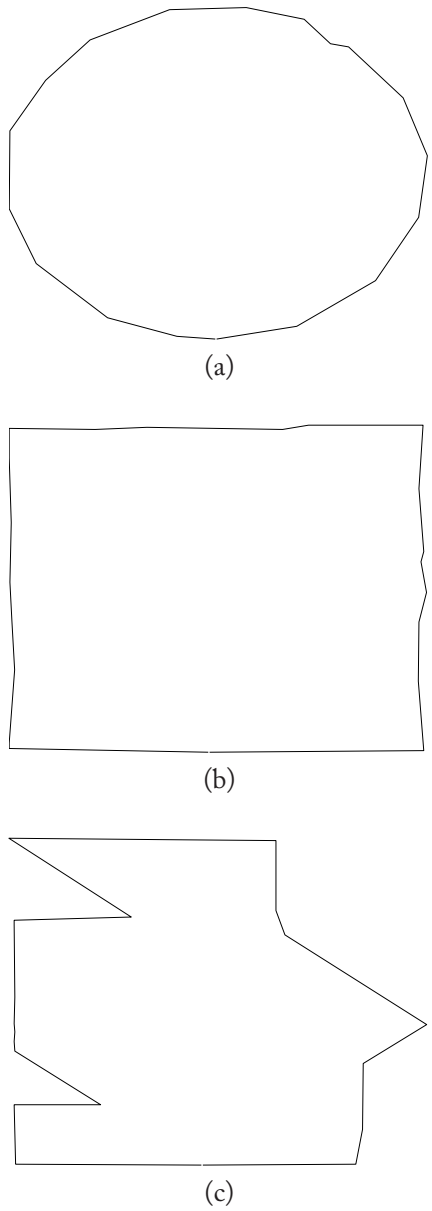


Figure 2.6: Polygonal models for the contours in Figure 2.3 using the method proposed by Pavlidis and Horowitz [27] with $E_{\max} = 50$: (a) polygonal model of the ellipse; (b) polygonal model of the rectangle; and (c) polygonal model of the nonconvex shape.

Table 2.1: Results obtained with the polygonal modeling method proposed by Pavlidis and Horowitz [27] with $E_{\max} = 50$ and $E_{\max} = 70$.

		$E_{\max} = 50$		$E_{\max} = 70$	
Contour	NC	$h(A, B)$ (pixels)	Cp	$h(A, B)$	Cp
Figure 2.3(a)	1900	27.29	0.010	31.16	0.007
Figure 2.3(b)	2533	23.40	0.013	24.24	0.015
Figure 2.3(c)	2705	25.42	0.018	26.38	0.011

Table 2.2: Results obtained with the polygonal modeling method proposed by Rangayyan et al. [14].

Contour	NC	$h(A, B)$ **(pixels)**	Cp
Figure 2.3(a)	1900	15.03	0.028
Figure 2.3(b)	2533	22.00	0.028
Figure 2.3(c)	2705	26.11	0.015

Table 2.3: Results obtained with the polygonal modeling method proposed by Guliato et al. [31] with $S_{\min} = 15$ and $\theta_{\max} = 150°$.

Contour	NC	$h(A, B)$ **(pixels)**	Cp
Figure 2.3(a)	1777	39.80	0.004
Figure 2.3(b)	2233	18.00	0.001
Figure 2.3(c)	2159	25.22	0.007

where A and B are the sets of points of the contours to be analyzed, were computed. The results are shown in Tables 2.1, 2.2, and 2.3. It is evident that the results of the method of Guliato et al. [31] provide the lowest compression ratio and the Hausdorff distance, except for the simple elliptic contour.

2.3 POLYGONAL APPROXIMATION OF CONTOURS BASED ON THE TURNING ANGLE FUNCTION

In the present section, the polygonal modeling method proposed by Rangayyan et al. [41] is described. The method is based on global polygonal approximation using the TAF [22, 24] of the given contour, with the aim of reduction of noise and artifacts, while preserving the relevant features. The method is controlled by the size of adjacent segments and by their turning angle [25]. The method described in the present section is different from the method described in Section 2.2 in the sense that, in the latter, the polygonal model is derived directly from the contour.

2.3.1 THE TAF OF A CONTOUR

The TAF, $T_C(s_n)$, of a contour, C, is the cumulative function of turning angles, and it may be obtained by deriving the counterclockwise angle between the tangent at the segment s_n and the x-axis, and expressing it as a function of the arc length of s_n [24]. The TAF is also known as the tangent function, and it has been used as a signature to represent the shape of a given contour (or its polygonal model) and in applications related to shape analysis and retrieval [22, 23, 24, 25, 30, 31, 44, 45, 46, 47, 48, 49]. The TAF keeps track of the turning angle of the contour, increasing with convex regions and decreasing with concave regions. The turning angle of a segment s_i is the difference or step between $T_C(s_i)$ and $T_C(s_{i+1})$. The turning angle ranges in the interval $(-180°, 180°)$. Negative values represent concave regions and positive values represent convex regions. For a convex contour, $T_C(s_n)$ is a monotonic function, starting at an arbitrary value ϕ and increasing to $\phi + 2\pi$. For a nonconvex polygon, $T_C(s_n)$ can become arbitrarily large, because it accumulates the total amount of turning angles, obeying the range of 2π between the starting point and the final point [24]. An example of a simple nonconvex shape and its TAF are shown in Figure 2.7.

Figure 2.8 shows a convex contour before and after the addition of noise as well as their TAFs. The monotonically increasing nature of the TAF is evident in this illustration, which is affected by the noise added. Figure 2.9 shows the contour of a mostly convex benign breast mass and its TAF. Random fluctuations are seen in the generally increasing TAF corresponding to small artifactual variations in the mostly convex contour.

Figure 2.10 shows a nonconvex contour and its TAF. For a contour with concave and convex regions, the TAF begins to decrease at the beginning of a concave portion and keeps on decreasing until the direction of the tangent to the contour changes at the beginning of the next convex portion.

The contour of a malignant breast tumor with several spicules and concave incursions is shown in Figure 2.11 along with its TAF. The TAF has several increasing and decreasing segments that correspond to the rough and jagged nature of the contour.

2.3.2 POLYGONAL MODEL FROM THE TAF

Contours drawn manually or derived automatically from a computational procedure could contain artifacts or noise related to hand tremor and other limitations. As a consequence, the corresponding TAFs could contain several small segments that are insignificant in the representation of the contours for further analysis. For this reason, it is necessary to filter TAFs in a selective manner, so as to remove the artifacts and noise, while preserving the significant details. Rangayyan et al. [41] proposed an iterative polygonal approximation method controlled by the size of the adjacent segments and their turning angle as represented in the TAF of the contour. The following two rules are applied to every linear segment s_i identified from the TAF in each iteration.

Rule 1: if the current segment s_i and the next segment s_{i+1} are both shorter than a threshold $S_{\min}$, then join s_i and s_{i+1}. The length of the combined segment is equal to the length of the straight line connecting the starting point of s_i and the ending point of s_{i+1}. The turning angle of the

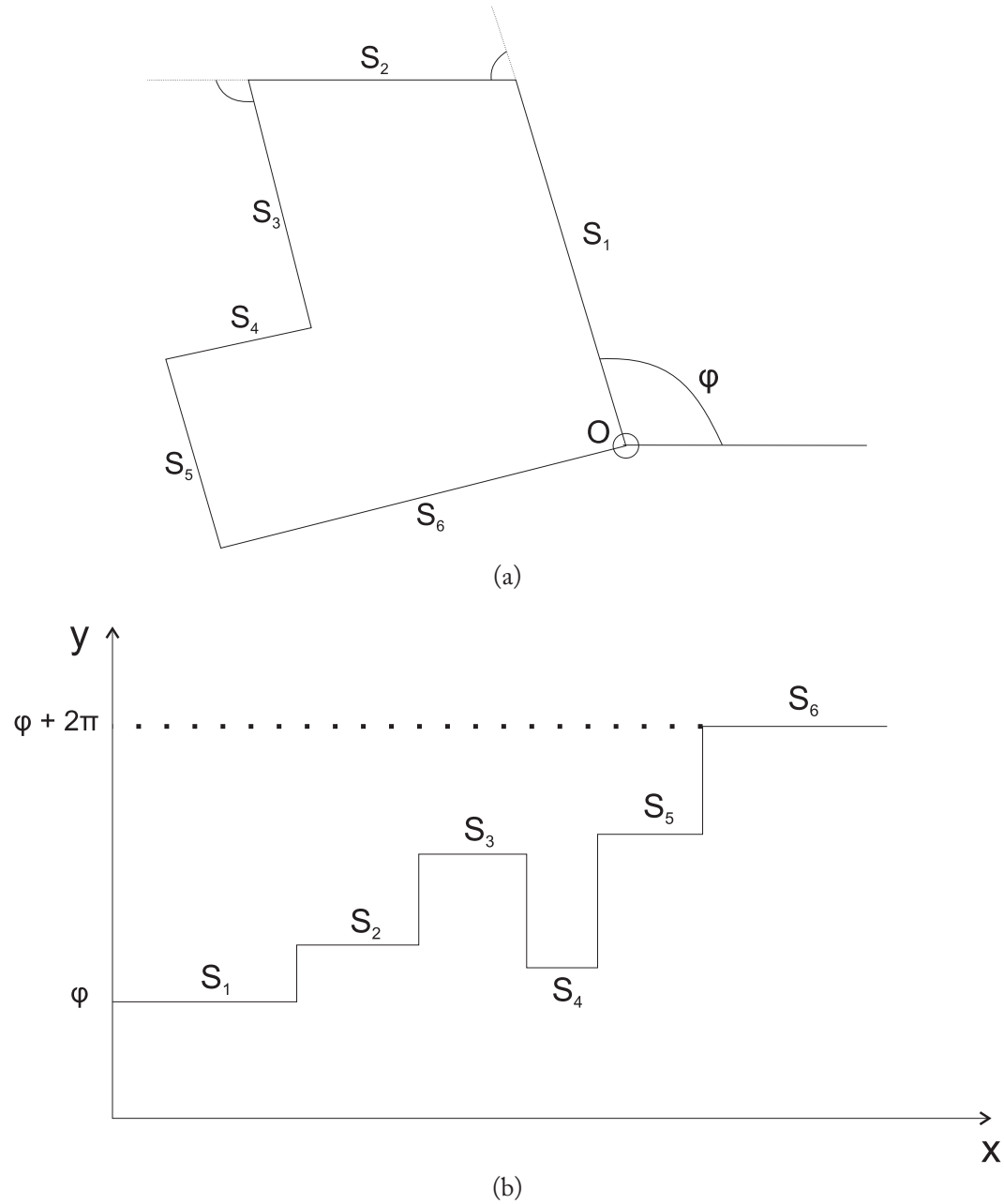
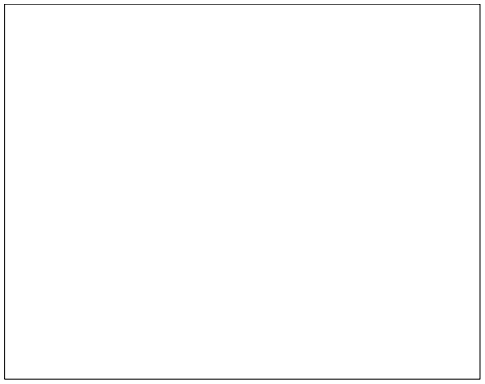
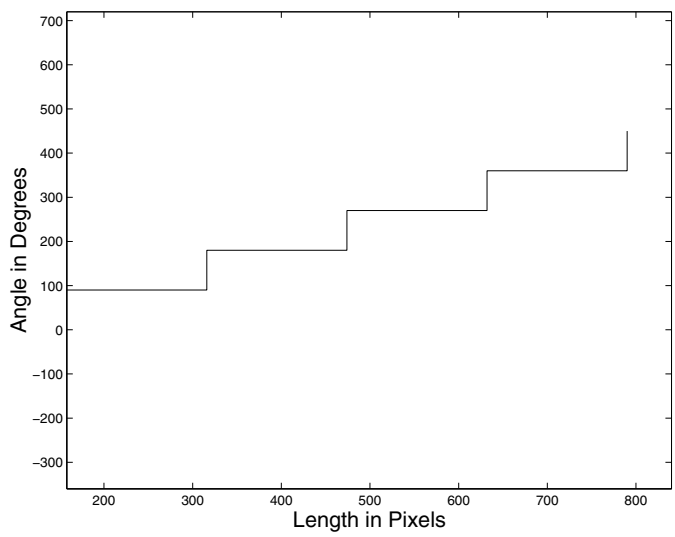


Figure 2.7: (a) A nonconvex contour. (b) The TAF of the contour. The horizontal axis (x) represents the segment length and the vertical axis (y) represents the turning angle.

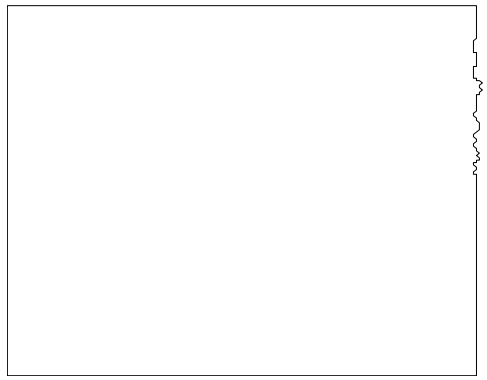


(a)

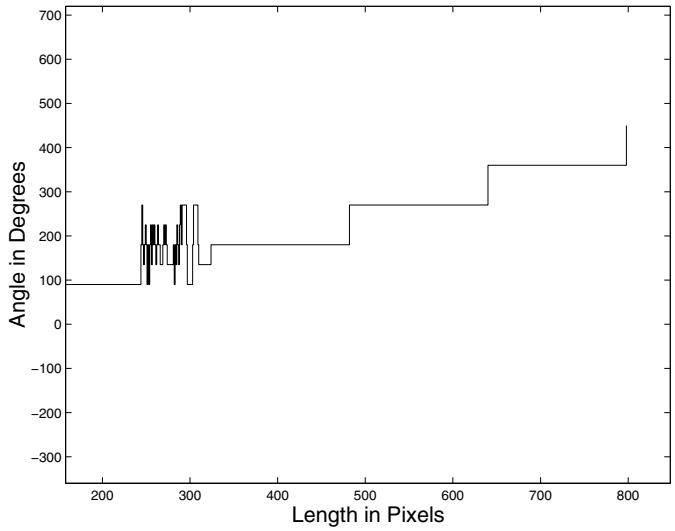


(b)

Figure 2.8: (*Continues.*)

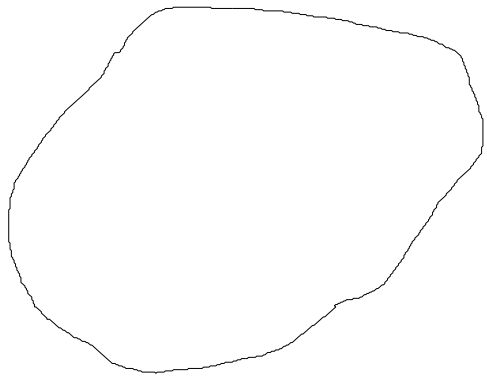


(c)

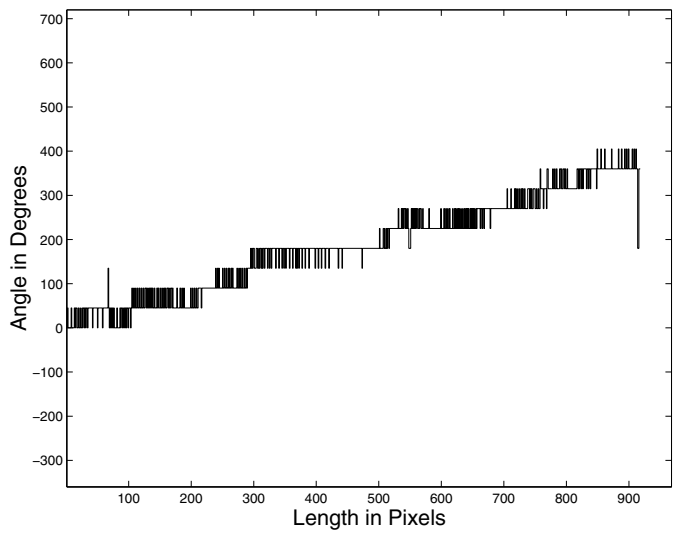


(d)

Figure 2.8: (*Continued.*) (a) A convex contour with $NC = 633$. (b) The TAF of the contour. (c) The contour with the addition of noise, $NC = 641$. (d) The TAF of the noisy contour. Reproduced with permission from R.M. Rangayyan, D. Guliato, J.D. de Carvalho, S.A. Santiago, "Polygonal approximation of contours based on the turning angle function", Journal of Electronic Imaging, 17(2), 023016:1-14, April – June 2008. ©SPIE and IS&T.

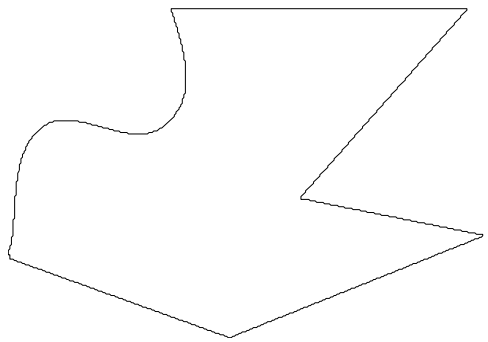


(a)

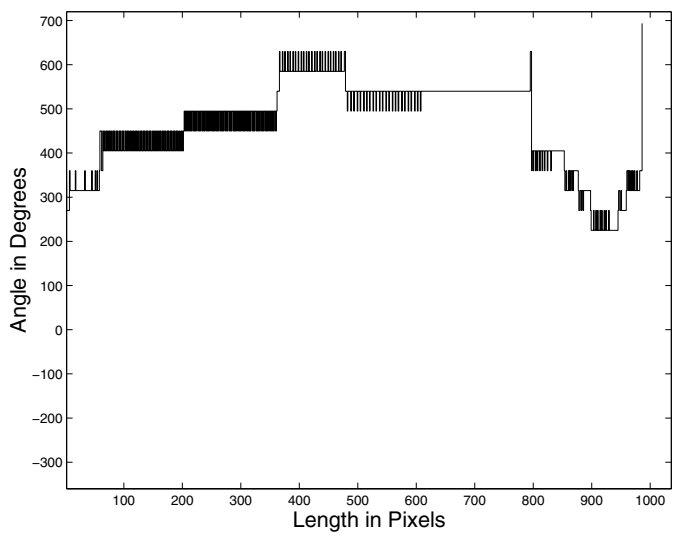


(b)

Figure 2.9: (a) The manually drawn contour of a benign breast mass with a relatively smooth and convex contour with $NC = 916$ and resolution of 50 μm per pixel. (b) The TAF of the contour. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

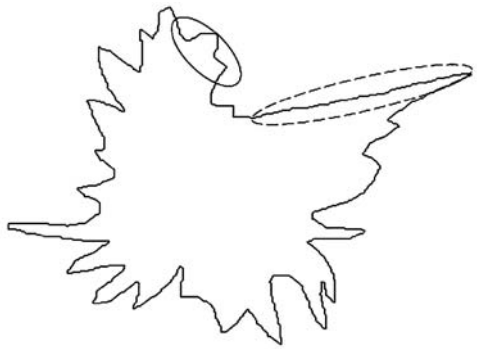


(a)

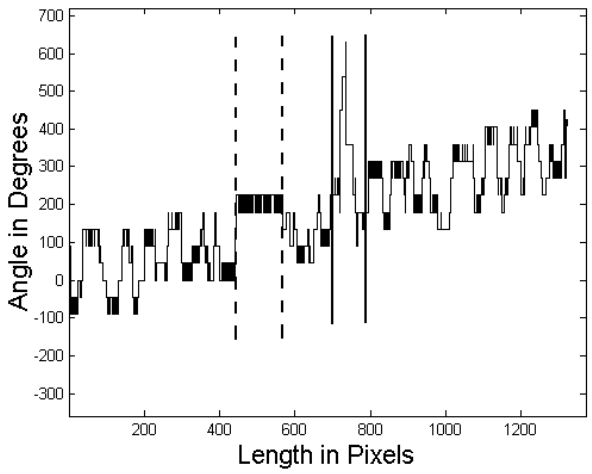


(b)

Figure 2.10: (a) A nonconvex contour with $NC = 983$. (b) The TAF of the contour. Reproduced with permission from R.M. Rangayyan, D. Guliato, J.D. de Carvalho, S.A. Santiago, "Polygonal approximation of contours based on the turning angle function", Journal of Electronic Imaging, 17(2), 023016:1-14, April – June 2008. ©SPIE and IS&T.



(a)



(b)

Figure 2.11: (a) The manually drawn contour of a malignant breast tumor. Adjacent artifactual segments within the dashed ellipse possess high internal angles and are small. Some adjacent segments within the solid ellipse present relevant internal angles. (b) The TAF of the contour. The region in the dashed ellipse is represented in the TAF as the region between the dashed lines with a sequence of small segments with different directions. The region in the solid ellipse is represented in the TAF as a sequence of segments of different sizes with large changes in direction, between the two solid vertical lines. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

combined segment is equal to the angle of the connecting straight line, with respect to the x axis, measured in the counterclockwise direction.

Rule 2: if the length of s_i or s_{i+1} is greater than the threshold $S_{\min}$, then analyze the turning angle between s_i and s_{i+1}. If $\{180° - \text{abs}[T_C(s_{i+1}) - T_C(s_i)]\} \geq \theta_{\max}$, then join s_i and s_{i+1}; else retain s_i and s_{i+1}. The procedure for joining two segments is described in *Rule 1*.

The threshold $S_{\min}$ represents the relevance of a segment and $\theta_{\max}$ indicates the relevance of the turning angle between the two adjacent segments of the contour being analyzed. The relevance of the segment is related to the resolution of the image and the requirements of the application. A high value for $\theta_{\max}$ means that when the internal angle between the two adjacent segments is large, then the segments should be joined. The procedure stops when no segments are joined in an iteration.

Figure 2.12 shows the filtered TAFs of the contours shown in Figures 2.8(a), 2.8(c), and 2.10(a), with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. Whereas the TAF of the rectangle with no noise has remained unaffected by the filtering procedure, all of the segments with length less than 10 pixels and/or internal angle greater than or equal to 170° have been removed in the two examples with noise.

2.3.3 POLYGONAL MODEL FROM THE FILTERED TAF

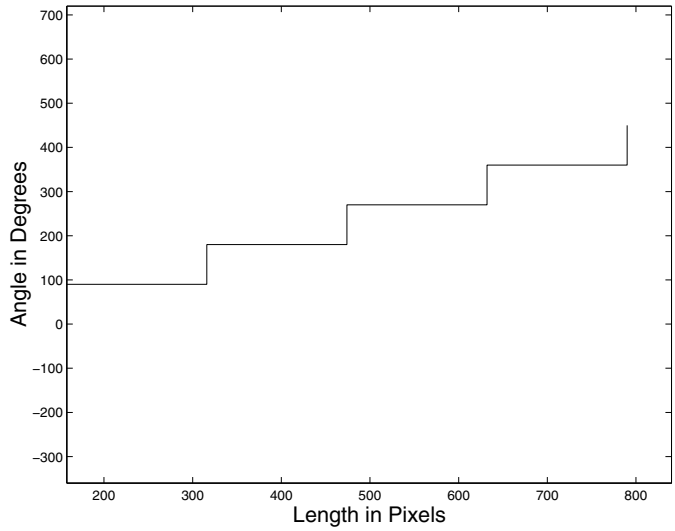
To reconstruct a polygon from its filtered TAF, an arbitrary point $p_1(x, y)$ is chosen as the initial coordinate of the reconstructed contour. The coordinate $p_{i+1}(x, y)$ of the new contour is obtained as

$$\begin{aligned} p_{i+1}(x) &= p_i(x) + d_i \cos[T_C(s_i)], \\ p_{i+1}(y) &= p_i(y) + d_i \sin[T_C(s_i)], \end{aligned} \tag{2.4}$$

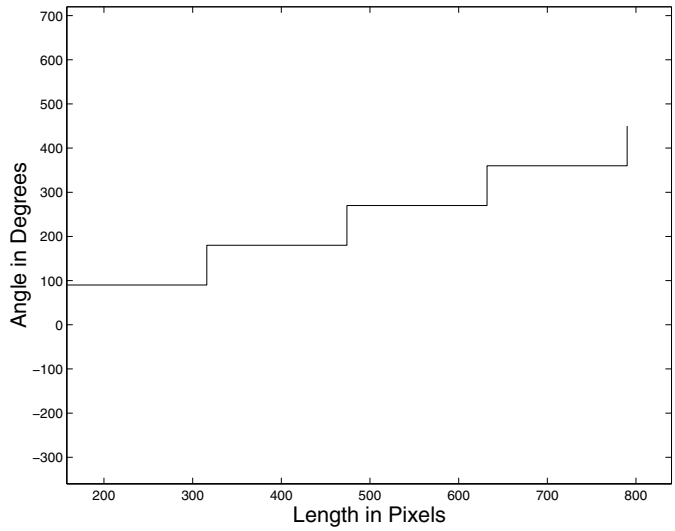
where d_i is the length of the ith segment in the TAF and $T_C(s_i)$ is the turning angle of the segment s_i.

Figure 2.13 illustrates the polygonal models reconstructed from the filtered TAFs shown in Figures 2.12(b) and (c). The number of vertices in each polygonal model (NP) is provided in the caption. Note that, in the case of the convex polygon with noise (see Figure 2.8(c)), the convex nature of the contour has been preserved.

Figures 2.14(a) and 2.15(a) illustrate the filtered versions of the TAFs corresponding to those in Figures 2.9(b) and 2.11(b), with $S_{\min} = 10$ pixels (equivalent to 0.5 mm) and $\theta_{\max} = 170°$. The filtered TAF maintains all of the relevant information required to reconstruct a polygonal model of a given contour with adequate detail [31, 50]. Figures 2.14(b) and 2.15(b) illustrate the polygonal models reconstructed from the respective TAFs. Note that the resulting polygonal models are free of major artifacts and noise; the model preserves important spicules and lobules in the contour of the malignant tumor.

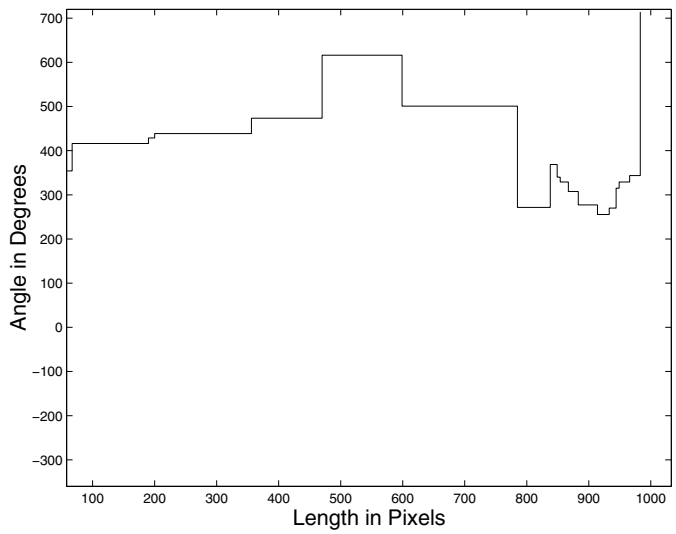


(a)



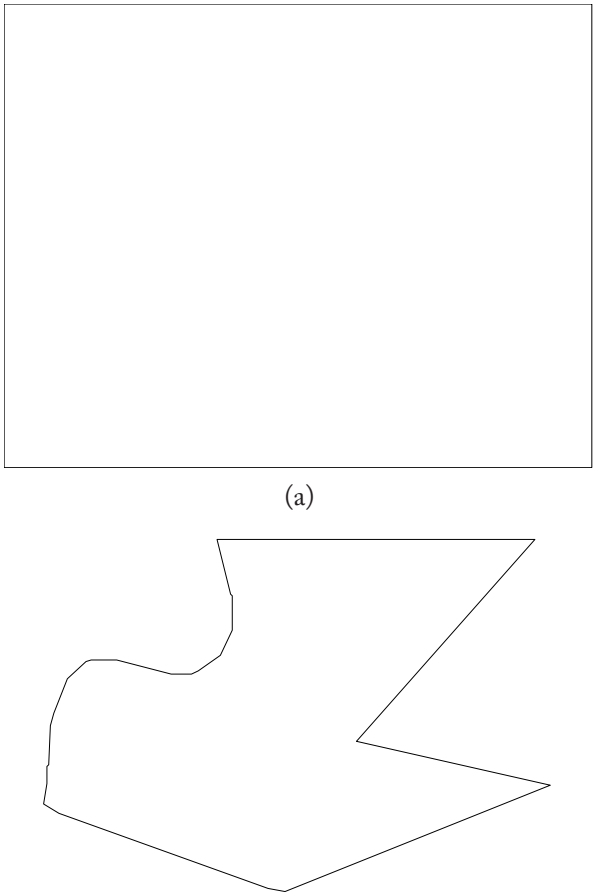
(b)

Figure 2.12: (*Continues.*)



(c)

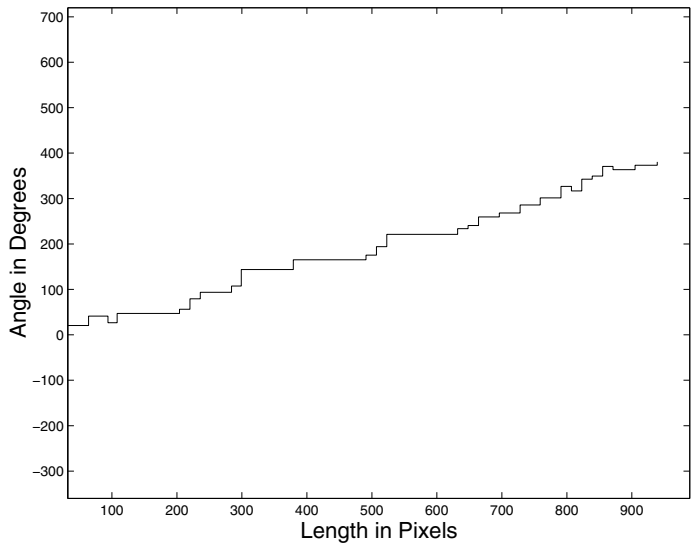
Figure 2.12: (*Continued.*) Filtered TAF for the derivation of a polygonal model with $S_{\min}$ = 10 pixels and $\theta_{\max} = 170°$. (a) The filtered TAF for the convex polygon in Figure 2.8(a). (b) The filtered TAF for the polygon with noise in Figure 2.8(c). Note that all the segments with length less than 10 pixels and/or internal angle greater than or equal to 170° have been removed. (c) The filtered TAF for the nonconvex contour in Figure 2.10(a). Reproduced with permission from R.M. Rangayyan, D. Guliato, J.D. de Carvalho, S.A. Santiago, "Polygonal approximation of contours based on the turning angle function", Journal of Electronic Imaging, 17(2), 023016:1-14, April – June 2008. ©SPIE and IS&T.



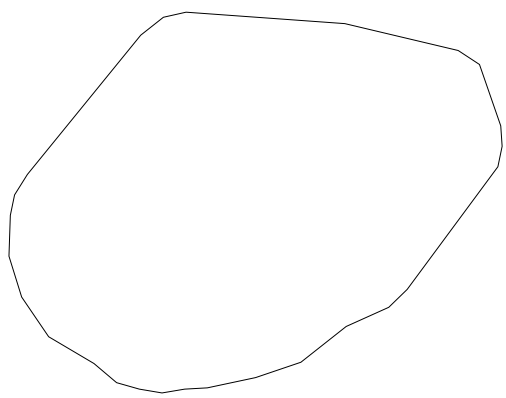
(a)

(b)

Figure 2.13: (a) The polygonal model of the convex polygon with noise obtained using the filtered TAF in Figure 2.12(b). $NP = 4$. See Figures 2.8(c) and 2.8(d) for the noisy contour and its TAF. (b) The polygonal model of the nonconvex contour obtained using the filtered TAF in Figure 2.12(c). $NP = 18$. See Figure 2.10 for the original contour and its TAF. Reproduced with permission from R.M. Rangayyan, D. Guliato, J.D. de Carvalho, S.A. Santiago, "Polygonal approximation of contours based on the turning angle function", Journal of Electronic Imaging, 17(2), 023016:1-14, April – June 2008. ©SPIE and IS&T.

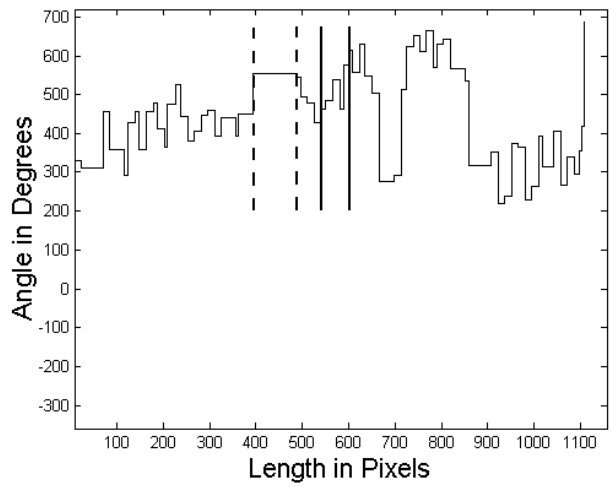


(a)

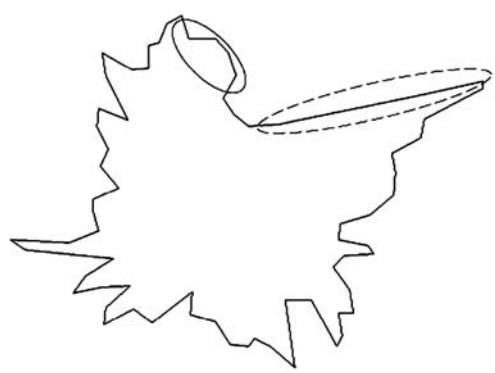


(b)

Figure 2.14: (a) Filtered version of the TAF in Figure 2.9(b) with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (b) Polygonal model of the contour in Figure 2.9(a) with reduced artifacts. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.



(a)



(b)

Figure 2.15: (a) Filtered version of the TAF in Figure 2.11(b) with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (b) Polygonal model of the contour in Figure 2.11(a) with reduced artifacts. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

2.3.4 ILLUSTRATIONS OF APPLICATION

In this section we present a comparison of polygonal models obtained by the TAF method and the methods proposed by Pavilidis and Horowitz [27] and by Rangayyan et al. [14], taking into account the compression rate and the Hausdorff distance. Figure 2.16 presents three noisy contours and their TAFs. Figure 2.17 presents the corresponding filtered TAFs and the polygonal models derived thereof, with $S_{\min} = 15$ pixels and $\theta_{\max} = 150°$. The results provided by the TAF approach are similar to those obtained by the polygonal modeling method described in Section 2.2. However, the polygonal modeling method based on the TAF provides a few advantages:

- it may be used for shape matching in content-based image retrieval (CBIR) systems, and
- it is suitable to derive shape descriptors, as shown in Chapter 3.

Table 2.4 presents the compression rate and the Hausdorff distance obtained for the polygonal models shown in Figure 2.17. The results for the polygonal models obtained via the TAF are better than those for the other methods listed in Tables 2.1 and 2.2. The results are similar to those obtained by the polygonal modeling method proposed by Guliato et al. [31], as listed in Table 2.3.

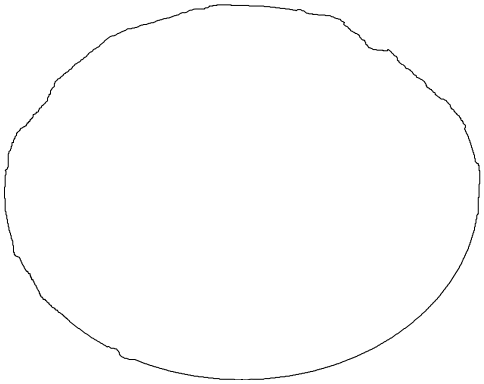
Table 2.4: Results obtained with the polygon modeling method based on the TAF with $S_{\min} = 15$ pixels and $\theta_{\max} = 150°$.

Contour	NP	$h(A, B)$ (pixels)	Cp
Figure 2.17(b)	1777	39.80	0.004
Figure 2.17(d)	2233	18.00	0.001
Figure 2.17(f)	2159	25.22	0.007

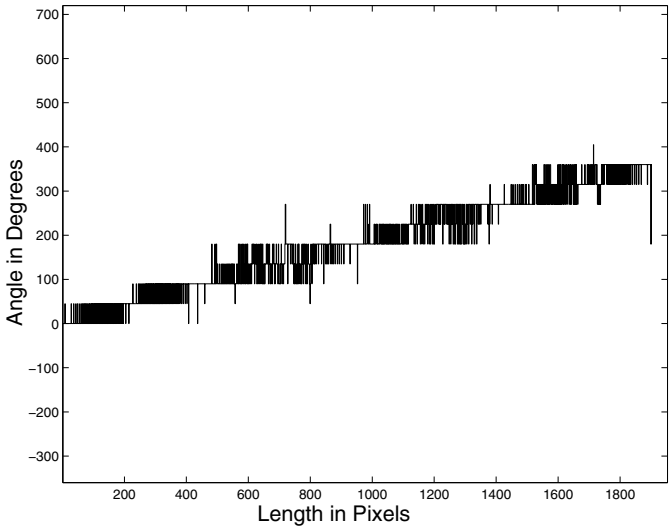
2.4 REMARKS

The polygonal models obtained using the method based on the given contour [31] and the method based on the TAF of the contour [41] may be easily tailored for a given application. By specifying appropriate parameters, both methods are able to remove noise and artifactual variations in contours. The methods have provided better results than those of the other methods described in the present chapter.

Independent of the polygonal modeling method used, the TAF of a polygonal model may be analyzed further to derive quantitative measures. Chapter 3 provides descriptions of several shape features derived from the TAF. The method proposed by Rangayyan et al. [41] is particularly suitable to derive polygonal models; the filtered TAF may be directly used to derive shape features.

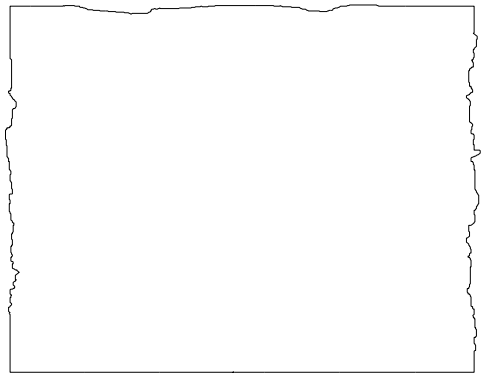


(a)

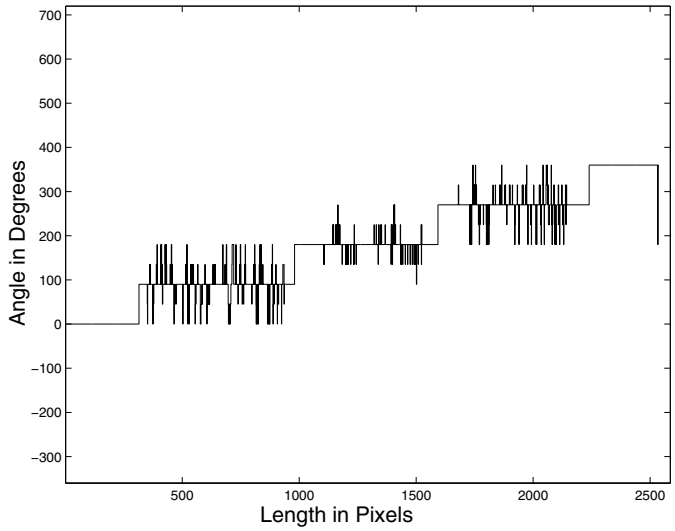


(b)

Figure 2.16: (*Continues.*)

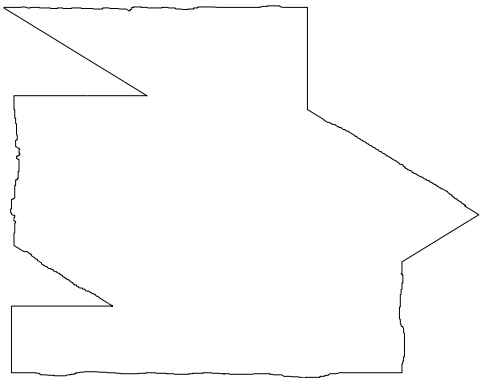


(c)

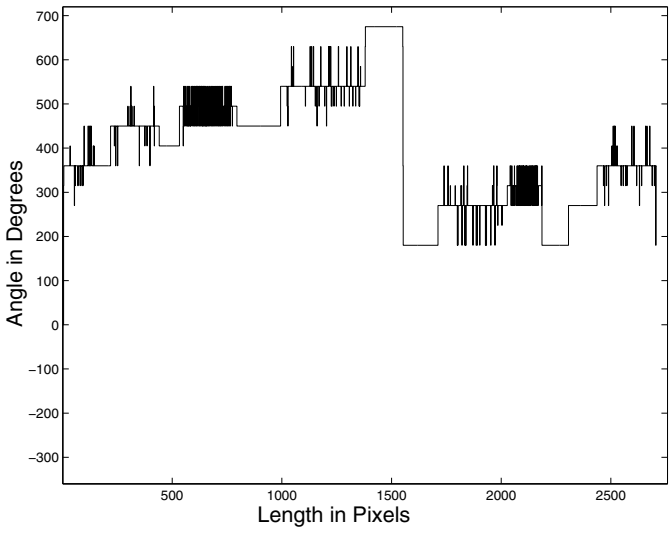


(d)

Figure 2.16: (*Continued.*) – (*Continues.*)

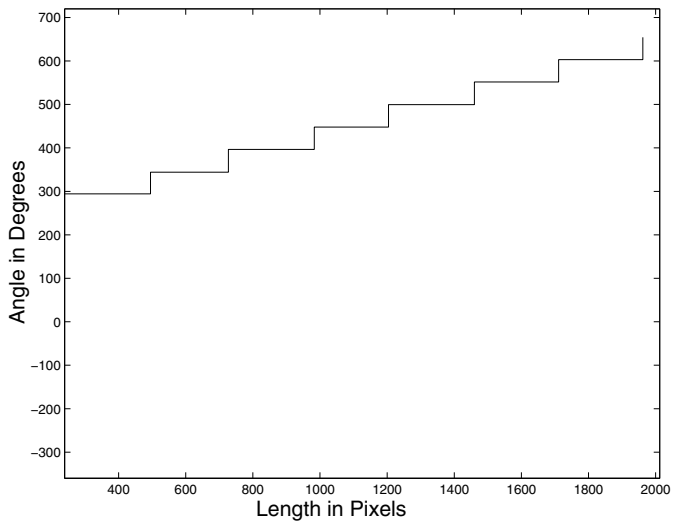


(e)

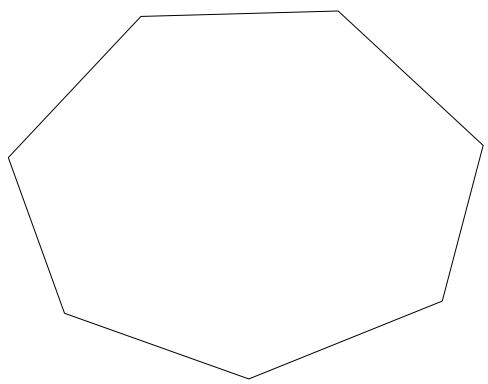


(f)

Figure 2.16: (*Continued.*) Three noisy contours and their TAFs.

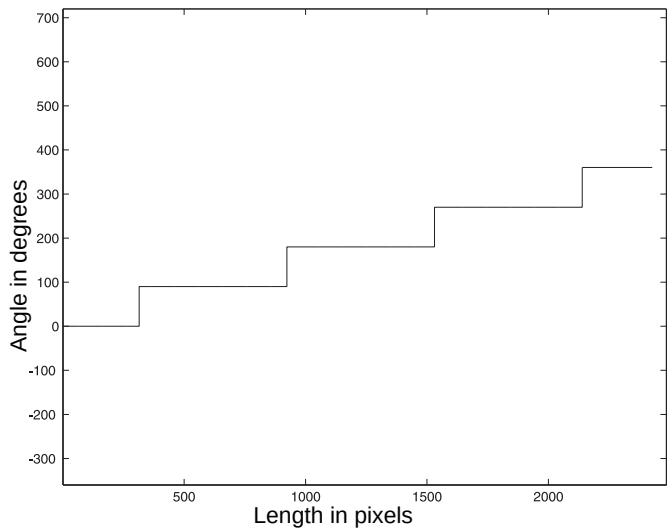


(a)

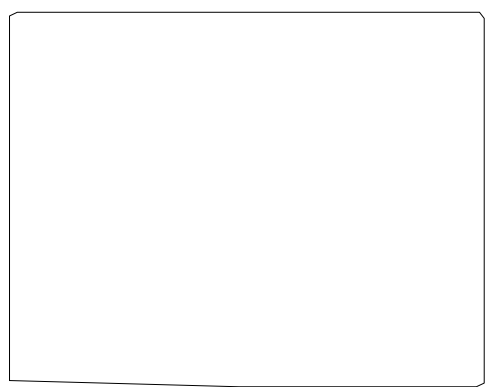


(b)

Figure 2.17: (*Continues.*)

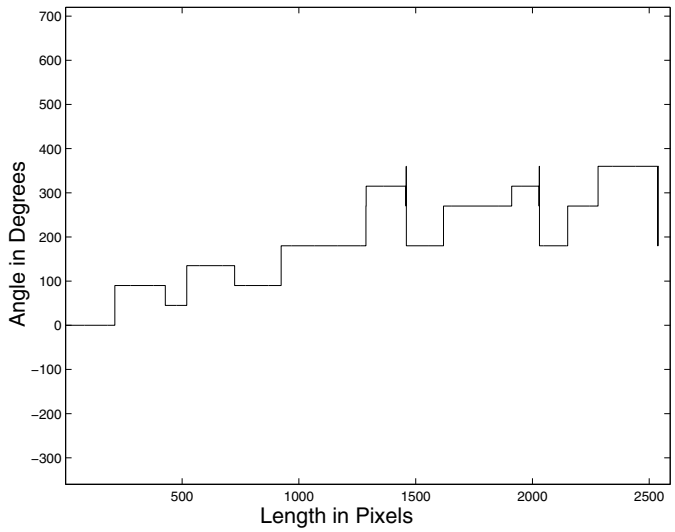


(c)

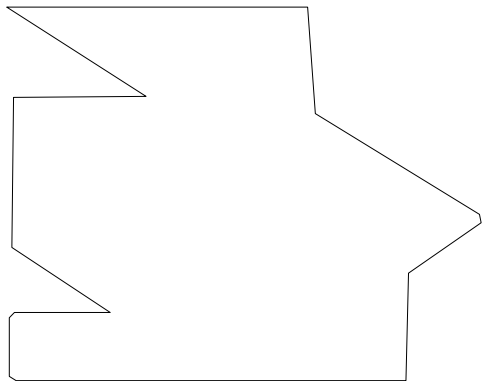


(d)

Figure 2.17: (*Continued.*) – (*Continues.*)



(e)



(f)

Figure 2.17: (*Continued.*) The filtered TAFs and the polygonal models derived thereof corresponding to the cases illustrated in Figure 2.16.

CHAPTER 3

Shape Factors for Pattern Classification

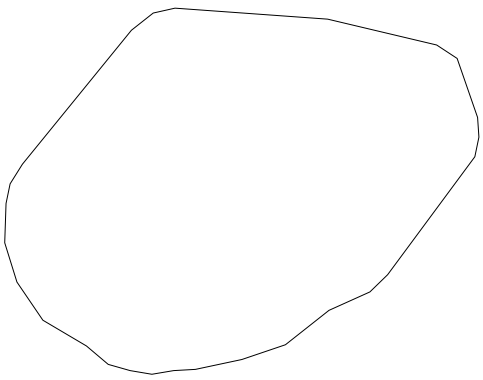
In this chapter, we describe the derivation of several measures of shape complexity from contours and their TAFs. The first two sections describe methods for the derivation of shape features from TAFs. The third section presents a brief review of a few shape factors of contours proposed in previous works in the literature.

3.1 SIGNATURE BASED ON THE FILTERED TAF

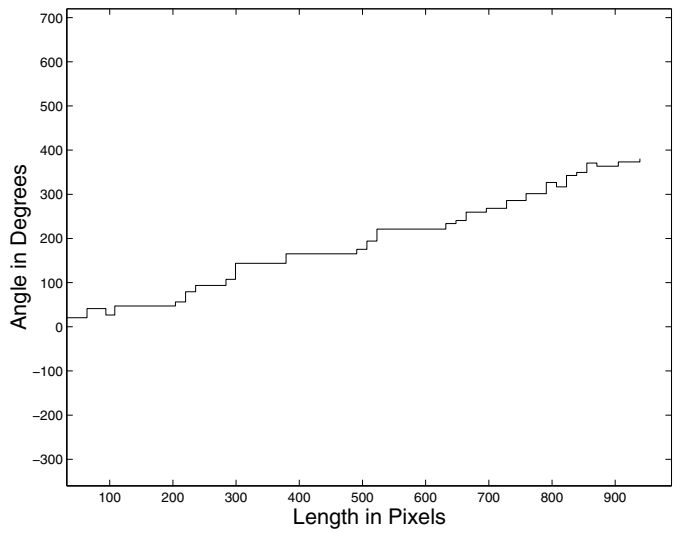
The filtered TAF, as seen in Chapter 2, maintains all of the relevant information required to reconstruct a polygonal model of a given contour [41]. The resulting polygonal model is free of major artifacts and noise, and preserves important spicules and lobules that are present in the contour of a breast tumor. Figures 3.1 and 3.2 illustrate the polygonal models of a nearly convex contour and a nonconvex contour with their respective filtered TAFs. Although a filtered TAF preserves only the significant angles and segments of the corresponding original contour, the successive increasing or decreasing sections do not give any extra information to derive shape factors related to the complexity of the contour, such as FD and index of convexity.

For this reason, Guliato et al. [50] proposed to process further the filtered TAF with the aim of retaining information only about the presence of concave and convex regions in the original contour. The smoothed filtered TAF, referred to as the signature based on the TAF (or STAF, for short), is obtained by replacing each monotonically increasing or decreasing section of the filtered TAF by a representative segment and its turning angle. The length of the new segment is obtained by summing the lengths of all of the related individual segments in the increasing or decreasing section, and the new turning angle is obtained by computing the average of the relative turning angles of the corresponding segments. The STAFs of the polygons in Figures 3.1 and 3.2 are shown in Figure 3.3.

Note that the STAF of a nearly convex contour is almost constant, as illustrated in Figure 3.3(a); on the other hand, the STAF of a contour with concavities possesses several variations, as shown in Figure 3.3(b). This nature of a STAF may be useful to characterize the shape of the related contour. The STAF, as computed above, does not permit the reconstruction of the original contour or any filtered version thereof.

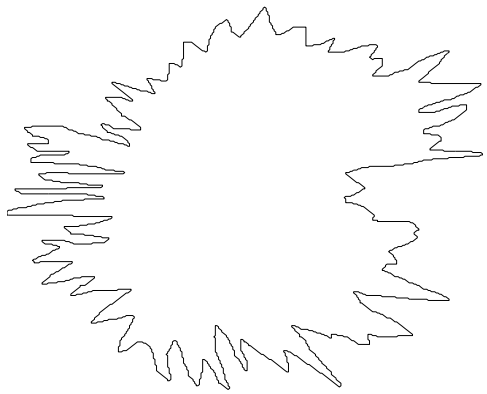


(a)

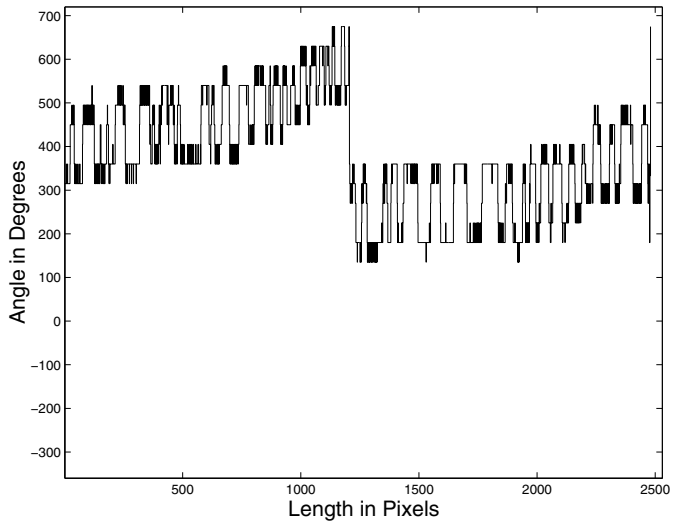


(b)

Figure 3.1: (a) The polygonal model of a benign breast mass. (b) The corresponding TAF. See Figure 2.9 for the original contour and its TAF. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

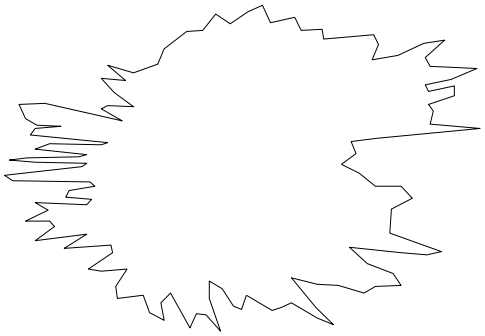


(a)

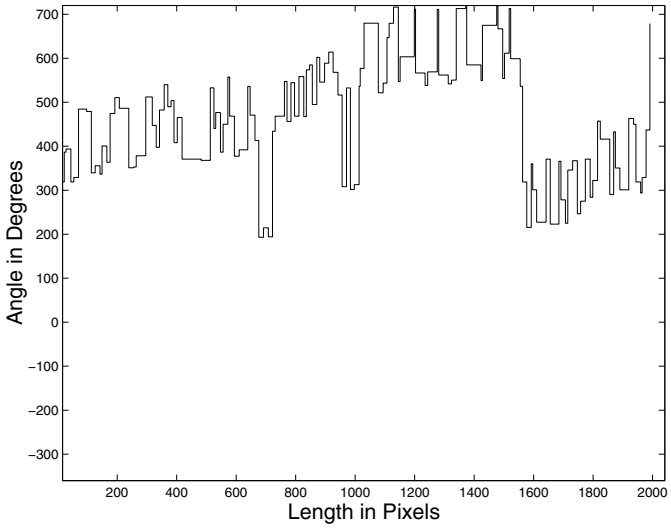


(b)

Figure 3.2: (*Continues.*)

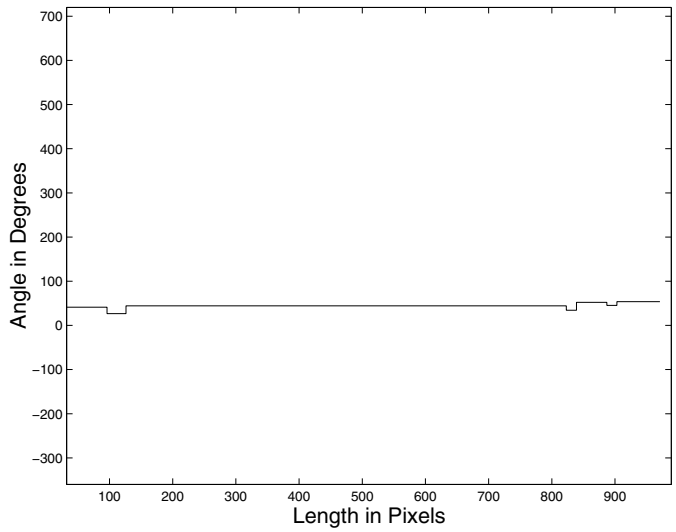


(c)

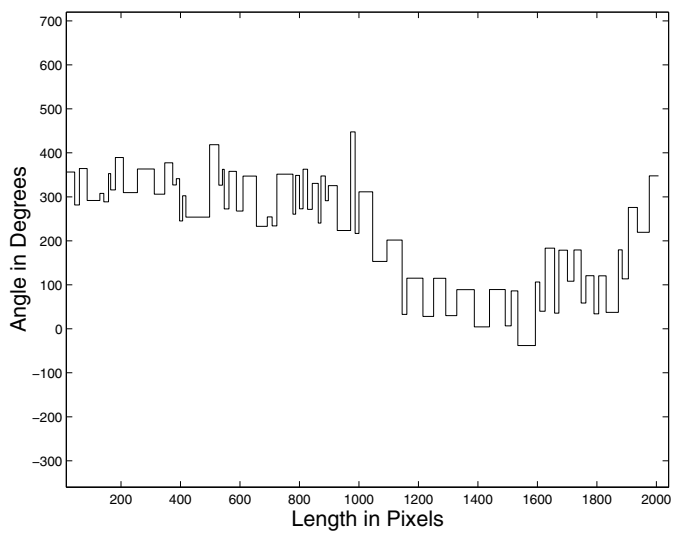


(d)

Figure 3.2: (*Continued.*) (a) The contour of a malignant breast tumor. (b) The corresponding TAF. (c) The polygonal model of the contour in part (a). (d) The TAF of the polygonal model in part (c). Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.



(a)



(b)

Figure 3.3: Signatures based on the TAF with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$ of: (a) the benign mass with a nearly convex contour shown in Figure 3.1, and (b) the malignant tumor with a spiculated contour shown in Figure 3.2. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

3.2 FEATURE EXTRACTION FROM THE STAF

In this section, we describe a set of shape descriptors or features derived from the STAF of the given contour. The features include two different measures of FD, indices that represent the presence of concave and convex regions in the contour, and an index of convexity.

3.2.1 DERIVATION OF AN INDEX OF SPICULATION FROM THE STAF

A spicule is represented in a STAF by a portion starting at a drop in the turning angle and ending at the next increase in the turning angle; see Figure 3.4.

For each spicule p, its length L_p and its angle ψ_p are computed as

$$L_p = \sum_{j=1}^{2} s_j, \tag{3.1}$$

where L_p is the length of the spicule p composed of two segments in the STAF, and

$$\psi_p = 180^\circ - |T_C(s_{p+1}) - T_C(s_p)|, \tag{3.2}$$

where ψ_p is the internal angle of the spicule p.

To derive the feature SI from the polygonal model based on the STAF (SI_{TA}), the length L_p of each possible spicule p is multiplied by $(1 + \cos \psi_p)$. The weighted lengths of the spicules are summed and normalized by twice the sum of their unweighted lengths as

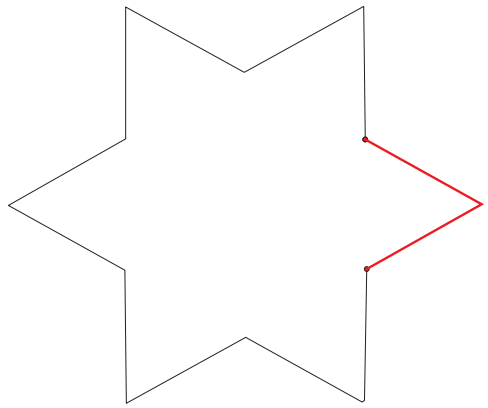
$$SI_{TA} = \frac{\sum_{p=1}^{k} (1 + \cos\ \psi_p)\, L_p}{2 \sum_{p=1}^{k} L_p}, \tag{3.3}$$

where k is the number of spicules in the contour. Note that $0 \leq SI_{TA} \leq 1$.

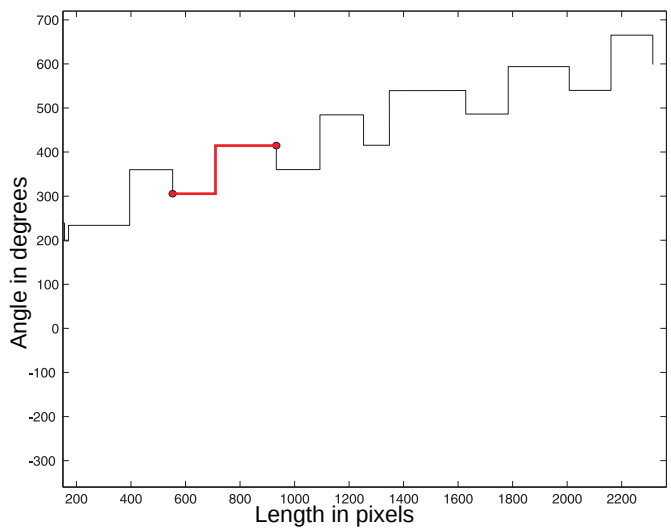
The rough contours of malignant tumors typically possess several narrow and long spicules, whereas the smooth contours of benign masses usually possess no spicules or may have a few broad spicule-like segments. These characteristics should lead to larger values of SI_{TA} for malignant tumors than for benign masses [14].

3.2.2 FRACTAL DIMENSION FROM THE STAF

Fractal analysis may be used to study the complexity and roughness of 1D functions, 2D contours, and images [17, 51, 52, 53, 54, 55, 56, 57]. Fractal analysis may be applied to classify breast masses based on the complexity of their contours [17]. Matsubara et al. [58] obtained 100% accuracy in the classification of 13 breast masses using FD. The method required the computation of a series of FD values for several contours of a given mass obtained by thresholding the mass at many levels; the variation in FD was used to categorize a given mass as benign or malignant. Pohlman et al. [16] obtained a classification accuracy of more than 80%, with fractal analysis of signatures of contours of masses based on the radial distance as described in Section 1.3. Rangayyan and Nguyen [17]



(a)



(b)

Figure 3.4: (a) A stellate or spiculated contour. (b) The STAF of the contour in (a). The red segments identify the parts that compose a spicule in the contour and the corresponding parts of the STAF.

estimated the FD of a set of 111 contours of breast masses and tumors using the ruler and the box-counting methods applied to the 2D contours as well as their 1D signatures ($d(n)$ as described in Section 1.3). The best classification performance with AUC = 0.89 was obtained with the ruler method applied to the 1D signatures of the contours.

In the method proposed by Guliato et al. [50] to obtain FD, the ruler method is applied to the STAFs of the contours of breast masses (referred to as FD_{TA}) and to the first derivative of the STAFs (FD_{dTA}). (See Section 3.3.5 for further details on fractal analysis.) Each STAF is normalized along both axes to the interval [0, 1]. The slope of the curve $\log(r)$ versus $\log(Nr)$, that is, the log of the size of the ruler length, r, versus the log of the number of times, N, that the ruler is used to measure the length (Nr) of the function, is obtained as an estimate of FD_{TA} or FD_{dTA}.

3.2.3 INDEX OF CONVEXITY

Measures of the presence of concave or convex regions may be used to characterize a given contour according to the relevant changes in the direction in the STAF. Related features may be used to classify contours of breast masses as benign or malignant [12, 14]. Such information could be used to discriminate between lobulated or spiculated contours, and between relatively smooth or convex contours. In order to characterize the roughness of a contour, Guliato et al. [50] proposed the features VR_{TA} and XR_{TA} to measure the presence of concave regions and convex regions in a given contour, respectively. Both features are normalized to the interval [0, 1]. The measure VR_{TA} is defined as

$$VR_{TA} = \frac{\sum_{i=1}^{N_d} \{1 + \cos[\theta(i)]\} L_a(i)}{2 \sum_{i=1}^{N_d} L_a(i)}, \tag{3.4}$$

where $L_a(i)$ is the sum of the lengths of two adjacent segments s_i and s_{i+1}, joined by a drop in the turning angle, $\theta(i)$, obtained from the STAF, and N_d is the number of drops in angle in the STAF. For a convex contour, the value for VR_{TA} is equal to zero.

The measure XR_{TA} is defined as

$$XR_{TA} = 1 - \left(\frac{\sum_{j=1}^{N_i} \{1 + \cos[\phi(j)]\} L_b(j)}{2 \sum_{j=1}^{N_i} L_b(j)} \right), \tag{3.5}$$

where $L_b(j)$ is the sum of the lengths of two adjacent segments s_j and s_{j+1}, joined by an increase in the turning angle, $\phi(j)$, obtained from the STAF, and N_i is the number of steps with increasing angles in the STAF. For a convex contour, the value for XR_{TA} is equal to 1.

The index of convexity, CX_{TA}, combines information regarding the presence of concave regions and convex regions in the contour, and is defined as

$$CX_{TA} = \frac{XR_{TA}}{2} + \frac{1 - VR_{TA}}{2}. \tag{3.6}$$

CX_{TA} is normalized to the interval [0, 1]. For a convex contour, the value for CX_{TA} is equal to 1. The index decreases as the presence of concave regions increases.

3.3 SHAPE FACTORS FROM CONTOURS

3.3.1 COMPACTNESS

Compactness (cf) is a measure of how efficiently a contour encloses a given area. A normalized measure of compactness is given by [59]

$$cf = 1 - \frac{4\pi A}{P^2}, \tag{3.7}$$

where P and A are the perimeter of the contour and the area enclosed, respectively. A high compactness value indicates a large perimeter enclosing a small area. Therefore, typical benign masses could be expected to have lower values of compactness as compared to typical malignant tumors [13, 14, 15].

3.3.2 SPICULATION INDEX

Spiculation index (SI) is a measure derived by combining the ratio of the length to the base width of each possible spicule in the contour of the given mass [14]. Let S_n and θ_n, $n = 1, 2, \ldots, N$, be the length and angle of N sets of polygonal model segments corresponding to the N spicule candidates of a mass contour. Then, SI is computed as

$$SI = \frac{\sum_{n=1}^{N} (1 + \cos\theta_n)\, S_n}{\sum_{n=1}^{N} S_n}. \tag{3.8}$$

The factor $(1 + \cos\theta_n)$ modulates the length of each segment (possible spicule) according to its narrowness. Spicules with narrow angles between 0° and 30° get high weighting, as compared to macrolobulations that usually form obtuse angles, and hence get low weighting. The degree of narrowness of the spicules is an important characteristic in differentiating between benign masses and malignant tumors. Benign masses are usually smooth or macrolobulated, and thus have lower values of SI as compared to malignant tumors, which are typically microlobulated or spiculated [14, 15].

3.3.3 FRACTIONAL CONCAVITY

Fractional concavity (f_{cc}) is a measure of the portion of the indented length to the total contour length; it is computed by taking the cumulative length of the concave segments and dividing it by the total length of the contour [14]. The given contour needs to be initially segmented into adjacent concave and convex parts via the detection of points of inflexion [14]. Benign masses have fewer, if any, concave segments than malignant tumors; thus, benign masses could be expected to have lower f_{cc} values than malignant tumors [14, 15].

3.3.4 FOURIER FACTOR

The Fourier factor (ff) is a measure related to the presence of roughness or high-frequency components in contours [59, 60]. The measure is derived by taking the sum of the normalized Fourier

descriptors of the coordinates of the contour pixels divided by the corresponding indices, dividing it by the sum of the normalized Fourier descriptors, and subtracting the result from unity, as follows [59]:

$$ff = 1 - \frac{\sum_{k=-N/2+1}^{N/2} |Z_o(k)|/|k|}{\sum_{k=-N/2+1}^{N/2} |Z_o(k)|}. \tag{3.9}$$

Here, $Z_o(k)$ are the normalized Fourier descriptors, defined as

$$Z_o(k) = \begin{cases} 0, & k = 0; \\ \frac{Z(k)}{Z(1)}, & \text{otherwise.} \end{cases}$$

The Fourier descriptors themselves are defined as

$$Z(k) = \frac{1}{N} \sum_{n=0}^{N-1} z(n) \exp\left[-j \frac{2\pi}{N} nk\right], \tag{3.10}$$

$k = -\frac{N}{2}, \ldots, -1, 0, 1, 2, \ldots, \frac{N}{2} - 1$, where $z(n) = x(n) + jy(n)$, $n = 0, 1, \ldots, N-1$, represents the sequence of contour pixel coordinates. The advantage of this measure is that it is limited to the range [0, 1], and it is not sensitive to noise, which would not be the case if weights increasing with frequency were used. The shape factor ff is invariant to translation, rotation, starting point, and contour size, and increases in value as the shape of the contour gets to be more complex and rough. Contours of malignant tumors are expected to be more rough, in general, than the contours of benign masses; hence, the ff value is expected to be higher for the former than the latter [13, 14, 19].

3.3.5 FRACTAL ANALYSIS

A fractal is a function or pattern that possesses self-similarity at all (or several) scales or levels of magnification [51, 52, 53, 54, 55, 56, 61]. The self-similarity dimension D is defined as follows [52]. Consider a self-similar pattern that exhibits a number of self-similar pieces at the reduction factor $1/s$ (the latter is related to the measurement scale). The power law expected to be satisfied is

$$a = \frac{1}{s^D}. \tag{3.11}$$

Then, we have

$$D = \frac{\log(a)}{\log(1/s)}. \tag{3.12}$$

Therefore, the slope (of the straight-line approximation) of a plot of $\log(a)$ versus $\log(1/s)$ provides an estimate of D. Due to practical limitations, it is important to limit the range of the reduction factor or measurement scale to a viable range [52, 62].

The most commonly used method for estimating FD is the box-counting method [52, 62, 63, 64, 65]. The box-counting method consists of partitioning the pattern or image space into square boxes of equal size, and counting the number of boxes that contain a part (at least one pixel) of the image. The process is repeated with partitioning of the image space into smaller and smaller squares. The log of the number of boxes counted is plotted against the log of the magnification index for each stage of partitioning, yielding a set of points on a line. The slope of the best-fitting straight line to the plot as above gives the FD of the pattern.

Another popular method for calculating FD is the ruler method (also known as the compass or divider method) [52]. With different lengths of rulers, the total length of a contour or pattern can be estimated to different levels of accuracy. When using a large ruler, the small details in a given contour would be skipped, whereas when using a small ruler, the finer details would get measured. The estimate of the length improves as the size of the ruler decreases. Similar to the box-counting method, FD is obtained from the linear slope of a plot of the log of the measured length versus the log of the measuring unit.

Let u be the length measured with the compass setting or ruler size s. The value $1/s$ is used to represent the precision of measurement. The power law expected to be satisfied in this case is

$$u = c\,\frac{1}{s^d}, \tag{3.13}$$

where c is a constant of proportionality, and the power d is related to D as [52]

$$D = 1 + d. \tag{3.14}$$

Applying the log transformation to Equation 3.13, we get

$$\log(u) = \log(c) + d\,\log(1/s). \tag{3.15}$$

Thus, the slope (of the straight-line approximation) of a plot of $\log(u)$ versus $\log(1/s)$ can provide an estimate of FD as $D = 1 + d$.

If we were to denote $u = ns$, where n is the number of times the ruler is used to measure the length u with the ruler of size s, we get

$$\log(n) = \log(c) + (1 + d)\,\log(1/s). \tag{3.16}$$

Then, the slope (of the straight-line approximation) of a plot of $\log(n)$ versus $\log(1/s)$ provides an estimate of D directly.

3.4 REMARKS

In this chapter, we have described methods for the derivation of several different measures of shape complexity from contours and their TAFs. The results of application of the methods to contours of

breast masses and tumors are provided in Chapter 4 along with comparative analysis with respect to shape factors proposed in previous works in the literature.

CHAPTER 4

Classification of Breast Masses

In this chapter, the results of application of the shape factors SI_{TA}, XR_{TA}, CX_{TA}, CV_{TA}, FD_{TA}, and FD_{dTA}, described in Chapter 3, to contours of breast lesions as seen in mammograms are presented, with the aim of evaluating their performance in the classification of breast masses for CAD of breast cancer.

4.1 DATASETS OF CONTOURS OF BREAST MASSES

The dataset of contours of breast masses used in this study includes contours obtained in two preceding studies. One set of contours was derived from mammograms of 20 cases obtained from Screen Test: the Alberta Program for the Early Detection of Breast Cancer [8, 15, 66]. The mammograms were digitized using the Lumiscan 85 scanner at a resolution of 50 μm with 12 b/pixel. The set includes 57 ROIs, of which 37 are related to benign masses and 20 are related to malignant tumors [15]. The sizes of the benign masses vary in the range $39 - 437$ mm^2, with an average of 163 mm^2 and a standard deviation of 87 mm^2. The sizes of the malignant tumors vary in the range $34 - 1122$ mm^2, with an average of 265 mm^2 and a standard deviation of 283 mm^2. Most of the benign masses in this dataset are smooth or macrolobulated, whereas most of the malignant tumors are spiculated or microlobulated.

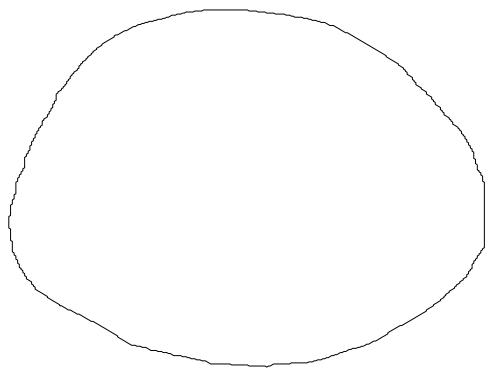
Another set of images was obtained from the Mammographic Image Analysis Society (MIAS, UK) database [67, 68] and the teaching library of the Foothills Hospital (Calgary) [13, 14]. The MIAS images were digitized at a resolution of 50 μm; the Foothills Hospital images were digitized at a resolution of 62 μm. This set includes smooth, lobulated, and spiculated contours in both the benign (28) and malignant (26) categories. The sizes of the benign masses vary in the range $32 - 1207$ mm^2, with an average of 281 mm^2 and a standard deviation of 288 mm^2. The sizes of the malignant tumors vary in the range $46 - 1244$ mm^2, with an average of 286 mm^2 and a standard deviation of 292 mm^2.

The contour of each mass was manually drawn by an expert radiologist specialized in mammography. The combined dataset has 111 contours, including both typical and atypical shapes of benign masses (65) and malignant tumors (46). The diagnostic classification was based upon biopsy. See Rangayyan and Nguyen [17] for illustrations of all of the contours.

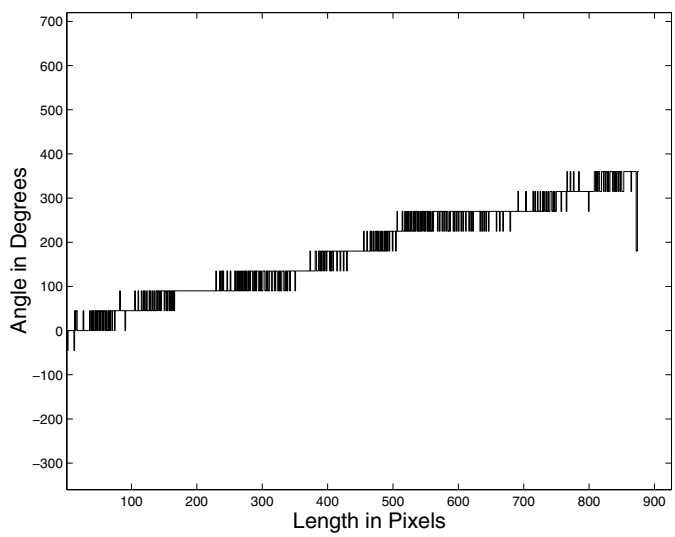
4.2 RESULTS OF SHAPE ANALYSIS AND CLASSIFICATION

To derive the shape factors, firstly, the polygonal model based on the TAF was derived for each of the 111 original contours. The values of $S_{\min}$ and $\theta_{\max}$ required to derive the TAF were set to 10

pixels and 170° for all the contours. Then, the STAF of each TAF and the respective shape factors were derived. Figures 4.1, 4.2, 4.3, and 4.4 show representative contours of benign breast masses and malignant tumors with different shapes, and their respective STAFs. It is worth noting that a convex contour, as shown in Figure 4.1(a), possesses a STAF represented by a constant, resulting in a value equal to zero for SI_{TA}, FD_{TA}, and VR_{TA}, and a value equal to 1 for XR_{TA} and CX_{TA}; see Chapter 3 for details.



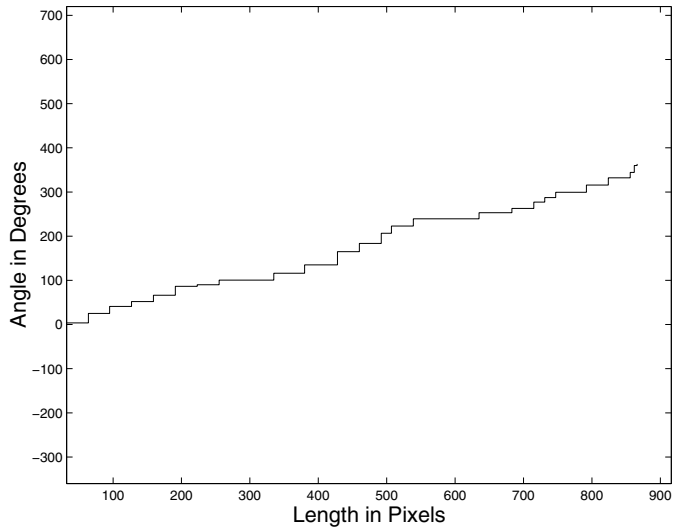
(a)



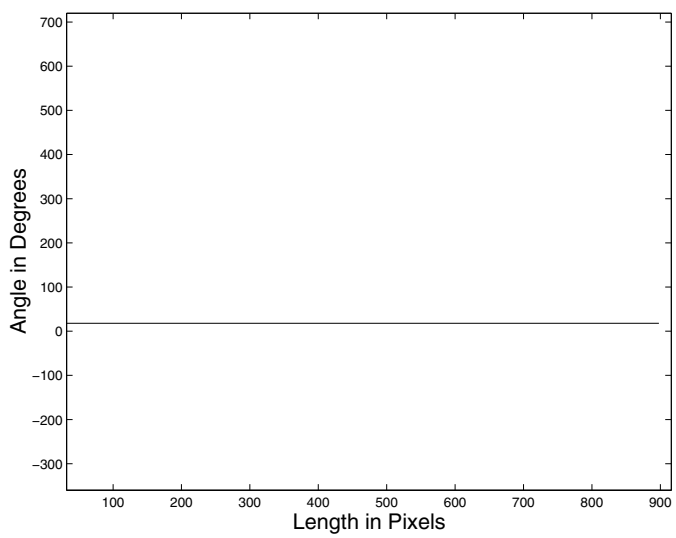
(b)

Figure 4.1: (*Continues.*)

In order to evaluate the efficiency of classification of the shape features, a sliding threshold was applied to each feature directly to classify the corresponding mass as benign or malignant.

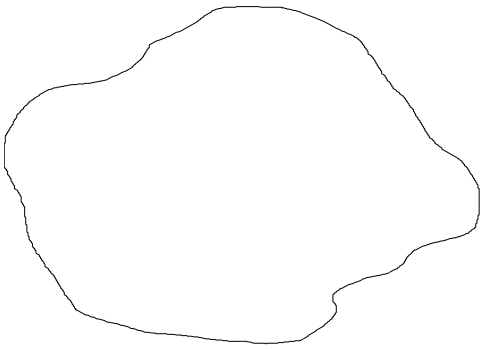


(c)

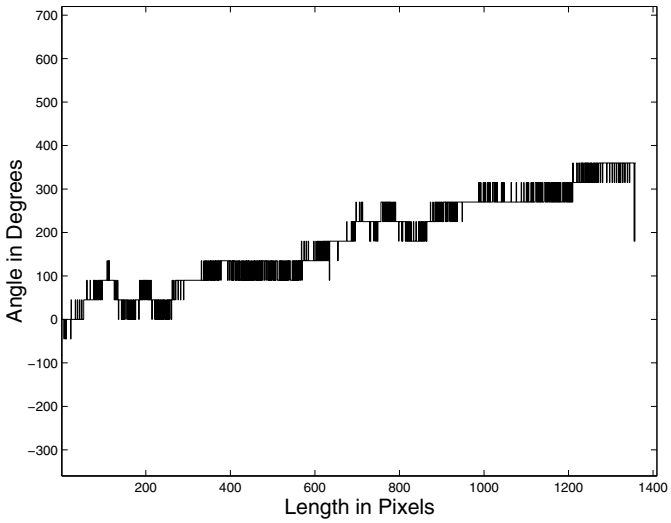


(d)

Figure 4.1: (*Continued.*) (a) The contour of a benign mass that is convex. (b) The TAF of the original contour. (c) The filtered TAF with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (d) The STAF. $SI_{TA} = 0.0$, $FD_{TA} = 0.0$, $VR_{TA} = 0.0$, $XR_{TA} = 1.0$, and $CX_{TA} = 1.0$. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

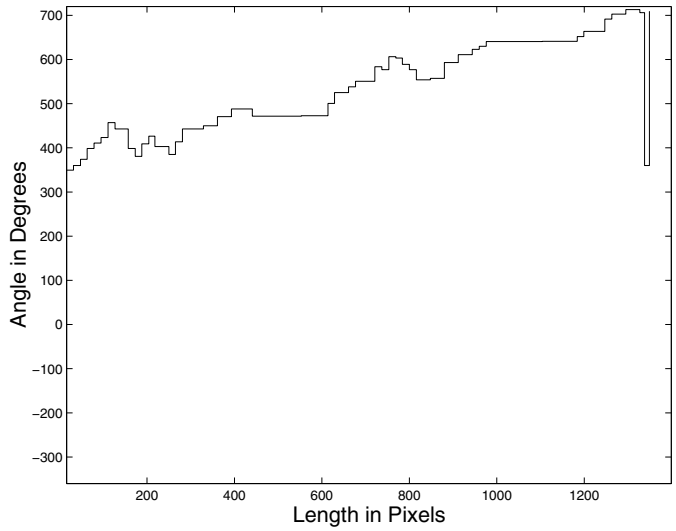


(a)

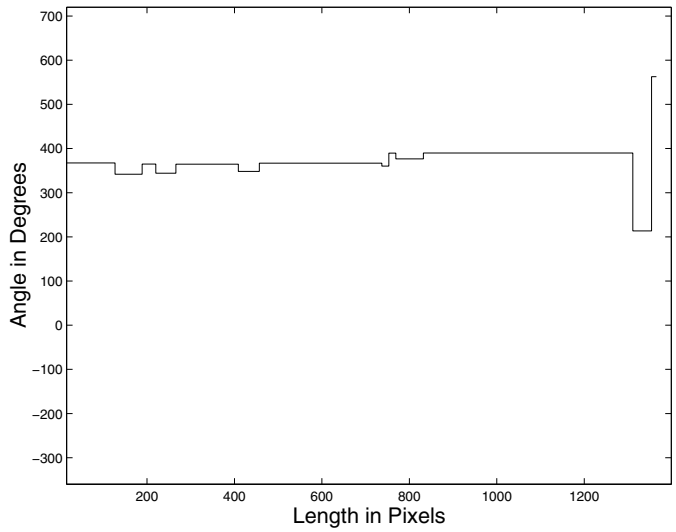


(b)

Figure 4.2: (*Continues.*)

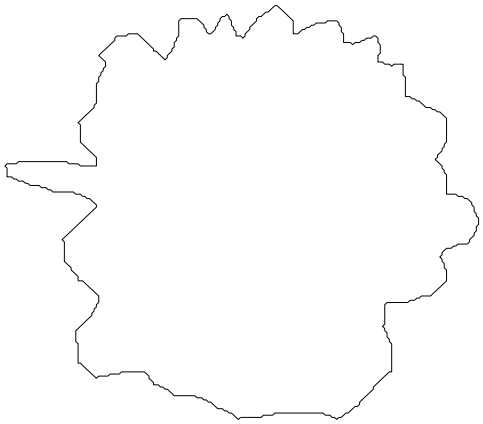


(c)

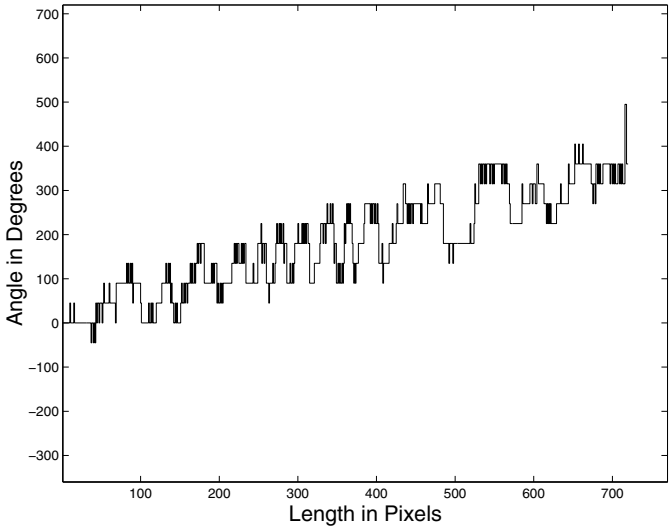


(d)

Figure 4.2: (*Continued.*) (a) The contour of a macrolobulated benign mass. (b) The TAF of the original contour. (c) The filtered TAF obtained with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (d) The STAF. $SI_{TA} = 0.23$, $FD_{TA} = 0.14$, $VR_{TA} = 0.45$, $XR_{TA} = 0.98$, and $CX_{TA} = 0.76$. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

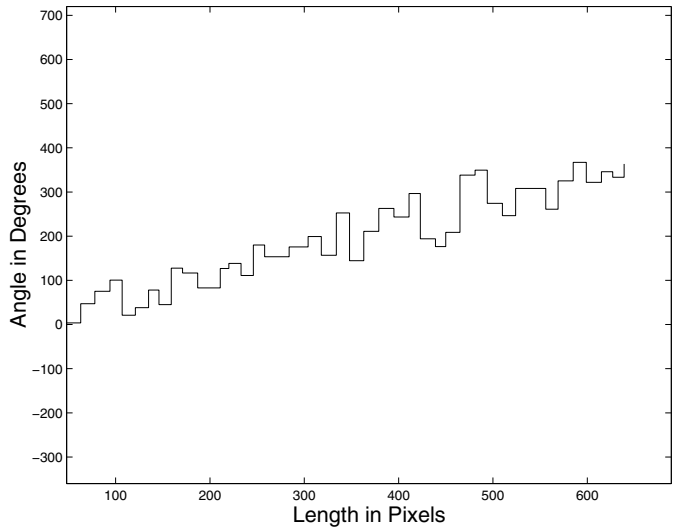


(a)

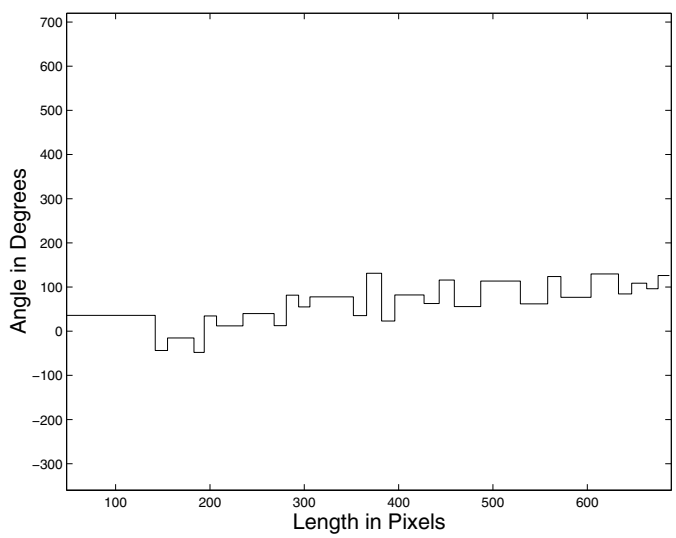


(d)

Figure 4.3: (*Continues.*)



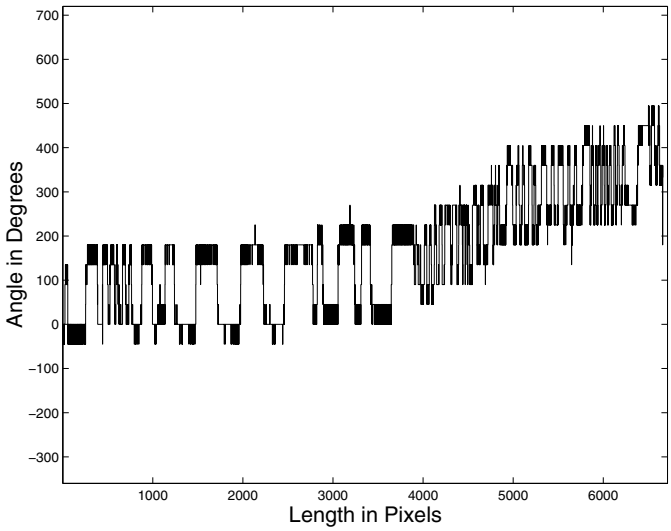
(c)



(d)

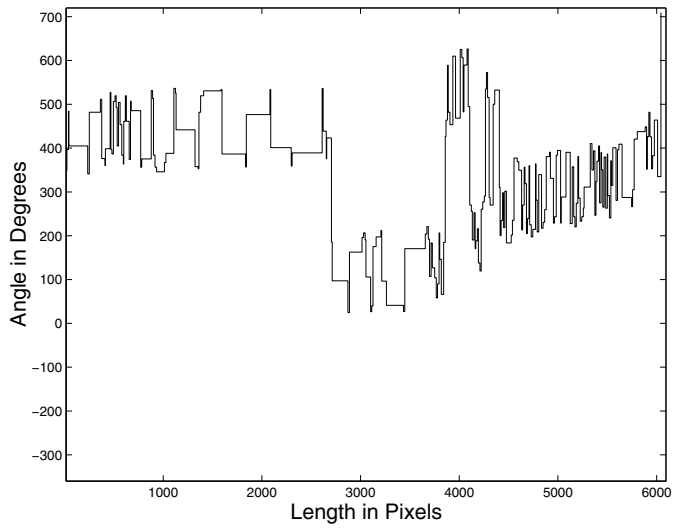
Figure 4.3: (*Continued.*) (a) The contour of a microlobulated malignant tumor. (b) The TAF of the original contour. (c) The filtered TAF obtained with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (d) The STAF. $SI_{TA} = 0.19$, $FD_{TA} = 0.32$, $VR_{TA} = 0.24$, $XR_{TA} = 0.78$, and $CX_{TA} = 0.77$. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129 144, 2008. ©Springer.

(a)

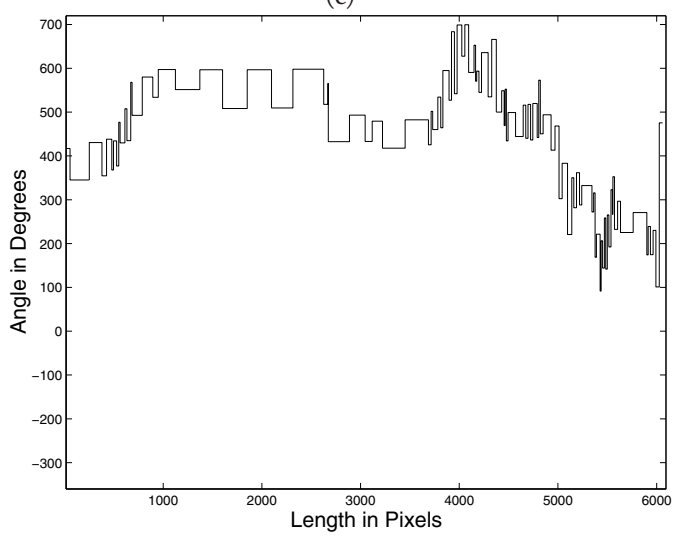


(b)

Figure 4.4: (*Continues.*)



(c)



(d)

Figure 4.4: (*Continued.*) (a) The contour of a spiculated malignant tumor. (b) The TAF of the original contour. (c) The filtered TAF with $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. (d) The STAF. $SI_{TA} = 0.37$, $FD_{TA} = 0.61$, $VR_{TA} = 0.42$, $XR_{TA} = 0.64$, and $CX_{TA} = 0.61$. Reproduced with permission from D. Guliato, J.D. de Carvalho, R.M. Rangayyan, and S.A. Santiago "Feature extraction from a signature based on the turning angle function for the classification of breast tumors", Journal of Digital Imaging, 21(2):129-144, 2008. ©Springer.

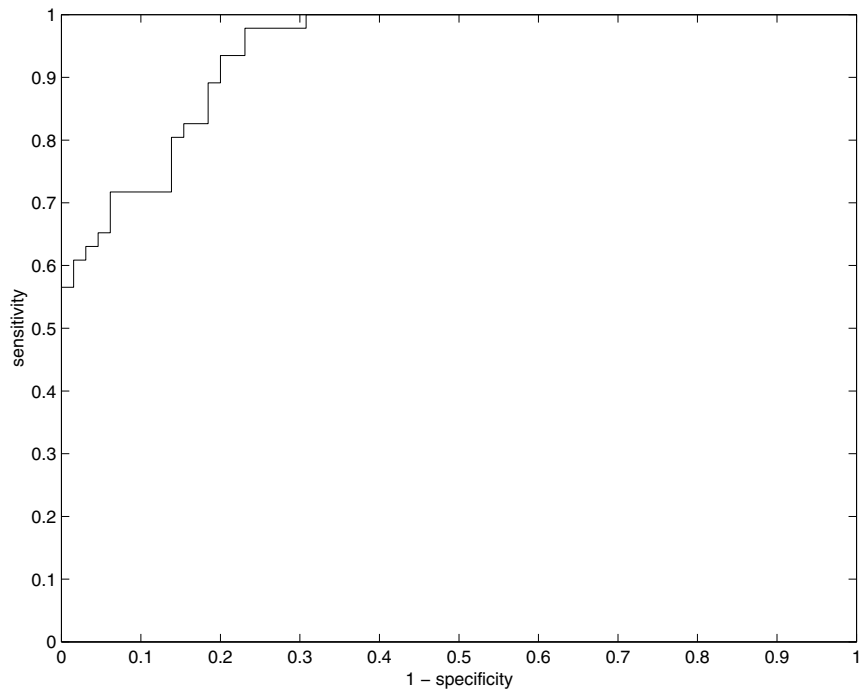


Figure 4.5: ROC curve for the feature SI_{TA} obtained with the parameters $S_{\min} = 10$ pixels and $\theta_{\max} = 170°$. AUC = 0.94. Reproduced with permission from D. Guliato, R. M. Rangayyan, J. D. de Carvalho, S. A. Santiago, "Polygonal modeling of contours of breast tumors with the preservation of spicules", IEEE Transactions on Biomedical Engineering, 55(1):14-20, January 2008. ©IEEE.

This classification strategy was used because, in such an experiment, each contour is represented by only one feature; consequently, the classifier does not require any training step. The diagnosis of each mass, as provided by biopsy, was used to validate the classification. The true-positive fraction (TPF) and false-positive fraction (FPF) were computed for each threshold using the results for all of the 111 contours in the dataset. An ROC curve [69] was generated for each experiment, with the sensitivity given by the TPF and the specificity given as 1− FPF. The procedure was repeated for each feature, and the results obtained are listed in Table 4.1. The area under each ROC curve (AUC) was computed to serve as a measure of the classification performance of the corresponding feature. Figure 4.5 illustrates the ROC curve for the feature SI_{TA}.

In order to compare the classification performance of the features based on the STAF with the performance of other features, the AUC values were obtained for the shape factors cf, ff, f_{cc}, SI, and FD derived using the ruler method applied to 1D signatures of contours [14, 17]. The results are shown in Table 4.1. It is seen that the shape features based on the STAF have provided better results than the other shape factors.

Table 4.1: Comparison of the classification performance of various shape factors. AUC = area under the ROC curve.

Shape factor	AUC
SI [14]	0.91
f_{cc} [14]	0.88
FD [17]	0.89
cf [17]	0.87
ff [17]	0.77
SI_{TA}	0.94
FD_{TA}	0.93
FD_{dTA}	0.92
VR_{TA}	0.92
XR_{TA}	0.92
CX_{TA}	0.93

4.3 REMARKS

We have described methods to obtain polygonal models of contours and to derive shape factors thereof. The procedures could be useful in modeling contours in a wide range of applications. The values of the parameters $S_{\min}$ and $\theta_{\max}$ in the modeling procedure should be defined in relation to the features of the original contours that need to be preserved in the polygonal models for further analysis. In particular, when applied to model contours of breast masses and tumors, the polygonal modeling method based on the TAF preserves diagnostically important features related to spicules and lobulations.

The results presented in this book were obtained using manually drawn contours of breast masses on mammograms. It is desirable to evaluate the methods for modeling and feature extraction with automatically detected contours of masses and with the new modalities of direct digital imaging of the breast. However, the contours of masses obtained automatically using image processing methods for the segmentation of masses [19, 70] may not possess adequate detail and accuracy to permit efficient analysis of shape. Although it is desirable to have the contours of masses drawn by several radiologists and also by the same radiologists several times, this is not feasible in practice.

It is desirable to select efficient combinations of several shape factors and other features of breast masses using feature selection methods to obtain improved and robust classification performance [15, 19, 71]. Advanced pattern classification methods could lead to further improvements in the classification accuracy [72].

Several shape factors derived from the TAFs of polygonal models have provided high classification accuracy in discriminating between benign breast masses and malignant tumors. The methods should be useful in CAD of breast cancer.

References

[1] Levine MD. *Vision in Man and Machine*. McGraw-Hill, New York, NY, 1985. 1

[2] Pavlidis T. *Algorithms for Graphics and Image Processing*. Computer Science Press, Rockville, MD, 1982. 1

[3] Bookstein FL. *The Measurement of Biological Shape and Shape Change*. Springer-Verlag, New York, NY, 1978. 1

[4] Marr D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. WH Freeman, San Francisco, CA, 1982. 1

[5] Rangayyan RM. *Biomedical Image Analysis*. CRC, Boca Raton, FL, 2005. 1, 2, 3, 9

[6] Costa LF and Cesar, Jr. RM. *Shape Analysis and Classification*. CRC Press, Boca Raton, FL, 2001. 1, 9

[7] Homer MJ. *Mammographic Interpretation: A Practical Approach*. McGraw-Hill, Boston, MA, 2nd edition, 1997. 1, 2

[8] Screen Test: Alberta Program for the Early Detection of Breast Cancer – 2001/03 Biennial Report. `http://www.cancerboard.ab.ca/screentest`, 2004. 1, 53

[9] Doi K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Physics in Medicine and Biology*, 51:R5–R27, June 2006. DOI: 10.1088/0031-9155/51/13/R02 1

[10] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31:198–211, 2007. DOI: 10.1016/j.compmedimag.2007.02.002 1

[11] Rangayyan RM, Ayres FJ, and Desautels JEL. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute*, 344:312–348, 2007. DOI: 10.1016/j.jfranklin.2006.09.003 1

[12] American College of Radiology. *Breast Imaging Reporting and Data System* BI-RADS. American College of Radiology, Reston, VA, 4th edition, 2004. 2, 48

[13] Rangayyan RM, El-Faramawy NM, Desautels JEL, and Alim OA. Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810, 1997. DOI: 10.1109/42.650876 2, 4, 9, 49, 50, 53

[14] Rangayyan RM, Mudigonda NR, and Desautels JEL. Boundary modelling and shape analysis methods for classification of mammographic masses. *Medical and Biological Engineering and Computing*, 38:487–496, 2000. DOI: 10.1007/BF02345742 2, 4, 9, 10, 11, 12, 17, 19, 33, 46, 48, 49, 50, 53, 62, 63

[15] Alto H, Rangayyan RM, and Desautels JEL. Content-based retrieval and analysis of mammographic masses. *Journal of Electronic Imaging*, 14(2):023016:1–17, 2005. DOI: 10.1117/1.1902996 2, 49, 53, 63

[16] Pohlman S, Powell KA, Obuchowski NA, Chilcote WA, and Grundfest-Broniatowski S. Quantitative classification of breast tumors in digitized mammograms. *Medical Physics*, 23(8):1337–1345, 1996. DOI: 10.1118/1.597707 3, 46

[17] Rangayyan RM and Nguyen TM. Fractal analysis of contours of breast masses in mammograms. *Journal of Digital Imaging*, 20(3):223–237, September 2007. DOI: 10.1007/s10278-006-0860-9 4, 46, 53, 62, 63

[18] Bruce LM and Adhami RR. Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. *IEEE Transactions on Medical Imaging*, 18(12):1170–1177, 1999. DOI: 10.1109/42.819326 4

[19] Sahiner BS, Chan HP, Petrick N, Helvie MA, and Hadjiiski LM. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics*, 28(7):1455–1465, 2001. DOI: 10.1118/1.1381548 4, 50, 63

[20] Huo Z, Giger ML, Vyborny CJ, Wolverton DE, and Metz CE. Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness. *Academic Radiology*, 7(12):1077–1084, 2000. DOI: 10.1016/S1076-6332(00)80060-4 4

[21] Huo Z, Giger ML, and Vyborny CJ. Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. *IEEE Transactions on Medical Imaging*, 20(12):1285–1292, 2001. DOI: 10.1109/42.974923 4

[22] Rangayyan RM, Guliato D, Carvalho JD, and Santiago SA. Feature extraction from the turning angle function for the classification of breast tumors. In *Proceedings of the International Special Topics Conference on Information Technology in Biomedicine - IEEE ITAB2006*, Ioannina, Greece, October 2006, 6 pages on CDROM. 4, 19, 20

[23] Guliato D, Rangayyan RM, Carvalho JD, and Santiago SA. Spiculation-preserving polygonal modeling of contours of breast tumors. In *Proceedings of the 28th Annual International Conference*

of the IEEE Engineering in Medicine and Biology Society, pages 2791–2794, New York, NY, September 2006. DOI: 10.1109/IEMBS.2006.260441 4, 9, 20

[24] Arkin EM, Chew LP, Huttenlocher DP, Kedem K, and Mitchell JSB. An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:209–216, March 1991. DOI: 10.1109/34.75509 4, 19, 20

[25] Carvalho JD, Rangayyan RM, Guliato D, and Santiago SA. Polygonal modeling of contours using the turning angle function. In *20th IEEE Canadian Conference on Electrical and Computer Engineering*, pages 1090–1267, Vancouver, BC, April 2007. DOI: 10.1109/CCECE.2007.278 4, 19, 20

[26] Pavlidis T and Ali F. Computer recognition of handwritten numerals by polygonal approximations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-5:610–614, November 1975. DOI: 10.1109/TSMC.1975.4309402 9

[27] Pavlidis T and Horowitz SL. Segmentation of plane curves. *IEEE Transactions on Computers*, C-23:860–870, August 1974. DOI: 10.1109/T-C.1974.224041 9, 10, 11, 12, 18, 19, 33

[28] Ramer U. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972. DOI: 10.1016/S0146-664X(72)80017-0 9

[29] Antoine JP, Barache D, Cesar, Jr. RM, and Costa L da F. Shape characterization with the wavelet transform. *Signal Processing*, 62(3):265–290, 1997. DOI: 10.1016/S0165-1684(97)00129-1 9

[30] Latecki LJ and Lakämper R. Application of planar shape comparisons to object retrieval in image databases. *Pattern Recognition*, 35(1):15–29, 2002. DOI: 10.1016/S0031-3203(01)00039-5 9, 20

[31] Guliato D, Rangayyan RM, Carvalho JD, and Santiago SA. Polygonal modeling of contours with the preservation of spicules. *IEEE Transactions on Biomedical Engineering*, 55(1):14–20, 2008. 9, 10, 11, 12, 16, 19, 20, 27, 33

[32] Mehrotra R and Gray JE. Similar-shape retrieval in shape data management. *IEEE Computer*, 28(9):57–62, 1995. DOI: 10.1109/2.410154 9

[33] Cesar, Jr. RM and Costa L da F. Piecewise linear segmentation of digital contours in o(n.log(n)) through a technique based on effective digital curvature estimation. *Real-Time Imaging*, 1:409–417, 1995. DOI: 10.1006/rtim.1995.1042 9

[34] Trimeche M, Cheikh FA, and Gabbouj M. Similarity retrieval of occluded shapes using wavelet-based shape features. In *Proceedings of SPIE 4210: International Symposium on Internet Multimedia Management Systems*, pages 281–289, 2000. DOI: 10.1117/12.403812 9

[35] Safar M, Shahabi C, and Sun X. Image retrieval by shape: a comparative study. In *IEEE International Conference on Multimedia and Expo (ICME)*, volume 1, pages 141–144, 2000. DOI: 10.1109/ICME.2000.869564 9

[36] Ventura JA and Chen JM. Segmentation of two-dimensional curve contours. *Pattern Recognition*, 25(10):1129–1140, 1992. DOI: 10.1016/0031-3203(92)90016-C 9, 10

[37] Menut O, Rangayyan RM, and Desautels JEL. Parabolic modeling and classification of breast tumours. *International Journal of Shape Modeling*, 3(3 & 4):155–166, 1998. 9

[38] Ivarinen J and Visa A. Shape recognition of irregular objects. In *Proceedings of SPIE 2904: Intelligent Robots and Computer Vision XV*, pages 25–32, 1996. DOI: 10.1117/12.256280 9

[39] Heesch D and Ruger S. Combining features for content-based sketch retrieval – a comparative evaluation of retrieval performance. In Crestani F, Girolarni M, and van Rijsbergen CJ, editors, *24th BCS-IRSG European Colloquium on IR Research*, volume LNCS 2291, pages 41–52, Heidelberg, Germany, 2002. Springer-Verlag. 9

[40] Computer Vision Information. 6.4.2.2 Polygonal representations from curves. `http://www.visionbib.com/bibliography/edge244.html`, accessed July 2007. 9

[41] Rangayyan RM, Guliato D, Carvalho JD, and Santiago SA. Polygonal approximation of contours based on the turning angle function. *Journal of Electronic Imaging*, 17(2):023016–1–023016–14, Apr-Jun 2008. DOI: 10.1117/1.2920413 10, 19, 20, 33, 41

[42] Costa L da F and Sandler MB. Effective detection of digital bar segments with Hough transform. *CVGIP: Graphical Models and Image Processing*, 55(3):180–191, May 1993. DOI: 10.1006/cgip.1993.1013 10

[43] Munkres JR. *Topology*. Prentice Hall, Upper Saddle River, NJ, 2000. 12

[44] Niblack W and Yin J. A pseudo-distance measure for 2D shapes based on turning angle. In *IEEE International Conference on Image Processing*, volume 3, pages 352–355, Washington DC, October 1995. DOI: 10.1109/ICIP.1995.537646 20

[45] Veltkamp RC and Tanase M. Part-based shape retrieval. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pages 543–546, Singapore, November 2005. DOI: 10.1145/1101149.1101272 20

[46] Veltkamp RC, Tanase M, and Sent D. Features in content-based image retrieval systems: a survey. In *State-of-the-Art in Content-Based Image and Video Retrieval [Dagstuhl Seminar, 5-10 December 1999]*, pages 97–124, Deventer, The Netherlands, 2001. Kluwer, B.V. 20

[47] Veltkamp RC. Shape matching: Similarity measures and algorithm. In *IEEE SMI 2001 International Conference on Shape Modeling and Applications*, pages 188–197, Genova, Italy, May 2001. DOI: 10.1109/SMA.2001.923389 20

[48] Veltkamp RC. Shape algorithmics. *Algorithmica*, 38(1):1–4, 2003. DOI: 10.1007/s00453-003-1039-0 20

[49] Latecki LJ, Lakämper R, and Wolter D. Shape similarity and visual parts. In *International Conference on Discrete Geometry for Computer Imagery (DGCI)*, pages 34–51, Naples, Italy, 2003. DOI: 10.1007/978-3-540-39966-7_3 20

[50] Guliato D, Carvalho JD, Rangayyan RM, and Santiago SA. Feature extraction from a signature based on the turning angle function for the classification of breast tumors. *Journal of Digital Imaging*, 21(2):129–144, September 2008. DOI: 10.1007/s10278-007-9069-9 27, 41, 48

[51] Mandelbrot BB. *The Fractal Geometry of Nature*. WH Freeman and Company, San Francisco, CA, 1983. 46, 50

[52] Peitgen HO, Jürgens H, and Saupe D. *Chaos and Fractals: New Frontiers of Science*. Springer, New York, NY, 2004. 46, 50, 51

[53] Deering W and West BJ. Fractal physiology. *IEEE Engineering in Medicine and Biology Magazine*, 11(2):40–46, June 1992. DOI: 10.1109/51.139035 46, 50

[54] Schepers HE, van Beek JHGM, and Bassingthwaighte JB. Four methods to estimate the fractal dimension from self-affine signals. *IEEE Engineering in Medicine and Biology Magazine*, 11(2):57–64, June 1992. DOI: 10.1109/51.139038 46, 50

[55] Fortin C, Kumaresan R, Ohley W, and Hoefer S. Fractal dimension in the analysis of medical images. *IEEE Engineering in Medicine and Biology Magazine*, 11(2):65–71, June 1992. DOI: 10.1109/51.139039 46, 50

[56] Goldberger AL, Rigney DR, and West BJ. Chaos and fractals in human physiology. *Scientific American*, 262:42–49, February 1990. DOI: 10.1038/scientificamerican0290-42 46, 50

[57] Rangayyan RM, Banik S, and Desautels JEL. Computer-aided detection of architectural distortion in prior mammograms of interval cancer. *Journal of Digital Imaging*, 23(5):611–631, October 2010. DOI: 10.1007/s10278-009-9257-x 46

[58] Matsubara T, Fujita H, Kasai S, Goto M, Tani Y, Hara T, and Endo T. Development of new schemes for detection and analysis of mammographic masses. In *Proceedings of the 1997 IASTED International Conference on Intelligent Information Systems (IIS'97)*, pages 63–66, Grand Bahama Island, Bahamas, December 1997. DOI: 10.1109/IIS.1997.645180 46

[59] Shen L, Rangayyan RM, and Desautels JEL. Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(6):1403–1416, 1993. DOI: 10.1142/S0218001493000686 49, 50

[60] Shen L, Rangayyan RM, and Desautels JEL. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*, 13(2):263–274, 1994. DOI: 10.1109/42.293919 49

[61] Liu SH. Formation and anomalous properties of fractals. *IEEE Engineering in Medicine and Biology Magazine*, 11(2):28–39, June 1992. DOI: 10.1109/51.139034 50

[62] Coelho RC, Cesar, Jr. RM, and Costa LF. Assessing the fractal dimension and the normalized multiscale bending energy for applications in neuromorphometry. In *Proceedings Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens (SIBGRAPI-96)*, pages 353–554, Caxambu, Brazil, November 1996. 50, 51

[63] Dubuc B, Roques-Carmes C, Tricot C, and Zucker SW. The variation method: a technique to estimate the fractal dimension of surfaces. In *Proceedings of SPIE, Volume 845: Visual Communication and Image Processing II*, volume 845, pages 241–248, 1987. 51

[64] Weinstein RS and Majumdar S. Fractal geometry and vertebral compression fractures. *Journal of Bone and Mineral Research*, 9(11):1797–1802, 1994. DOI: 10.1002/jbmr.5650091117 51

[65] Sedivy R, Windischberger Ch, Svozil K, Moser E, and Breitenecker G. Fractal analysis: An objective method for identifying atypical nuclei in dysplastic lesions of the cervix uteri. *Gynecologic Oncology*, 75:78–83, 1999. DOI: 10.1006/gyno.1999.5516 51

[66] Alto H, Rangayyan RM, Paranjape RB, Desautels JEL, and Bryant H. An indexed atlas of digital mammograms for computer-aided diagnosis of breast cancer. *Annales des Télécommunications*, 58(5-6):820–835, 2003. DOI: 10.1007/BF03001532 53

[67] The Mammographic Image Analysis Society digital mammogram database. `http://peipa.essex.ac.uk/info/mias.html`, accessed October, 2006. 53

[68] Suckling J, Parker J, Dance DR, Astley S, Hutt I, Boggis CRM, Ricketts I, Stamatakis E, Cerneaz N, Kok SL, Taylor P, Betal D, and Savage J. The Mammographic Image Analysis Society digital mammogram database. In Gale AG, Astley SM, Dance DR, and Cairns AY, editors, *Proceedings of the 2nd International Workshop on Digital Mammography*, volume 1069 of *Excerpta Medica International Congress Series*, pages 375–378, York, UK, July 1994. 53

[69] Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, VIII(4):283–298, 1978. DOI: 10.1016/S0001-2998(78)80014-2 62

[70] Mudigonda NR, Rangayyan RM, and Desautels JEL. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging*, 20(12):1215–1227, 2001. DOI: 10.1109/42.974917 63

[71] Sahiner B, Chan HP, Petrick N, Wagner RF, and Hadjiiski L. Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. *Medical Physics*, 27(7):1509–1522, July 2000. DOI: 10.1118/1.599017 63

[72] Mu T, Nandi AK, and Rangayyan RM. Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers. *Journal of Digital Imaging*, 21(2):153–169, June 2008. DOI: 10.1007/s10278-007-9102-z 63

Authors' Biographies

DENISE GULIATO

Denise Guliato is an Associate Professor with the Faculdade de Computação at the Universidade Federal de Uberlândia, Uberlândia, Minas Gerais, Brazil. She received the Bachelor of Computer Science degree from the Universidade Federal de São Carlos, São Carlos, Brazil, the Master's degree in Computer Science from the Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, and the Ph.D. degree in Electrical Engineering from Universidade de São Paulo, São Paulo, Brazil. Her research interests include digital image processing, image analysis, pattern recognition, medical imaging and image analysis, computer-aided diagnosis, content-based image and video retrieval, e-learning systems, and Web services. Her current research projects are on feature extraction, analysis and classification of mammograms; computer-aided diagnosis of breast cancer; and content-based image and video retrieval using bags of visual features.

RANGARAJ M. RANGAYYAN

Rangaraj M. Rangayyan is a Professor with the Department of Electrical and Computer Engineering, and an Adjunct Professor of Surgery and Radiology, at the University of Calgary, Calgary, Alberta, Canada. He received the Bachelor of Engineering degree in Electronics and Communication in 1976 from the University of Mysore at the People's Education Society College of Engineering, Mandya, Karnataka, India, and the Ph.D. degree in Electrical Engineering from the Indian Institute of Science, Bangalore, Karnataka, India, in 1980. His research interests are in the areas of digital signal and image processing, biomedical signal analysis, biomedical image analysis, and computer-aided diagnosis. He has published more than 140 papers in journals and 220 papers in proceedings of conferences. His research productivity was recognized with the 1997 and 2001 Research Excellence Awards of the Department of Electrical and Computer Engineering, the 1997 Research Award of the Faculty of Engineering, and by appointment as a "University Professor" in 2003, at the University of Calgary. He is the author of two textbooks: Biomedical Signal Analysis (IEEE/Wiley, 2002) and Biomedical Image Analysis (CRC, 2005); he has coauthored and coedited several other books. He was recognized by the IEEE with the award of the Third Millennium Medal in 2000, and he was elected as a Fellow of the IEEE in 2001, Fellow of the Engineering Institute of Canada in 2002, Fellow of the American Institute for Medical and Biological Engineering in 2003, Fellow of SPIE: the International Society for Optical Engineering in 2003, Fellow of the Society for Imaging Informatics in Medicine in 2007, Fellow of the Canadian Medical and Biological Engineering Society in 2007, and Fellow of the Canadian Academy of Engineering in 2009. He has been awarded the Killam Resident Fellowship thrice (1998, 2002, and 2007) in support of his book-writing projects.

Index

breast tumor, 1, 2

centroid, 3
classification of breast masses, 53
compactness, 49
compression rate, 12
concavity, 48, 49
contour, 2, 3
contours of breast masses, 53
convexity, 48

Euclidean distance, 3

feature extraction, 46
filtered turning angle function, 27
Fourier descriptors, 50
Fourier factor, 49
fractal analysis, 50
fractal dimension, 46
fractional concavity, 49

Hausdorff distance, 12

inflection
    points of, 10

noise in contours, 36

polygonal modeling, 9, 33
    compression rate, 19
    rules for, 12, 20
    using the turning angle function, 19, 20

shape
    breast tumor, 2
    importance of, 1
shape factors, 49
    comparison, 63
signature, 3, 41
spiculation index, 46, 49

turning angle function, 20
    filtering of, 27