Louisiana State University
LSU Scholarly Repository

LSU Master's Theses | Graduate School

2006

Computer-aided diagnostic systems for digital mammograms

Nicholas Jabari Lee
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_theses

 Part of the Computer Sciences Commons

Recommended Citation

Lee, Nicholas Jabari, "Computer-aided diagnostic systems for digital mammograms" (2006). *LSU Master's Theses*. 336.
https://repository.lsu.edu/gradschool_theses/336

This Thesis is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.

COMPUTER-AIDED DIAGNOSTIC SYSTEMS FOR DIGITAL MAMMOGRAMS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science in Systems Science

in

The Interdepartmental Program in Computer Science

by
Nicholas Jabari Lee
B.S., Jackson State University, 1996
December, 2006

ACKNOWLEDGEMENTS

I feel privileged to have had the opportunity to work on this project. I would like to thank Dr. Tyler for serving as my advisor as well as Drs. Priya Vashishta, Rajiv Kalia, and Aiichiro Nakano for their support, encouragement and advice. I would like to sincerely thank Dr. S. Sitharama Iyengar, Dr. Bijaya B. Karki and Dr. Jianhua Chen for serving on my thesis committee. I would like to thank my mother Verna Lee for her help in preparing this thesis as well as my family for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A computer-aided diagnostic (CAD) system that uses a unique shape-based classification scheme, the Ellipse-Closed Curve Fitting (ECCF) algorithm, is developed for digital mammogram image analysis. The system is developed to work as a post-processing extension to a previously developed CAD system that locates and segments mass lesions, or tumors, found in digital mammograms into separate images. The ECCF system is implemented in the MATLAB mathematical scripting language and is thus capable of running on multiple platforms.

The ECCF algorithm detects edges in tumor images and casts them into closed curve functions. Parameters for an ellipse of best fit for a closed curve function are computed in a way analogous to that in linear regression, where a line of best fit is determined to fit a set of data points.

In addition to the shape-fitting algorithm, the ECCF system comprises several other independently functioning components, including auxiliary algorithms and techniques that perform image cropping and edge detection, employed initially to prepare the images for efficient processing, and self-test tools that calculate R^2, area matching ratios, and a "shape conformity value" to determine the "goodness of fit". Output generated by the ECCF system for sufficiently large image sets may contain correlations between malignant tumors and their shape that may be captured with data mining techniques, the implementation of which may result in an improved integrated CAD system.

CHAPTER 1 - INTRODUCTION

Breast Cancer

Breast cancer is the uncontrolled growth of abnormal cells in the breast. As with other forms of cancer, breast cancer is considered to be a result of malfunctioning DNA due to damage or inherited mutation. Breast cancer is a disease that typically develops in women; however, it is also possible, although rare, for breast cancer to develop in men. According to the World Health Organization, more than 1.2 million people worldwide will learn they have breast cancer this year. The American Cancer Society estimates women in the United States will account for approximately 213,000 of these cases. The National Cancer Institute (NCI) reports breast cancer as the most common type of cancer among women in the US, second only to skin cancer [1].

Breast Cancer Statistics

Breast cancer ranks second to lung cancer as the leading cause of death in women diagnosed with cancer in the US. About 41,000 women in the US are expected to die from the disease in 2006.[2] The number of cases of women with breast cancer has been increasing. In 2005, 211,240 women in the US were diagnosed with breast cancer, compared to ~7,522 women in 1975, which comes out to an average increase of about 0.4% per year. However, over the last decade, due to increased awareness, screening, and improved treatments, the number of deaths due to breast cancer has been decreasing overall.[3] Figure 1.1 shows death rates due to breast cancer in comparison to other types of cancer over the last seven decades.

Currently there is no cure for breast cancer, and it is also not possible to predict when or if a person will develop the disease. Early detection and treatment are currently

the only means proven to reduce breast cancer related mortality rates. It is therefore important for women especially to avoid the risk factors for breast cancer, monitor themselves for its symptoms, and get screened periodically for the disease.

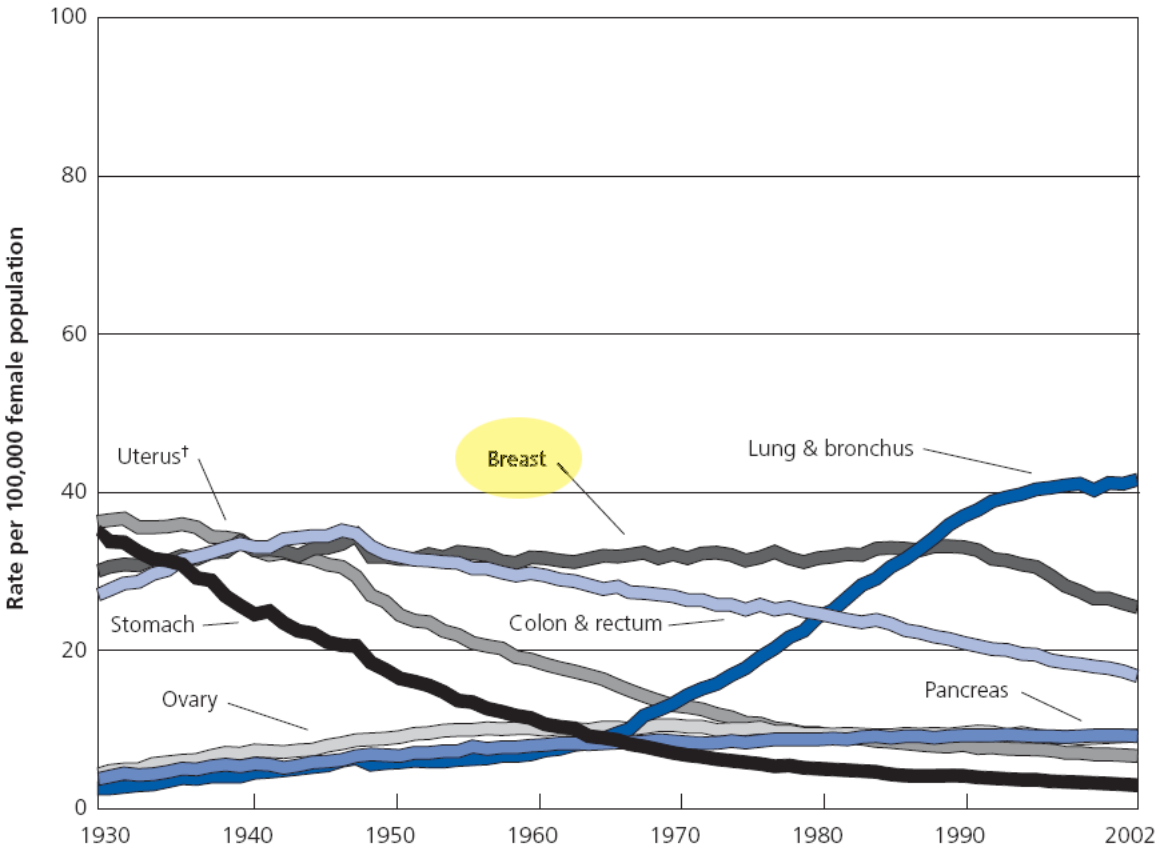


Figure 1.1: Age-adjusted cancer death rates of women in the US between 1930-2002. Overall, deaths due to breast cancer have been declining since 1990 yet it remains the second leading killer of women diagnosed with cancer in the US.

Mammography and Breast Cancer Screening

Mammography is the study that involves identifying structures within the breast to classify them as benign or malignant. A mammogram is an image obtained by using X-rays to probe the breast. During a screening for breast cancer radiologists, or readers, inspect mammogram for areas that may indicate further investigation through biopsy, a surgical procedure from which the diagnosis and subsequent prognosis is obtained.

Routine mammography reduces the breast cancer mortality rate by 25% to 35% in asymptomatic middle-aged women. The National Cancer Institute recommends annual mammograms for women above 50 years of age. However, medical science is not exact. Readings are subject to error, thus diagnostics are not 100% accurate. Nevertheless, the field of mammography has benefited from continuous advances in technology over time.

Today, two kinds of mammography are in clinical use, conventional film mammography and digital mammography. Each has comparative advantages and disadvantages with respect to the other. Despite its drawbacks, digital mammography is however potentially superior to film mammography particularly as it can be used along with CAD (computer-aided detection or computer-aided diagnostic) systems. CAD systems have been demonstrated as effective tools for helping radiologist identify malignancies in mammograms. Research activity in digital mammography and CAD systems is relatively new, beginning in the late 1980's. The area is now nevertheless recognized as an important area of research in and across a number of industrial and academic sites around the world.

Basically, the goal of development in these areas is to improve the precision of the CAD process. In this thesis we present a CAD and an attempt to improve the performance of this CAD system using a tumor shape analysis algorithm.

Chapter Outline

The outline of the remaining portion of this thesis is as follows: Chapter 2 will cover aspects of breast cancer treatment. Conventional film mammography, digital mammography, and computer-aided diagnostic systems will be covered in Chapter 3. Descriptions of the CAD system developed by Sample and Tyler and the post-screening

shape analysis (PSA) system will be described in Chapter 4. Results and conclusions are presented in Chapter 5.

CHAPTER 2 – BREAST CANCER TREATMENT

Risk Factors and Prevention

As previously stated, gender is the most significant factor affecting the likelihood of one developing breast cancer. High blood levels of estrogen have been associated with breast cancer. Women are about 90 times more likely to develop breast cancer than men. Besides gender, age is the next most strongly correlated risk factor of breast cancer. The odds of a female developing breast cancer before age 39 is 1 in 209. By the time she becomes 40 those odds increase by almost 9-fold to 1 in 24, as can be seen in Figure 2.1. The US National Cancer Institute suggests that women have their first mammogram at age 35, every two years after age 40, and then annually after 50.

Genetics also plays a role. Women with a personal or family history of breast cancer are being advised to have more frequent and even extensive examinations. It is usually suggested to start screening when a women reaches 10 years less than the age at which the relative was diagnosed with breast cancer. Exposure to radiation or various carcinogenic chemicals can increase the likelihood of developing breast cancer. Reproductive factors also play a role in the chances of developing breast cancer. Women that start menstruating or enter menopause later than average are more likely to get breast cancer. Women who have children later in life or who do not breast-feed are more likely to get breast cancer. Women have control over some significant risk factors of breast cancer in their modifiable habits. It is suggested that woman quit, or never start, smoking and reduce their exposure to 2nd and 3rd hand smoke as much as possible. Excessive alcohol consumption, more than one alcoholic beverage per day, has also been shown to have a positive correlation with the occurrence of breast cancer. Also women who are

obese are more likely to develop breast cancer. Thus a healthy diet and regular exercise may also decrease the risk of developing cancer.

In some cases, knowing the risk factors can help women reduce their likelihood of developing breast cancer by making lifestyle choices. In cases where genetics or the environment creates a predisposition to the disease, women should have earlier and more frequent exams as well as closer monitoring for symptoms.

		Birth to 39 (%)	40 to 59 (%)	60 to 69 (%)	70 & Older (%)	Birth to Death (%)
All sites†	Male	1.43 (1 in 70)	8.57 (1 in 12)	16.46 (1 in 6)	39.61 (1 in 3)	45.67 (1 in 2)
	Female	1.99 (1 in 50)	9.06 (1 in 11)	10.54 (1 in 9)	26.72 (1 in 4)	38.09 (1 in 3)
Urinary bladder†	Male	.02 (1 in 4375)	.40 (1 in 250)	.93 (1 in 108)	3.35 (1 in 30)	3.58 (1 in 28)
	Female	.01 (1 in 9513)	.12 (1 in 816)	.25 (1 in 402)	.96 (1 in 104)	1.14 (1 in 88)
Breast	Female	.48 (1 in 209)	4.11 (1 in 24)	3.82 (1 in 26)	7.13 (1 in 14)	13.22 (1 in 8)
Colon & rectum	Male	.07 (1 in 1399)	.90 (1 in 111)	1.66 (1 in 60)	4.94 (1 in 20)	5.84 (1 in 17)
	Female	.06 (1 in 1567)	.70 (1 in 143)	1.16 (1 in 86)	4.61 (1 in 22)	5.51 (1 in 18)
Leukemia	Male	.15 (1 in 650)	.22 (1 in 459)	.35 (1 in 284)	1.17 (1 in 85)	1.50 (1 in 67)
	Female	.13 (1 in 788)	.14 (1 in 721)	.19 (1 in 513)	.78 (1 in 129)	1.07 (1 in 93)
Lung & bronchus	Male	.03 (1 in 3244)	1.00 (1 in 100)	2.45 (1 in 41)	6.33 (1 in 16)	7.58 (1 in 13)
	Female	.03 (1 in 3103)	.80 (1 in 125)	1.68 (1 in 60)	4.17 (1 in 24)	5.72 (1 in 17)
Melanoma of skin	Male	.13 (1 in 800)	.51 (1 in 195)	.51 (1 in 195)	1.25 (1 in 80)	1.94 (1 in 52)
	Female	.21 (1 in 470)	.40 (1 in 248)	.26 (1 in 381)	.56 (1 in 178)	1.30 (1 in 77)
Non-Hodgkin lymphoma	Male	.14 (1 in 722)	.47 (1 in 215)	.56 (1 in 178)	1.57 (1 in 64)	2.18 (1 in 46)
	Female	.09 (1 in 1158)	.31 (1 in 320)	.42 (1 in 237)	1.29 (1 in 77)	1.82 (1 in 55)
Prostate	Male	.01 (1 in 10149)	2.66 (1 in 38)	7.19 (1 in 14)	14.51 (1 in 7)	17.93 (1 in 6)
Uterine cervix	Female	.15 (1 in 657)	.28 (1 in 353)	.15 (1 in 671)	.22 (1 in 464)	.74 (1 in 135)
Uterine corpus	Female	.06 (1 in 1641)	.72 (1 in 139)	.83 (1 in 120)	1.36 (1 in 74)	2.61 (1 in 38)

Figure 2.1: Probability of developing invasive cancers over selected age intervals and by gender US populations based on cases diagnosed during 2000 to 2002. Subjects are cancer free at the beginning of an age interval. 1 in 24 women between the ages of 40 and 59 are likely to develop breast cancer. The chance of developing breast cancer increases with age. 1 out of 8 women will develop breast cancer at some point over their life time.

Symptoms

Breast cancer may be accompanied by any of a number of symptoms. Abnormal lumps persisting in the breast are perhaps one of the most commonly associated symptoms. Skin dimpling, unusual changes in texture or skin color on the breast, change in the shape of the nipple, and discharge of blood through the nipple may also be

symptoms of breast cancer as shown in Figure 2.2. In some cases of breast cancer, however, there are no noticeable symptoms. In fact, half of women who get breast cancer experience no obvious symptoms and discover their breast cancer only after undergoing a medical examination.[4] Therefore, it is important for women to have periodic screenings for breast cancer.

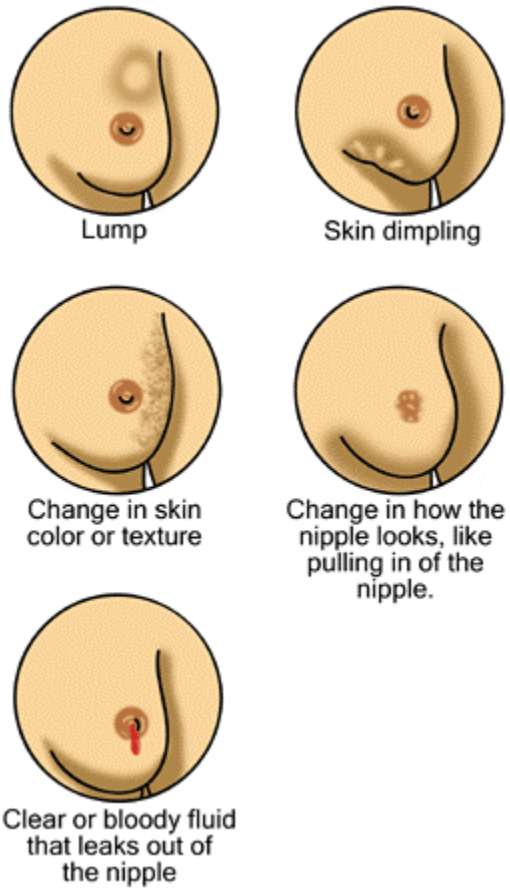


Figure 2.2: Illustrations of early signs of breast cancer. (Image originally from the National Institutes for Health.)

Breast Cancer Screening

Self-examinations for breast cancer are a very important part of health maintenance. They should be performed regularly and frequently as a first line of defense

against breast cancer. However, since about 50% of breast cancers go undetected by self-examination, professional medical screenings should also be used in conjunction with self-examination for precaution. Breast cancer screening is a professional medical examination performed to check women's breasts for abnormalities such as tumors and cysts and identify malignancies where they exist. It is widely available in the U.S. and is highly recommended as it has been proven to significantly reduce fatalities due to breast cancer.

Several techniques can be used to examine the breast including ultrasound, which uses a band of high frequency sound waves to probe the breast; magnetic resonance imaging, which probes the breast using powerful magnetic fields; and mammography, which is essentially producing X-ray photographs of the breast. Mammography will be discussed in more detail in Chapter 3. A screening will basically involve mammography or getting a mammogram, but some cases may involve all three examination methods for thorough investigations. In any case, the screening process is simply used to find abnormalities in the breast, not to obtain a diagnosis.

Diagnosis

A diagnosis is the process of identifying a disease by its signs, symptoms and results of various diagnostic procedures. The conclusion reached through that process as to whether a tumor is malignant (cancerous) or benign (non-cancerous) is also called a diagnosis. Diagnoses can only be obtained through biopsy with present technologies and procedures. The term biopsy refers in general to a procedure where tissue is removed for examination from areas in the body where cancer is suspected. There are specific biopsies for different sites in the body: liver, skin, bone marrow, prostate and various others. For

the breast alone there are several kinds of biopsies including fine-needle aspiration, nipple aspirates, ductal lavage, core needle biopsy, and local surgical biopsy. Each has comparative advantages and limitations. No method of biopsy is perfect. Sometimes cancers can still be missed if an ample amount of tissue is not sampled or if the tissue is sampled from the wrong place altogether.

Particular questions a biopsy is intended to resolve are: What's the size of the tumor? Is the tumor cancerous? Are lymph nodes involved? Once cancer has been found at a site, there is the question of metastasis, or in other words, has the cancer spread to other places in the body from that site and how aggressively is it spreading. The spreading of cancer is called metastasis. Methods used to check for metastasis are chest x-ray, bone scan, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography. Tumor markers are also used to trace where cancer has spread in the body. Data collected from biopsies is used to determine the progress of breast cancer, which is described in stages. The stage of breast cancer is defined in terms of inspected tumor size, tumor grade, hormone receptor status, HER2/neu oncogene over expression, and margins of resection. The TNM system is one of the most commonly used staging systems, accepted by the International Union Against Cancer (UICC) and the American Joint Committee on Cancer (AJCC). The system assesses tumors based on their tumor size,

There are four stages of breast cancer as shown in Table 2.1 In stage 0, the cancer is said to be in situ. That is, the cancer is only present at the site where it was first detected and has not spread to other places in the body. A tumor is described as stage I if it has spread beyond the margins into the surrounding tissue. A tumor is categorized as

stage II if it is invasive between 2 and 5 centimeters and has spread from the original site to other parts of the body. Stage III tumors have sizes greater than 5 centimeters. If a cancerous tumor has grown into the chest wall or spread to lymph nodes, then it is classified as stage IV. T category is another classification scheme used to indicate the development of cancer. The two classification systems are similar as they both relate to the size of the tumor. The chart below shows how the stage and T category relate.

Table 2.1: Physicians use staging to indicate the size and location of a patient's cancer. The AJCC-TNM classification system is shown in the table above.

Stage	**T category**	**Tumor size t_s (cm)**	**Description**
0	T0	$t_s < 2$	Size of tumor is less than 2 centimeters in diameter and is in situ.
I	T1	$t_s < 2$	Size of tumor is less than 2 centimeters in diameter and has spread beyond margins.
II	T2	$2 < t_s < 5$	Size of tumor is between 2 and 5 t centimeters in diameter and has spread beyond margins.
III	T3	$t_s > 5$	Size of tumor is greater than 5 centimeters in diameter and has spread beyond margins.
IV	T4	$t_s > 0$	Tumor is any size, has attached to the chest wall and spread to the pectoral lymph nodes.

Treatment

A number of treatments are available for breast cancer including surgery, radiation therapy, chemotherapy, hormonal therapy and immunotherapy. The kind of treatment carried out basically depends on the stage of cancer. It is not uncommon for more than one kind of treatment to be used in combination. There are two types of surgical procedures, the lumpectomy and the mastectomy. In a lumpectomy, the tumor and its surrounding tissue, the margin, is removed from the breast. The second kind of

surgical procedure is the mastectomy, where the entire breast is removed. Lymph nodes are also removed in a mastectomy to determine the stage of the disease more precisely. As a precautionary measure, radiation therapy is usually recommended for invasive cancers that have spread into the surrounding tissue in addition to lumpectomy or mastectomy. Radiation therapy involves using high-powered X-rays to bombard the site of the cancer. Radiation is intended to halt, or decrease the rate of growth, of a cancerous tumor by denaturing the DNA of cells. Chemotherapy is another form of treatment that uses cytotoxic, or cell killing, drugs to rid cancer from the body. In some cases it is administered before surgery to reduce the size of the tumor. More often it is given after surgery to help reduce the chances of the tumor reforming. Chemotherapy is responsible for hair-loss during cancer treatments, although a common misperception is that hair loss comes as a result of other forms of therapy.

Since breast cancer has been correlated with high levels of estrogen in the blood, it is thought that decreasing estrogen levels can reduce the risk of recurring cancer. Thus hormonal therapy is sometimes given after chemotherapy to patients in cases where tumors have estrogen positive receptors, which are susceptible to hormone treatment. Tumor cells are not attacked by the body's immune system since they are recognized as the body's own cells and not as foreign antigens. Immunotherapy is a form of cancer treatment that causes the immune system to respond to the tumor cells as if they were bad cells, specifically targeting them for destruction. Indicators for all these kinds of therapy are always changing. It has not been established which treatment works best. Many times combinations of therapies are used.

Prognosis

A prognosis is a prediction that entails how a patient's disease may progress, their expected chance for recovery and projected life expectancy post-diagnosis. The prognosis is based on previous health records and response to treatments. The basis for a prognosis is formed from statistics gathered from case studies on a particular disease. The accuracy of a prognosis is based on the extent of research conducted and will also depend on the experience and expertise of the physician(s) rendering the prognosis. Prognosis for women with breast cancer generally declines as the stage progresses. The odds of a woman living beyond 10 years after being diagnosed with breast cancer are roughly 95%, 88%, 66%, 36%, and 7% for stages 0, I, II, III, and IV respectively. It is important to keep in mind, however, that a prognosis can never carry a predicted outcome for an individual with 100% certainty, but knowing a prognosis can help physician make informed decisions as to whether certain treatments are worthwhile, necessary, too dangerous, etc. The prognosis plays an important role in end-of-life decisions.

Misdiagnosis in Breast Cancer Screening

The importance of breast cancer screenings should not be underestimated. Increasing survival rates have been attributed to healthcare awareness campaigns and improving medical technologies, particularly in the U.S. Regular examinations are key to early detection; however, as with any examination, there exists the possibility of misdiagnosis due to error. The goal is to reduce error as much as possible.

There are two categories of scientific error, systematic error and statistical error. A systematic error is the difference between what has been computed, estimated or a measured state of something and its actual state. A systematic error is not random; it

arises from an unknown source and can be eliminated or resolved once the source is identified. Human error is an example of systematic error that can be a factor in misdiagnosis. Results can be mixed up or misinterpreted as the wrong type of breast cancer resulting in a patient receiving the wrong type of treatment.

A statistical error is also the difference between what has been computed, estimated or measured state of something and its actual state. However, statistical errors are due to fluctuations in the measurement apparatus that are not predictable. Statistical errors are further subdivided into type I errors and type II errors. A type I error is also known as a false positive. As illustrated in the confusion matrix below, a false positive occurs when a hypothesis states something is true when it is actually not true. An example of this is when a patient receives a false alarm and is told she has cancer when she in fact does not. Type II errors, also referred to as false negatives, occur when the status of something is reported as false when it is actually true. An example of this is when a malignant tumor is diagnosed as benign and the patient is told that she is clear of cancer when she actually is not. The false negative is indicated in the upper right corner of the confusion matrix below.

The performance of a diagnostic system is measured by its sensitivity and specificity. Sensitivity is the measure of how reliable a system is at making positive identifications, or, in other words, correctly identifying that which is inspected as being specifically that which is sought. A highly sensitive system will recognize what it is looking for most of the time, and rarely produce a false negative. Thus sensitivity is expressed as the ratio of number of true positives, TP, to the sum of true positives and false negatives as shown in the equation,

$$Sensitivity = \frac{TP}{TP + FN} \qquad \text{Eq. 1.1}$$

Specificity is a measure of how well a system can make a negative identification, or indicate when something inspected is not what is being sought, but something else. A classification system with high specificity will rarely make the mistake of identifying what is being inspected as what is being sought. Specificity is thus defined as the ratio of the number of true negatives, TN, to the sum of true negatives and false positives as in the equation,

$$Specificity = \frac{TN}{TN + FP} \qquad \text{Eq. 1.2}$$

Figure 2.3 displays a confusion matrix, which illustrates the four possible outcomes of an evaluation. The results of a test performed with perfect sensitivity and specificity will all be either TP or TN' and never FP or FN. In actuality, no inspection method or tool can evaluate with perfect accuracy. Typically an instrument is considered to have acceptable performance if its specificity and sensitivity are above 0.90.

In cases where cancers go undetected, or in other words, when a screening results in a false negative (FN), adequate treatment can be delayed, allowing cancers to advance to later stages where more drastic treatments may be required. More frequent examinations may reduce the chances of missing a malignancy, but there are issues with this as well. Apart from the financial aspect, (assuming cost is not a prohibitive factor, although it is in some cases) the possibility of having a false positive (FP) result increases with the number of screenings carried out.

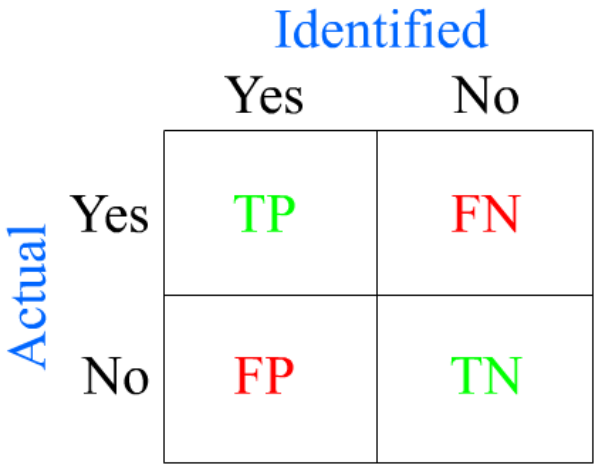


Figure 2.3: The confusion matrix. A "perfect" system detects with 100% sensitivity and 100% specificity. That is, it only makes true positives (TP's) and true negatives (TN's) in green and never false positives (TP's) and false negatives (TN's) in red.

A study conducted by the Cancer Registry of Norway reported on Norwegian women who started having mammograms bi-annually near age 50 and continued having mammograms for 20 years had approximately a 1 in 5 chance of getting a false positive result at some point during the 20 years.[5] Nevertheless, one may raise the question, "Are the risks of getting a misdiagnosis worth having regularly scheduled examinations?" The question is fair and important.

The pronouncement of cancer can be psychologically distressing. Perhaps even more traumatic is the experience of needlessly enduring breast cancer treatments, which are not completely harmless. Moreover, medical expenses can mount an unnecessary financial burden on the patient that they might not be able to recoup. Misdiagnosis also encumbers healthcare systems with malpractice lawsuits that adversely affect health care providers, their services, and ultimately the patients.

The Norwegian study found that although 1 out of 5 women have false positive exams, only 93.8% of their follow-up biopsies returned true positive results for cancer.

Thus, the study concluded that regular breast cancer screenings are advisable despite the odds of misdiagnosis. Nevertheless, the aim in breast cancer screening is to perform diagnoses keeping the odds of missed and false detections to a minimum. In Chapter 3 we discuss the role technology plays in mammography.

CHAPTER 3 – DIGITAL MAMMOGRAPHY

Film Mammography, Digital Mammography and Computer-Aided Systems

Mammography and the quality of overall healthcare can be improved by technology to a great extent. Film mammography and digital mammography are concurrent technologies used for detecting breast cancer. Film mammography also, known as traditional mammography, is the conventional method that stores x-ray images on film as described earlier. Digital mammography developed subsequently with advances in computer technology and optical electronics. Digital mammography involves producing X-ray images of the breast and storing them directly to a machine in electronic form. Each has its own advantages and disadvantages, which will be described in more detail later. However, one of the advantages in digital mammography is that mammogram images can be analyzed and manipulated using computer-aided diagnostics, or CAD. CAD systems aid radiologists in locating malignancies in digital mammograms. They were first put to use in the 1990's, and they continue to develop with increasing promises to benefit the field of clinical mammography.

Industrial research is primarily responsible for many of the advances seen in digital mammography. At present, the only CAD system approved by the Food and Drug Administration (FDA) for use in screening, diagnostic and digital mammography is ImageChecker® manufactured by R2 Technology Inc. in Los Altos, CA. ImageChecker® was the first CAD system approved by the FDA in 1998. Since then it has been installed in more than 1400 clinical sites worldwide. Two other CAD systems that have also been approved by the FDA are Second Look™ by CADx Medical Systems, Inc. in Northborough, MA and the MammoReader™ by Intelligent Systems Software, Inc.

Clearwater, FL. CADx Medical Systems merged with another company called Qualia Computing, Inc. in 2002, shortly after receiving FDA approval. However R2 Technology sued CADx that same year for infringing on three of its patents. CADx lost the lawsuit and is no longer in operation. It appears Intelligent System Software, Inc. was not successful in establishing a foothold in the industry after its MammoReader™ was accepted by the FDA in 2002. Thus R2 Technology Inc. remains standing as the sole industry force leading the development of technology in the area of digital mammography.

Academic research also plays a significant role in the development of digital mammography and CAD systems. A group in the radiology department at the University of Chicago was one of the first to bring computer-assisted diagnostics into the clinical arena. In January 2005, R2 Technology and the University of Chicago Medical Center announced their exclusive agreement to collaborate on the development of a mammographic CAD workstation reference library. The multi-faceted collaboration is looking into a number of different aspects involving mammography including: image processing, image storage, CAD mark interpretation and recall through PACS (Picture Archiving Interface) and DICOM (Digital Imaging and Communications in Medicine) connectivity. [6]

The eDiamond project is a joint effort headed by Oxford University and supported by IBM, Hurley and the UK government. The aim of the project is to use grid computing to facilitate information distribution among health centers in the UK. The hope is to relieve a burdened healthcare system experiencing a shortage of radiologists along with

an increasing patient population and improve the reliability of the diagnoses using the redundant systems engineering concept.[7]

The University of California at San Diego (UCSD) is one of 38 centers in the United States to hold a National Cancer Institute designation as a Comprehensive Cancer Center. The university has several well-established programs involved in cancer research and is expanding. As of the spring 2005, digital mammography machines have been installed at two UCSD locations, the Rebecca and John Moores Cancer Center and the UCSD's Breast Imaging Center in Hillcrest. The research there is promising to have far reaching impact on the San Diego community and beyond.[8]

Louisiana State University (LSU) serves as home base for UniPACS (Universal Picture Archiving and Communication Systems). UniPACS is a software company focused in medical imaging and staffed with LSU researchers headed by Dr. John M. Tyler—former CEO and president, now vice-president of the company and professor emeritus of the computer science department at LSU. The company offers a suite of applications that allow medical images to be archived and shared securely via the Internet. The company's products are used in several health care centers in the southeast and New England area. They save cost, eliminating the need for expensive equipment previously required by radiologists for viewing the images.

Sample and Tyler CAD System

Although a CAD package from UniPACS is not yet available for commercial use, Professor Tyler has performed extensive academic research in digital medical image processing. He has developed a CAD system with one of his former doctoral students, Dr. John Sample.

Sample's dissertation, entitled "Computer Assisted Screening of Digital Mammogram Images", sought to address the needs of CAD systems designed for digital mammography. In the thesis, Sample describes a computer-aided detection system comprised of algorithms he designed with Tyler to identify and outline suspected cancerous masses in digital mammograms. The system is self-testing and also performs comparative analyses of its results.[9]

The Sample-Tyler CAD system displays excellent performance in identifying malignant tumors. Their algorithm detects masses with a sensitivity ranging from 92.8% to 100%. However, Sample reports in his thesis that it also misclassifies an unacceptable proportion of non-cancerous masses as malignant. Thus the specificity of the algorithm is undesirably low. In the closing statements of his thesis he proposes developing a post-processing algorithm to help further discriminate between malignant masses and benign masses.

Shape Analysis System

Tyler suggests that tumor shape can be used as a discriminating factor. The shape analysis system presented here is an algorithm designed to geometrically classify the shapes of tumors identified in mammograms by the Sample-Tyler CAD system. This shape classification scheme involves approximating tumor shapes with ellipses and then categorizing the shapes of the ellipses based on their parameters: size, eccentricity, and goodness of fit. The elliptical parameters are then compared with diagnostic data for each tumor to determine correlations between tumor shape in order to identify and discard false positive verdicts rendered by the Sample-Tyler CAD system. The Sample-Tyler CAD system and shape analysis system will be discussed in Chapter 4 in greater detail.

CHAPTER 4 – TUMOR SHAPE ANALYSIS

Introduction

The purpose of a computer-aided diagnostic (CAD) system in digital mammography is to aid radiologists in identifying possibly malignant tumors in digital mammograms using computer algorithms. As stated in Chapter 3, all such systems are subject to statistical error. In this work we aim to improve the performance of a CAD system developed by Tyler and Sample by increasing its specificity using tumor shape analysis and categorization.

Tumor Shape Analysis Premise and Methodology

The approach to tumor shape analysis presented here involves approximating the shapes of tumors by fitting them with ellipses and then categorizing the tumors based on the dimensions of the ellipses and the closeness of the approximation. We propose that tumors warranting investigation are localized, globular masses (as opposed to say stringy objects or objects with highly irregular, concave, or angular features). Thus, the shapes of these masses projected onto two-dimensional images would appear round or near circular when segmented apart from the rest of the image.

The Sample-Tyler algorithm performs graphical segmentation of the digital mammogram images. The images of the tumors it segments appear in a variety of shapes. Referring to the assumption that tumors suggested for biopsy tend to be round to some degree, and have edges that form closed curves, the tumor shape algorithm is intended to be used to define a criterion by which false positives generated by the Sample-Tyler CAD system can be identified based on shape. We refer to the tumor shape-fitting algorithm as the Ellipse-Closed Curve Fitting (ECCF) system. The ECCF is comprised of an algorithm

that performs edge detection on the overlay images produced by the Sample-Tyler CAD system and then attempts to fit an elliptical curve to the shape of the tumor in a way that is similar to regression techniques used in statistics where a simple equation is parameterized to fit a set of real data points. In fact, we adapt a regression analysis technique (R-square) to measure the closeness of the fit as well as other structural quantities. The output of the ECCF is in the form of text and graphic images that show processing of the images step-by-step. The tumor shape algorithm can then be used post-operatively to the Sample-Tyler CAD system to analyze tumors based on shape and then reject shapes that do not match the derived criteria. The aim is to increase the specificity of the overall detection process.

Elliptical Quantities

Since the structural analysis of the ECCF system is based on elliptical quantities, we briefly discuss the two elliptical quantities we consider, which are area and eccentricity. Any 2D ellipse can be completely described given its area and the ratio between its major and minor axes. The area of an ellipse is given as π times the product of its semi-major axis and its semi minor axis as shown in equation 4.1, where a and b are the major axis and minor axis respectively.

$$A = \pi ab \qquad \text{Eq. 4.1}$$

The eccentricity is a measure of how round or needle-shaped an ellipse is. Here we denote eccentricity as e and show how it is defined in terms of the semi-major and minor axis.

$$e = \left(1 - \frac{a^2}{b^2}\right)^{\frac{1}{2}} \qquad \text{Eq. 4.2}$$

For a circle, the major and minor axes are equal lengths. If the major and minor axes are the same length, then the quotient $\frac{a^2}{b^2}$ is 1, subtracted from 1 gives zero. Thus, a circle has zero eccentricity. As an ellipse becomes more needle-shaped the minor axes length of an ellipse a approaches zero and the eccentricity e approaches 1. By approximating the shapes of tumors with ellipses, the ECCF system can then use the area and eccentricity of the ellipses to quantify tumor shape.

MATLAB Implementation

The shape analysis algorithm was implemented in MATLAB. MATLAB has a number of capabilities that make it suitable for digital image processing. The MATLAB application operates on a variety of platforms: DEC Alpha, HP 9000, IBM RS/6000, PC & MAC, Open VMS, SGI (Silicon Graphics), and SUN Sparc. All diagnostics presented in this thesis research were implemented in MATLAB and performed on a PC equipped with a 1.6 GHz- Intel Pentium processor and 512 MB of RAM. MATLAB's built-in I/O functions support a number of video, text, audio and image formats including JPEG, PNG and TIFF. The medical images are in a format using Lossless-JPEG compression or LJPEG. The LJPEG format is not supported by MATLAB; however, the images are converted from LJPEG format to TIFF format as described in Appendix A. The TIFF images are then read into MATLAB. In MATLAB, images are represented as two-dimensional matrices allowing low-level image analysis and manipulation using MATLAB's efficient matrix and vector computations. MATLAB also has graphics and visualization capabilities that make useful for rendering scientific and engineering

graphics quickly and easily. Lastly, MATLAB is programmable; it has a mathematical scripting language, which we have used to implement an automated shape analysis system that takes advantage of the functions and features just described.

Overview Sample Tyler CAD System

The Sample-Tyler CAD system is designed to detect regions in digital mammograms that may possibly be tumors. The digital mammograms for this work are taken from actual patient cases stored in the DDSM (Digital Database for Screening Mammography (DDSM), at the University of South Florida) (http://marathon.csee.usf.edu/Mammography/Database.html). The digital mammograms have already been classified as normal, cancerous or benign. Sample and Tyler used the database as a test bed for their CAD system. The system is quite accurate. It performs with sensitivity above 0.928 (or 92.8%). However, it returns an average of 36.6 FP's per image, which is undesirable. So the goal is to develop an algorithm that reduces the number of false positives it detects to increase the CAD systems specificity.

Figure 4.1 shows a flow diagram of the Sample-Tyler CAD system. Input patient cases are represented on the left by the numbered by panes. Each pane contains a set of 4 digital mammogram images belonging to one patient. There are two views for each breast: left medial lateral oblique (L-MLO), right medial lateral oblique (R-MLO), left cranial caudal (L-CC), right cranial caudal (R-CC). The Sample-Tyler CAD system, represented at the center of the diagram, processes the digital mammograms and generates output shown in the green and yellow areas on the right:

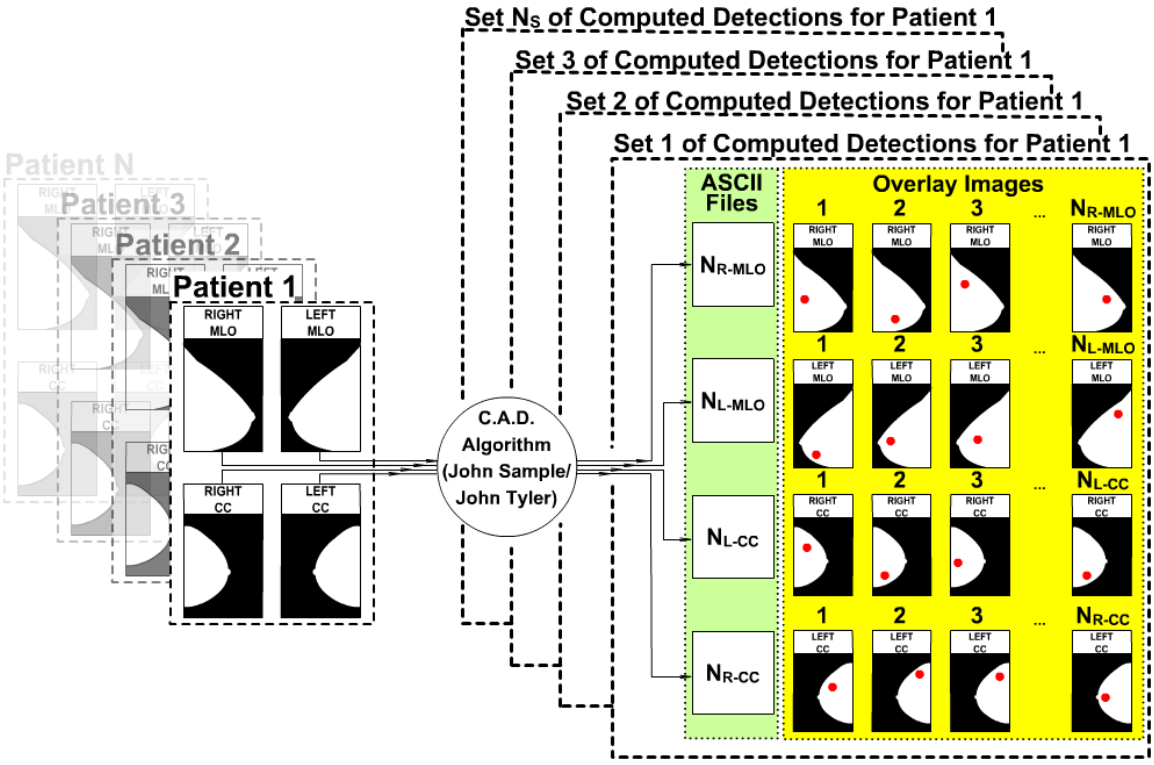


Figure 4.1: Schematic of the Sample-Tyler CAD system (center). The algorithm takes digital mammograms from patient cases as input, displayed as numbered panes on the left and produces images containing areas where suspected cancerous tumors have been marked.

1. Four text files (represented in the green area of the diagram). Each file contains one integer indicating the number of suspected areas in the breast detected by the algorithm for each mammographic view. $N_{L\text{-}MLO}$, $N_{R\text{-}MLO}$, $N_{L\text{-}CC}$, and $N_{R\text{-}CC}$ represent the integers in each text file corresponding to left medial lateral oblique, right medial lateral oblique, left cranial caudal, right cranial caudal view respectively.

2. Four sets of images (represented in the yellow area of the diagram). Each set of images corresponds to a different mammographic view: left

MLO, right MLO, left CC, right CC. The number of images in each set depends on the number of cancers detected for each mammography view. Each image contains a single detection indicated here by the red dot.

To conserve disk space, the images are reduced in color depth to 1-bit (stored in monochrome black and white) as well as reduced in width and height (maintaining the aspect ratio of the original digital mammogram image it was taken from).

Overview of Ellipse-Closed Curve Fitting (Extending the Sample-Tyler CAD system)

The shapes of the tumors detected by the Sample-Tyler CAD system have various multi-diametered shapes the outlines of which are closed curves. As stated earlier in the premise, we suspect true positive tumors are more likely to have round or elliptical shapes. The ellipse-closed curve fitting (ECCF) system was developed to serve as a post-screening component to perform shape analysis on the overlay images generated by the Sample-Tyler CAD system. The ECCF system is comprised of several sub-components, represented in the blue rectangle in Figure 4.2. The subcomponents are organized in the diagram in three divisions: pre-processing, curve-fitting, and fit evaluation. Overlay images from Sample-Tyler system, framed in yellow borders, enter the ECCF system as indicated by the red arrow. Each sub-component performs a particular operation on the overlay images. The green arrows point to the output of the ECCF system. The item in the top right corner represents an ASCII file generated by the ECCF system, which contains several columns of data calculated by the ECCF system- including the ratio between the areas of the tumor and the ellipse that has been fitted to it, the eccentricity of the fitted ellipse and the R-square of the regression. Each of these subcomponents will be discussed in more detail in the following sections.

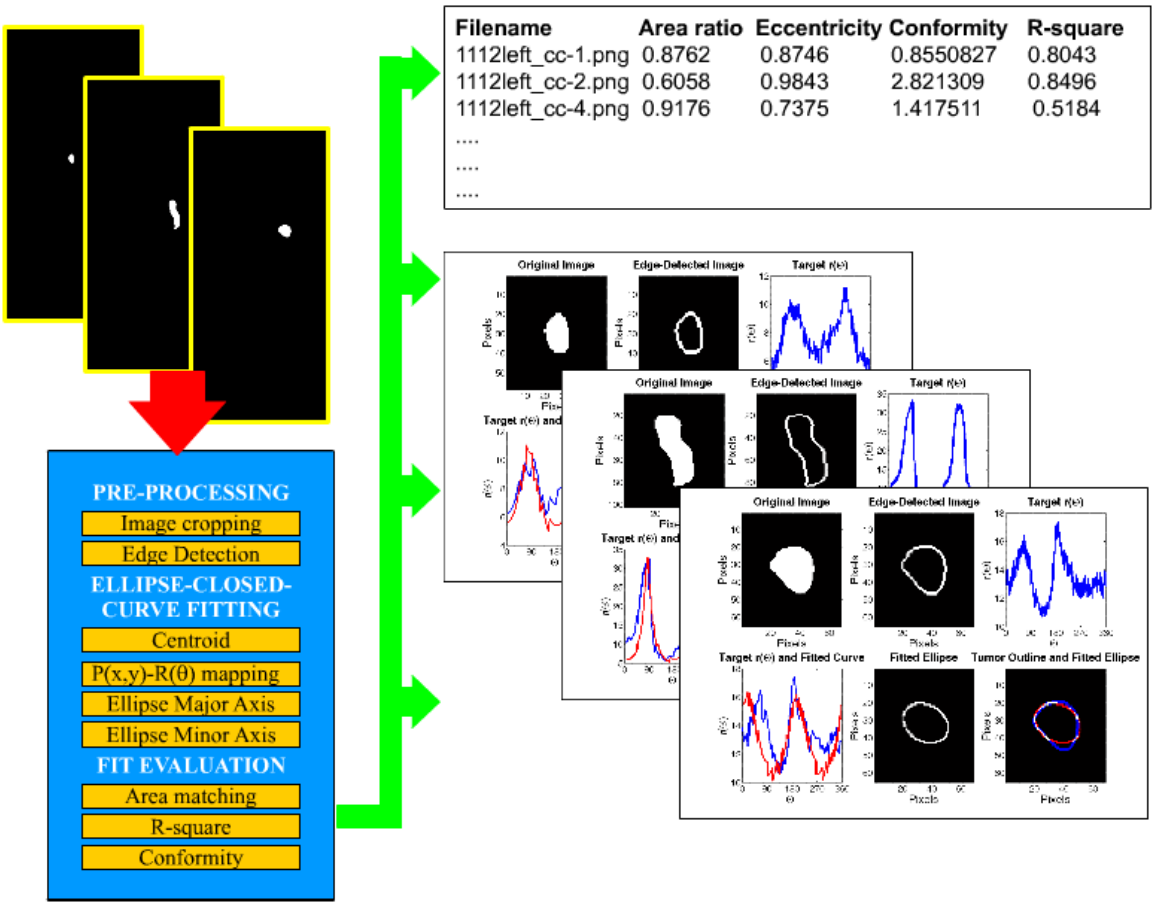


Figure 4.2: Schematic of ECCF system (Ellipse-Closed-Curve Fitting) system. The ECCF computes in three phases pre processing, where the images are cropped and edge-detected, curve-fitting, where the shapes of the tumor are fitted by ellipses and fit-evaluation, where that goodness of the fit is calculated and performance information is output.

Pre-formatting

Image cropping

The overlay images are initially generated with the same size and aspect ratio (length-to-width) of the digital mammogram it is created from. These images are relatively large, ~4000 × ~4000 pixels. One image alone can have a file size of 30-40 MB. Thus the images are cropped for the sake of efficiency. Figure 4.3 shows an actual left mediolateral mammogram, outlined by a blue border, where the tumors detected by the Sample-Tyler have been marked in red. The dotted white arrows show how each

tumor detected appears in a separate overlay file, framed in yellow. The areas marking the tumors are typically a small fraction of the overlay image area ~3% of the entire image. The images are cropped relieving all but a 20-pixel buffer on each side of the tumors. The cropped images, framed in green, can be processed with much greater efficiency as the search space has been reduced.

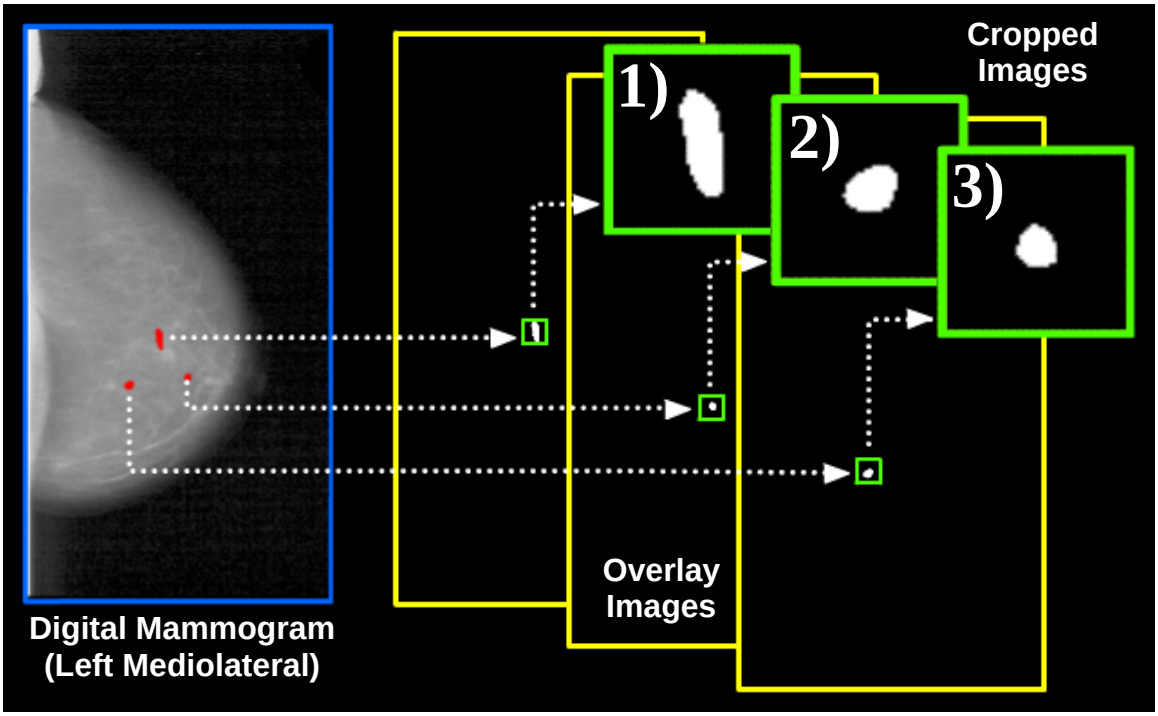


Figure 4.3: Digital mammogram of the left breast mediolateral view is shown outlined by blue rectangle. Detected tumors indicated in red are shown in overlay images outlined by yellow rectangles. Cropped images are shown in green frames. The tumors are numbered 1, 2, and 3 for future reference.

Edge detection

Only pixels defining the shape of the tumor are of interest. The algorithms used to approximate tumor shape operate only the pixels lying on the edge of the tumors. All pixels inside the edge boundary must be removed. Thus, after cropping, the next operation performed is edge detection.

The edges in an image can be found by computing the Laplacian on that image. An image can be represented as a two dimensional surface $I(x,y)$, where $I(x,y)$ represents pixel intensity and (x,y) are pixel coordinates. Edges in the image exist in places where the intensity of the pixels is varying. More precisely, edges exist where the partial derivate $\frac{\partial I(x,y)}{\partial x}$ or $\frac{\partial I(x,y)}{\partial y}$ is not equal to zero. Thus edges detected can be detected by computing the derivative of $I(x,y)$ with respect to x and y or the gradient of $I(x,y), \nabla I(x,y) = \frac{\partial I(x,y)}{\partial x} + \frac{\partial I(x,y)}{\partial y}$. The gradient will have negative or positive values at edges, but the absolute value of the gradient is the only thing of interest. We take the dot product of $\nabla I(x,y)$ with itself. Thus the edge-detected images are essentially the Laplacian of the original image $I(x,y)$, $\nabla^2 I(x,y)$.

The creation of elementary and specialized arrays and matrices as well as basic array operators and operations are provided in MATLAB. Since digital images are represented in MATLAB as two-dimensional matrices, basic linear algebra operations and manipulations can be performed on them including: addition, subtraction, multiplication, transposing, left/right divide and others. We construct a gradient operator from these simple linear algebra operations in MATLAB to perform edge detection on images. We compute the Laplacian using discrete derivatives.

$$\nabla^2 I = \nabla I \cdot \nabla I \approx \Delta I_e \cdot \Delta I_e \qquad \text{Eq. 4.3}$$

where,

$$\Delta I_e = \Delta I_u + \Delta I_d + \Delta I_r + \Delta I_l \quad \text{Eq. 4.4}$$

$$\Delta I_u = I - S_m I \quad \text{Eq. 4.5}$$

$$\Delta I_d = I - S_m^T I \quad \text{Eq. 4.6}$$

$$\Delta I_r = I - I S_n \quad \text{Eq. 4.7}$$

$$\Delta I_l = I - I S_n^T \quad \text{Eq. 4.8}$$

ΔI_u, ΔI_d, ΔI_r, and ΔI_l, are matrices containing the values of the discrete differences between adjacent pixels in the "up", "down", "left" and "right" directions respectively. Superscript, T, indicates a transposed matrix. The variables m and n are the dimensions of the image $I(x, y)$. S_m and S_n are circular permutation matrices described in more detail in Appendix C. Figure 4.4 shows an example of the three images of the cropped tumors, from Figure 4.3, in relation to the edge detected images.

Ellipse Closed Curve Fitting

The next section entails descriptions of the ECCF subcomponent, which actually fits ellipses to the closed-curve outlines of the tumors. The basic objective here is to calculate parameters for ellipses that match the location, shape, and size of the tumors as closely as possible. First, since the tumor shape is actually a constellation of points, we must describe the location of the tumor in a way that is non-arbitrary. We require that the center of the ellipse be fitted to the tumor and the center of the tumor itself must coincide. So we have to find the center of the tumor. We do that by calculating its centroid.

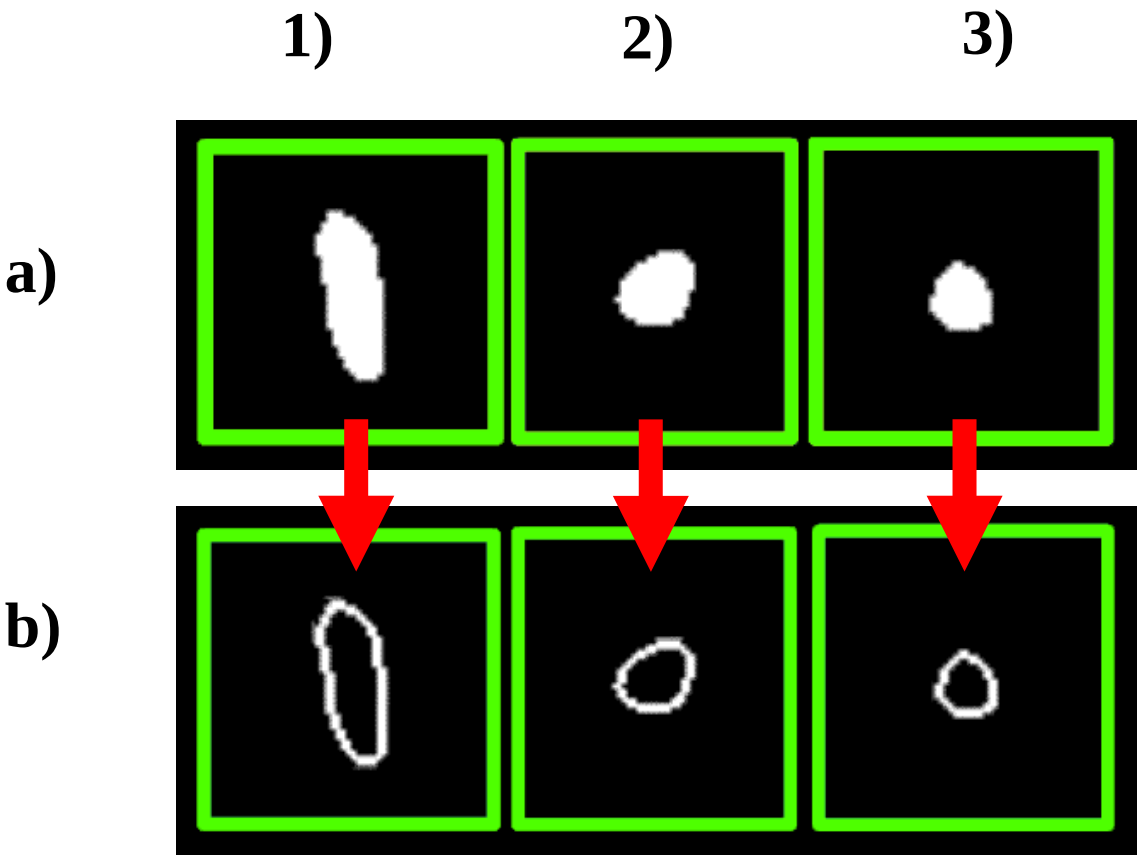


Figure 4.4: Example of tumors detected by the Sample-Tyler CAD are shown in a). Red arrows relate corresponding edge-detected images in b) as rendered edged-detecting component of the shape analysis algorithm.

Centroid calculation

The Cartesian coordinates of a centroid are the means of the coordinates of the set of vertices. Figure 4.5 contains a simple illustration of the centroid corresponding to a set of three points P_i, where $i = 1,2,3$.

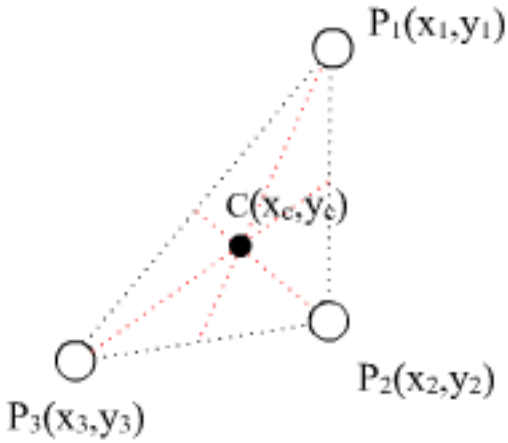


Figure 4.5: Calculating the centriod, C (x_c,y_c), for three points $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, and $P_3(x_3, y_3)$

If the three vertices are located at $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, and $P_3(x_3, y_3)$ then the centroid is a point $C(x_c, y_c)$ at which,

$$x_c = \frac{x_1 + x_2 + x_3}{3} \qquad \text{Eq. 4.9}$$

$$y_c = \frac{y_1 + y_2 + y_3}{3} \qquad \text{Eq. 4.10}$$

In general, for any number of points N, the centroid is calculated as shown in the following equation:

$$x_c = \bar{x} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i} \qquad \text{Eq. 4.11}$$

$$y_c = \bar{y} = \frac{\sum_{i=1}^{N} m_i y_i}{\sum_{i=1}^{N} m_i} \qquad \text{Eq. 4.12}$$

, where m_i is a weighting factor generalizing the expression to apply to non-uniform systems. In this case all pixels are identical, $m_i = m \neq 0$. Thus, all the weights factor out of the expression.

The vertices $P_i(x_i, y_i)$ are the coordinates of the pixels forming the shape of tumor. $C(x_c, y_c)$ represents the location of the tumor and serves as a unique and non-arbitrary point of origin for constructing the fitted ellipse.

Mapping Cartesian coordinates to polar coordinates

With the centroid serving as a reference position for the tumor, we now seek to match the size, shape, and orientation of the tumor shape with an ellipse as closely as possible. Just as in linear regression, where the objective is to find the linear equation that comes closest to fitting a collection of data points, we wish to find the equation for an ellipse that comes closest to fitting a constellation of points lying on a closed curve.

To begin, our method requires a special mapping of the points on a Cartesian plane to a polar coordinate plane. The mapping is a bijection. All points $P_i(x_i, y_i)$ get mapped to exactly one point $\vec{R}_i(\theta_i)$

$$P_i(x_i, y_i) \rightarrow \vec{R}_i(\theta_i) \qquad \text{Eq. 4.13}$$

Figure 4.6a illustrates points $P_i(x_i, y_i)$ on a Cartesian plane as well as their centroid. The function $\vec{R}_i(\theta_i)$ is obtained by calculating the vector that extends from the centroid $C(x_c, y_c)$ to each point $P_i(x_i, y_i)$. The corresponding angles, θ_i are computed as,

$$\theta_i = \arccos\left(\frac{\vec{R}_i \cdot \hat{x}}{\left|\vec{R}_i \cdot \hat{x}\right|}\right), \qquad \text{Eq. 4.14}$$

where $\hat{x}$ is the unit vector appearing in Figure 4.6a. The arccosine is one-to-one, but not onto. Two points bisected by the horizontal axis $\hat{x}$ and equidistant from the vertical axis $\hat{y}$ would have the same angle, even though they are in different half planes. The problem

is circumvented by introducing an asymmetry where the points, $\vec{R}_i(\theta_i)$, in the lower half plane are phase shifted by 180˚. After that, $\vec{R}_i(\theta_i)$ is sorted to obtain a function that describes tumor shape as a radial function of theta exampled in Figure 4.6b.

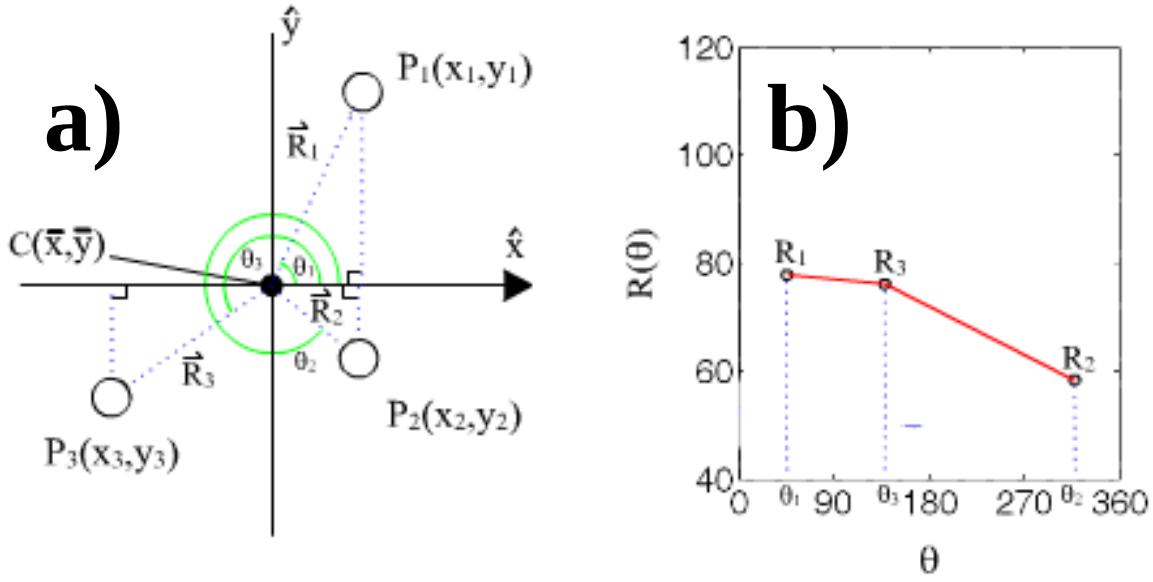


Figure 4.6: Cartesian-Polar coordinate mapping. Three points, $P_1(x_{1i}, y_1)$, $P_{2i}(x_{2i}, y_2)$, and $P_{3i}(x_{3i}, y_{3i})$ in a) are expressed as a radial distribution in polar coordinates in b).

Both representations $P_i(x_i, y_i)$ and $\vec{R}_i(\theta_i)$ are used to obtain the metrics calculated to fits ellipses to tumor shape.

As stated in the previous section, an ellipse can be completely described in terms of its major axis, minor axis, orientation and position. We have to some how extract these quantities from our point constellation.

First, a major axis and orientation for the ellipse is chosen large enough to extend over the entire tumor shape, from end-to-end. This is accomplished by calculating a vector $\vec{v}_{major}$ between points on the tumor that are farthest apart as expressed in the equation,

$$\left|\vec{v}_{major}\right| = \max\left(\left|P_i\left(x_i, y_i\right) - P_j\left(x_j, y_j\right)\right|\right) \qquad \text{Eq. 4.15}$$

The norm of $\vec{v}_{major}$ serves as the length of the major axis of the fitted ellipse. The angle $\vec{v}_{major}$ makes with the horizontal axis $\hat{x}$, is used to compute θ_υ, which is used for the orientation of the ellipse as:

$$\theta_\upsilon = \arccos\left(\frac{\vec{v}_{major} \cdot \hat{x}}{\left|\vec{v}_{major} \cdot \hat{x}\right|}\right) \qquad \text{Eq. 4.16}$$

We compute the semi-minor axis for the fitted ellipse by finding $\min\left|\vec{R}_i\left(\theta_i\right)\right|$ as shown in Figure 4.7 b. In reference to the tumor shape, the point $\min\left|\vec{R}_i\left(\theta_i\right)\right|$ is the point on the edge of the tumor that is closest to the centroid.

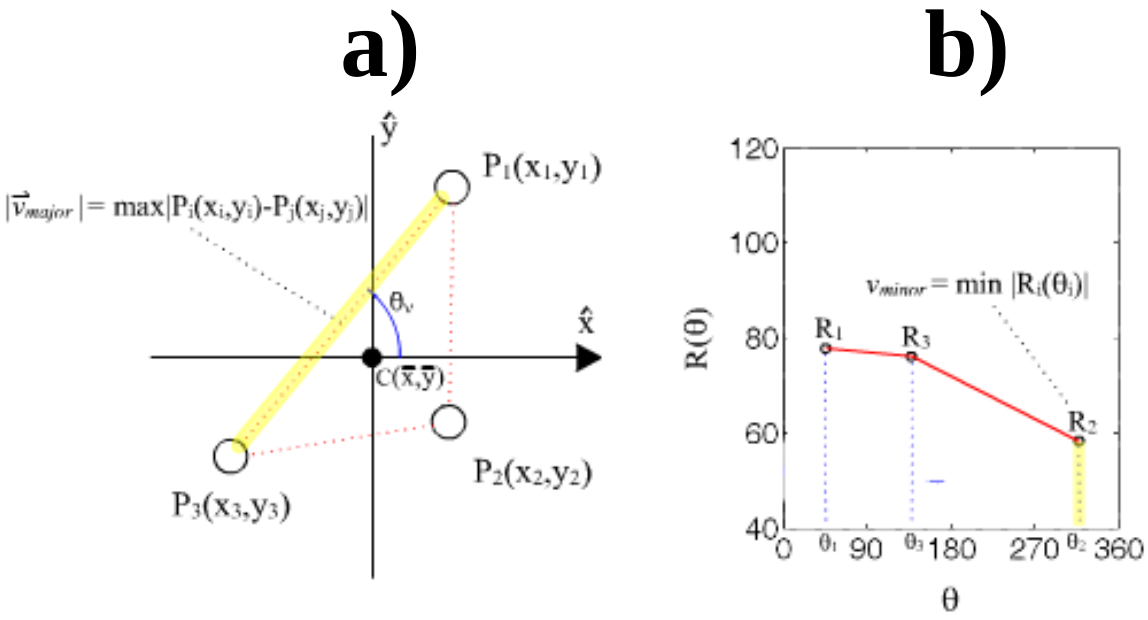


Figure 4.7: a) The major axis of the fitted ellipse, $\vec{v}_{major}$, outlined in yellow, is computed by finding the longest distance between two points $P_i\left(x_i, y_i\right)$ and $P_j\left(x_j, y_j\right)$. b) The minor axis of the fitted ellipse, $\vec{v}_{\min or}$, also outlined in yellow, is computed as the minimum of the function $\vec{R}_i\left(\theta_i\right)$.

Constructing the fitting ellipse

With the major axis, the semi-minor axis, the position and orientation for the fitted ellipse obtained; the images of the fitted ellipses are generated using:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = [\Re] \cdot \vec{\upsilon} + \vec{C} = \begin{bmatrix} \cos\theta_\upsilon & \sin\theta_\upsilon \\ -\sin\theta_\upsilon & \cos\theta_\upsilon \end{bmatrix} \begin{bmatrix} \upsilon_{major} \cdot \cos(i) \\ \upsilon_{\min or} \cdot \sin(i) \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix} \qquad \text{Eq. 4.17}$$

The ellipses are constructed as a set of points $P_i'(x_i, y_i)$ from the vector $\vec{\upsilon}$. The variable i is a real number varying from 0˚ to 360˚ in adjustable discrete steps. The vector $\vec{\upsilon}$ consists of the parametric equations for which the major axis is chosen to extend across the horizontal axis of the image plane and the minor axis to align with the vertical axis of the image plane. The orientation of the ellipse is set by rotating the points $P_i'(x_i, y_i)$ in the plane about the center of the ellipse by θ_υ as specified in the matrix $\Re$. The fitting ellipse is the shifted by the vector $\vec{C}(x_c, y_c)$ so that its center coincides with the centroid of the tumor.

Fit Evaluation

The last subcomponent of the ECCF system measures how closely a fitting ellipse matches the shape of a tumor. We perform three calculations to assess how the ECCF system has performed: area match ratio, $\mathbf{R}^2$ or the coefficient of determination (adapted from statistics), and a shape conformity value. These quantities will be discussed in more detail in the following sections.

Area matching

Area matching is one of comparing the sizes of the tumor to the ellipse that has been fitted to it. The area of the ellipse is computed directly from the equation:

$$A_{fittedellipse} = \frac{\pi ba}{2} \qquad \text{Eq. 4.18}$$

The area of the tumor is computed by counting the number of pixels in the image. It is simply the number of white pixels in the original overlay image.

$$A_{tumor} = \sum_{i}^{m \cdot n} I(x, y) \qquad \text{Eq. 4.19}$$

where $I(x, y) = 1$ if the pixel is white and $I(x, y) = 0$ if the pixel is black. When the ratio between $A_{\text{fitted ellipse}}$ and A_{tumor} is close to 1, it is an indicator that the tumors are close in area or size. The area match ratio is a rough indicator of the closeness of the fit.

R-square (coefficient of determination)

In statistics, the coefficient of determination $\mathbf{R}^2$ is the proportion of a sample variance of a response variable that is "explained" by the predictor (explanatory) variables when a linear regression is done. It is used as a quantitative measure of the "goodness of fit" of linear data. $\mathbf{R}^2$ is a descriptive measure between 0 and 1. The closer it is to one, the better the fit. The ECCF technique involves fitting the parametric and canonical equations of ellipses to real data points. We adapt $\mathbf{R}^2$ from statistics and use it as a means of more precisely measuring how well fitted ellipses have matched the shape of the tumor. We use quantity $\mathbf{R}^2$ as defined in Equation 4.20,

$$R^2(X,Y) = \frac{\operatorname{cov}(X,Y)}{std(X)std(Y)} = \frac{\sum_{i}^{N}\left(X_i - \frac{1}{N}\left(\sum_{j}^{N} X_j\right)\right)\left(Y_i - \frac{1}{N}\left(\sum_{j}^{N} Y_j\right)\right)}{\sqrt{\sum_{i}^{N}\left(X_i - \frac{1}{N}\left(\sum_{j}^{N} X_j\right)\right)^2}\sqrt{\sum_{i}^{N}\left(Y_i - \frac{1}{N}\left(\sum_{j}^{N} Y_j\right)\right)^2}} \qquad \text{Eq. 4.20}$$

, where the terms X_i are the target values set and Y_i fitted. We adapt the quantity R-square for our purposes as shown in Equation 4.21.

$$R^2(X,Y) = \frac{\operatorname{cov}(X_{\theta_i}, Y(\theta_i))}{std(X_{\theta_i})std(Y(\theta_i))} = \frac{\sum_i^N \left(X_{\theta_i} - \frac{1}{N}\left(\sum_j^N X_{\theta_{ji}} \right) \right)\left(Y(\theta_i) - \frac{1}{N}\left(\sum_j^N Y(\theta_j) \right) \right)}{\sqrt{\sum_i^N \left(X_{\theta_i} - \frac{1}{N}\left(\sum_j^N X_{\theta_j} \right) \right)^2} \sqrt{\sum_i^N \left(Y(\theta_i) - \frac{1}{N}\left(\sum_j^N Y(\theta_j) \right) \right)^2}} \quad \text{Eq. 4.21}$$

here X_{θ_i} and $Y(\theta_i)$ are θ_i is the dependent variable for corresponding profiles of the target and the fitted ellipse respectively. The independent variable $Y(\theta_i)$ is taken from a continuous function, the radial coordinate of the fitted ellipse given by,

$$Y(\theta_i) = \sqrt{x'^2 + y'^2} \quad \text{Eq. 4.22}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{Eq. 4.23}$$

x and y are obtained from the parametric equation for and ellipse

$$\begin{aligned} x &= a\cos(\theta_i) \\ y &= a\sin(\theta_i) \end{aligned} \quad \text{Eq. 4.24}$$

By using $\mathbf{R}^2$ we obtain how the approximating function $Y(\theta_i)$ varies with our target function X_{θ_i}.

Ellipse shape conformity value

The ellipse shape conformity is another value we calculate to assess the "goodness of fit." It is a unit of pixel length, in this case pixel length, and is calculated as

the average of deviations of the points of the fitted ellipse $P_i'(x_i, y_i)$ from the tumor shape $P_i(x_i, y_i)$ as shown in Equation 4.23 and illustrated on Figure 4.8.

$$\frac{\sum_{i}^{N} \min\left|P_i(x_i, y_i) - P_i'(x_i, y_i)\right|}{N} \qquad \text{Eq. 4.25}$$

Figure 4.8 shows a schematic of how the ellipse shape conformity is calculated. The fitted ellipse is shown in red. The shape of the tumor is represented by the point set P_i connected by blue lines. The center of the ellipse and centroid of the tumor are positioned to coincide. The smaller the ellipse shape conformity value, the closer the fitted ellipse is to the shape of the target. An ellipse shape conformity value equal to zero means the target has an identical shape as the ellipse. The higher the ellipse shape conformity value, the more dissimilar the target shape is from the fitted ellipse. The shape conformity coefficient calculated for two identical shapes would be zero.

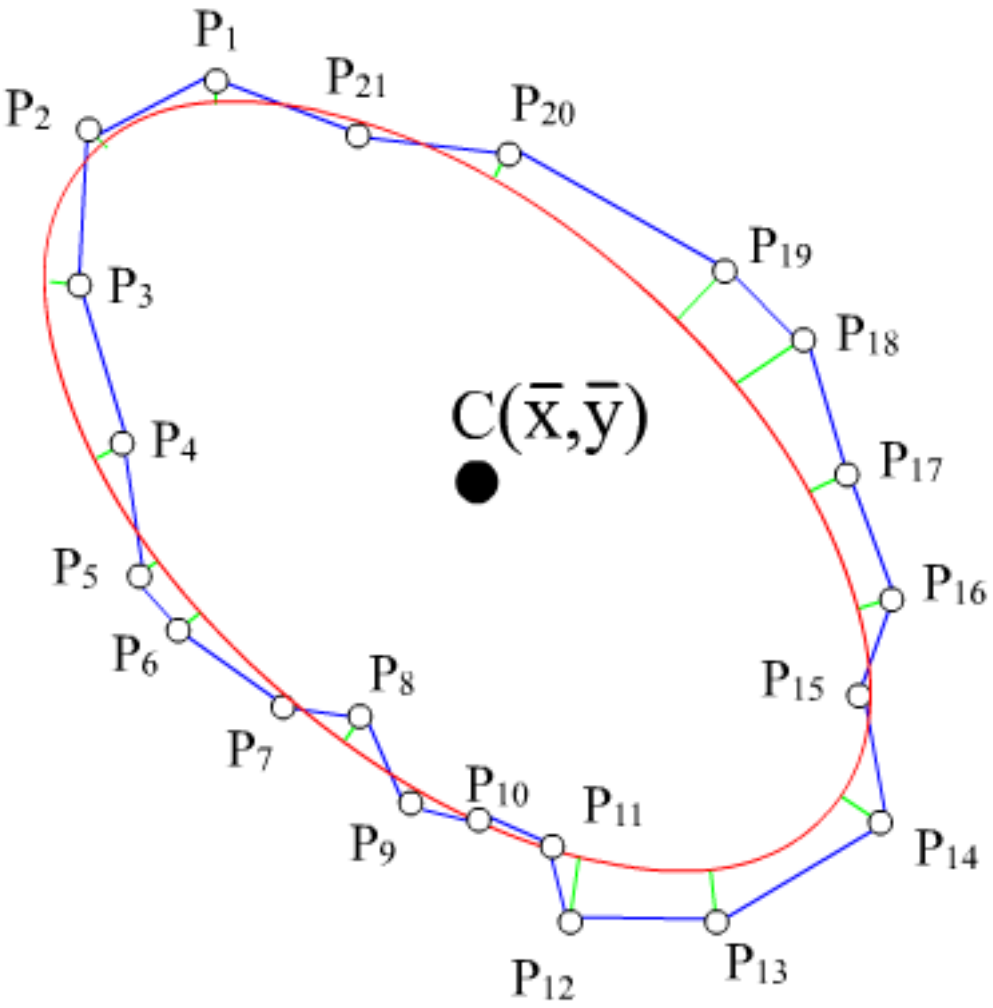
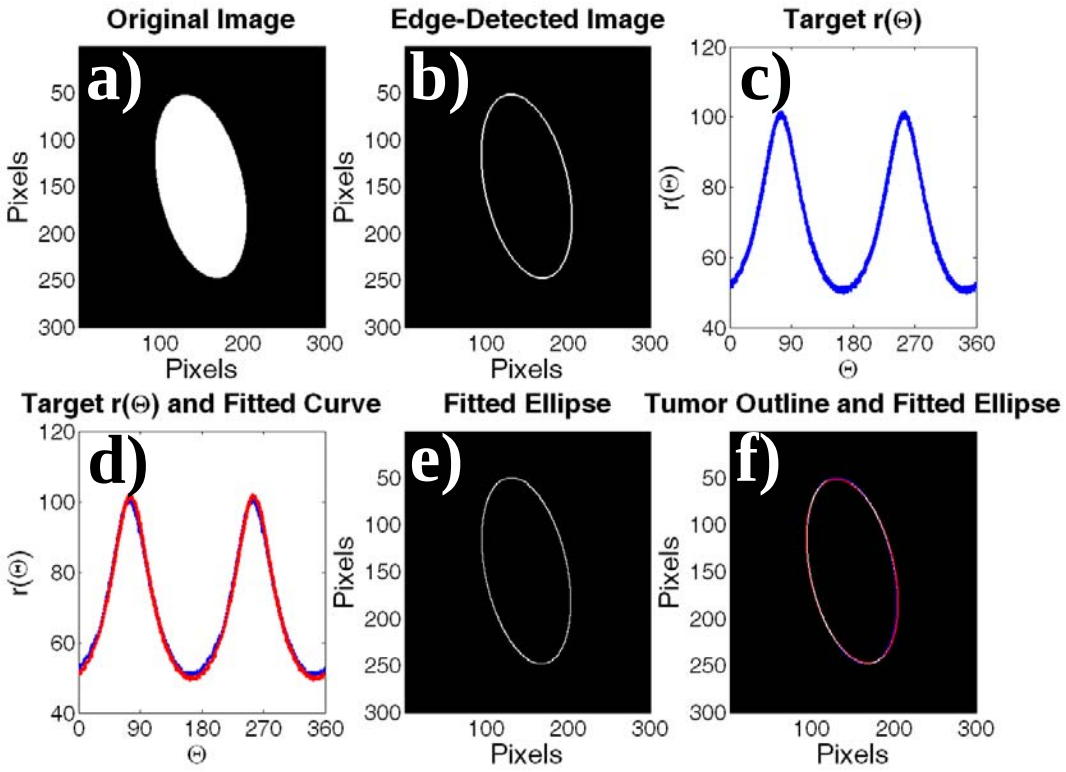


Figure 4.8: The ellipse shape conformity value is calculated by comparing the shapes of the tumor and the ellipse that has been fitted to it by the ECCF system. The fitted ellipse is represented in red. The set of points connected by blue lines represent the edge of the tumor. The green lines represent the amount of error between the two shapes.

CHAPTER 5 – RESULTS AND CONCLUSION

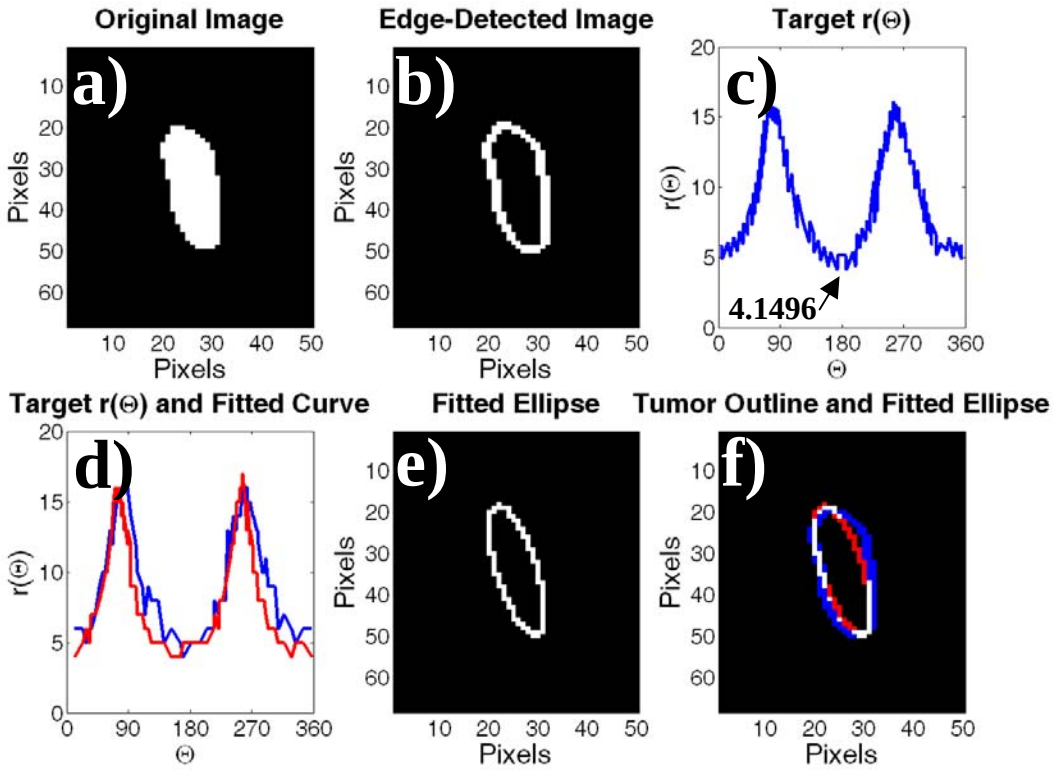
The ECCF system was tested on a battery of test images and then actual tumors generated by the Sample-Tyler CAD system. The battery of test images consisted of ellipses of various sizes, eccentricities and orientations generated by a testing program and is organized into four sets in Appendix B. Set 1 contains images of an ellipse with a semi-major axis of 100 pixels, a semi-minor axis of 50 pixels, rotated through 180° in steps of 15°. Set 2 contains test images with vertical ellipses at various eccentricities where the semi-major axis length is maintained at 100 pixels. Set 3 contains test images with horizontal ellipses at various eccentricities where the semi-major axis length is also maintained at 100 pixels. Tests on ellipses varying in size, with equal major and minor axes, appear in Set 4. An example of one test result for an ellipse with a semi-major axis of 100 pixels, semi-minor axis of 50 pixels, orientated at 105° off the horizontal axis appears in Figure 5.1. The original test image generated initially by the graphics routine is shown in part A. Part B shows the edge-detected test image. In part C, the radial function $r(\theta)$ of the ellipse with respect to its center is plotted in blue. A fitted ellipse is drawn from the major and minor axes computed by the component algorithms of the ECCF system. The radial functions, $r(\theta)$, of the original ellipse and its fitted ellipse are plotted together for comparison in blue and red respectively in part D. Part E shows the actual 2D image of the fitted ellipse alone. Part F shows the fitted ellipse (red) plotted over the original ellipse (blue) for comparison. The white pixels indicate where the points of the fitted and original ellipse match exactly.



Calculated quantities - ellipse_angle_at_15.tif	
Semi-major axis (pxw)	101.5012
Semi-minor axis (pxw)	49.8448
Angle	103.9678
Area of Fitted Ellipse	1.589427e+004
Area of tumor (px)	16014.0000
Ratio of difference in area	0.9925
Eccentricity	0.8711
Shape conformity (pxw)	9.596332e-001
R-squared	0.9985

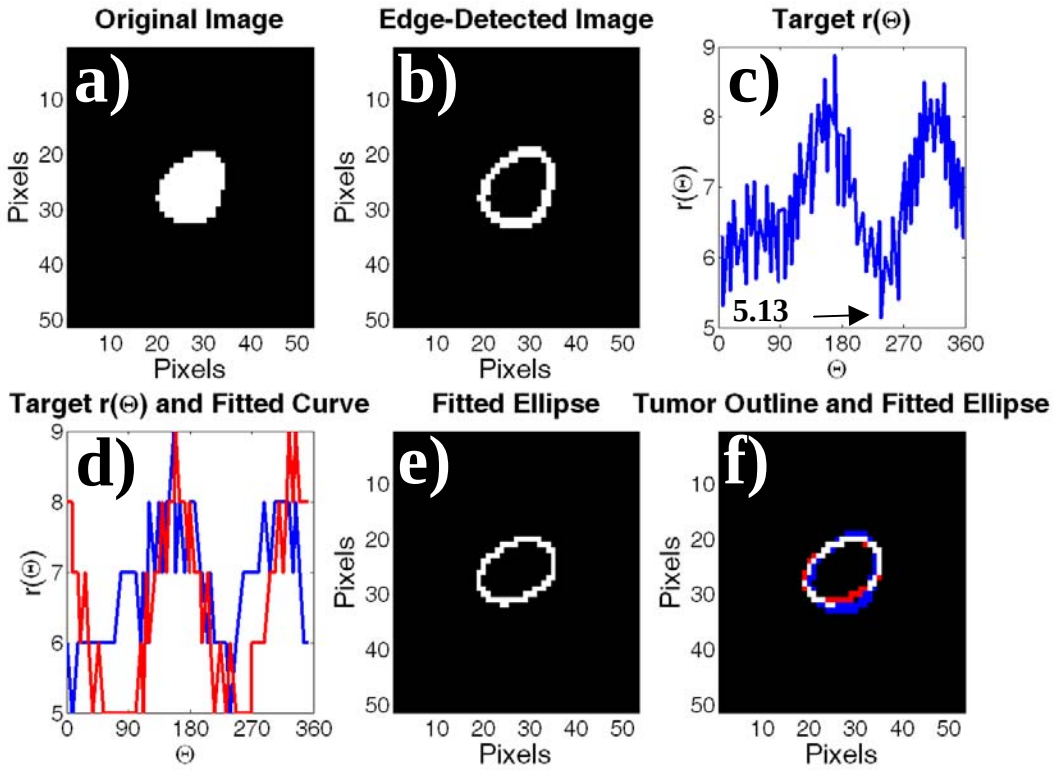
Figure 5.1: Output from the ECCF system for an ellipse with 100 pixel semi-major axis and 50 pixel semi minor axis at an angle of 105° off the horizontal axis. The elliptical quantities calculated are in close agreement with the actual dimensional parameters of the ellipse. The shape conformity is very good, slightly less than one pixel. The $\mathbf{R^2}$ calculation also indicates close agreement.

The ECCF system performs satisfactorily for the test images. Figures 5.2, 5.3 and 5.4 are demonstrations of the ECCF system output for actual tumor overlay images generated by the Sample-Tyler CAD system which have arbitrary shapes.



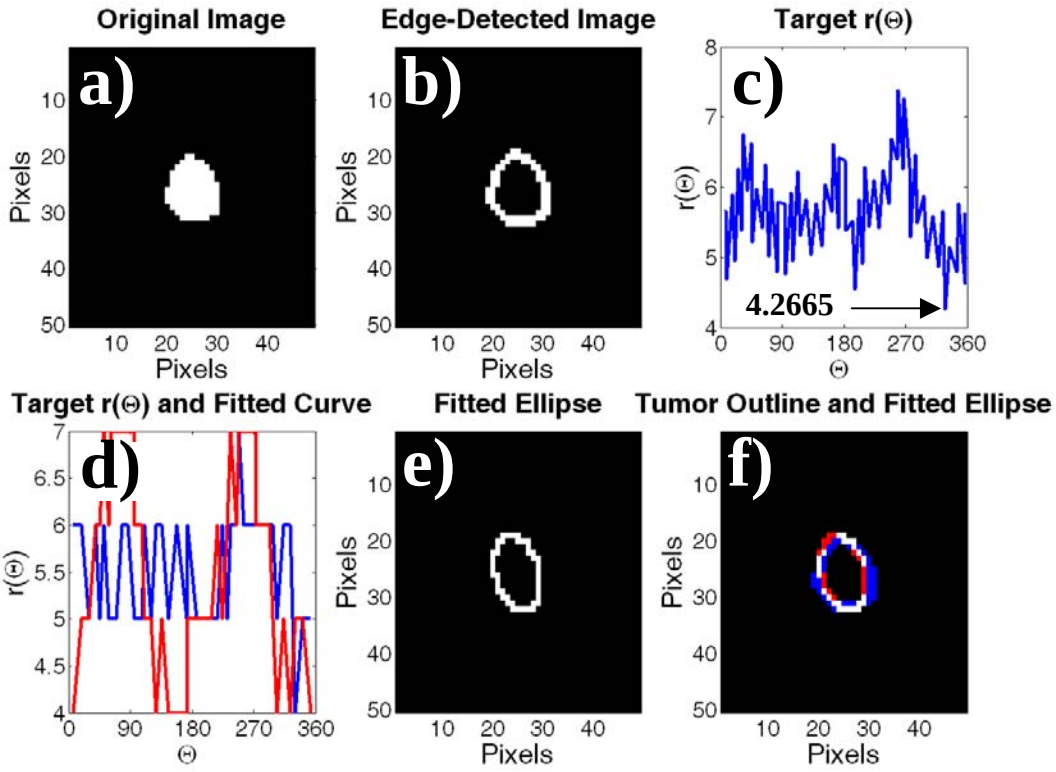
Semi-major axis (pxw)	16.0078
Semi-minor axis (pxw)	4.1496
Angle	104.4703
Area of Fitted Ellipse	208.6852
Area of tumor (px)	263.0000
Ratio of difference in area	0.7697
Eccentricity	0.9658
Shape conformity (pxw)	1.0129
R-squared	0.9017

Figure 5.2: Output data from the ECCF system for tumor shape 1. The original tumor image appears in a). Part b) show the tumor after edge-detection. The radial distribution function $r(\theta)$ is plotted in c). The vertical axis is in units of pixel length. The minor axis for the fitted ellipse is obtained from the absolute minimum of $r(\theta)$ which is 4.149 pixels. Tumor shape 1 is more elliptical in comparison to the tumor shape 2 and 3. This is reflected in the peaks appearing at close to 90˚ and 270˚. d) Shows the radial functions for the tumor and its fitted ellipse are in fairly closed agreement. A 2D image of the fitted ellipse is shown in e). f) shows an overlaying comparison of the fitted ellipse (red) with the outline of the tumor in (blue). The white pixels indicate exact overlap.



Semi-major axis (pxw)	8.5440
Semi-minor axis (pxw)	5.1330
Angle	20.5560
Area of Fitted Ellipse	137.7800
Area of tumor (px)	147.0000
Ratio of difference in area	0.9352
Eccentricity	0.7994
Shape conformity (pxw)	0.7877
R-squared	0.3497

Figure 5.3: Output data from the ECCF system for tumor shape 2. The original tumor image appears in a). Part b) show the tumor after edge-detection. The radial distribution function $r(\theta)$ is plotted in c). Tumor 2 is almost round as $r(\theta)$ fluctuate remains roughly a radius of 7 pixels. The minor axis for the fitted ellipse is obtained from the absolute minimum of $r(\theta)$ which is 5.13 pixels. d) Shows the radial functions for the tumor and its fitted ellipse are in fairly closed agreement. The agreement also bears out in e) and f).



Semi-major axis (pxw)	6.8007
Semi-minor axis (pxw)	4.2665
Angle	107.1027
Area of Fitted Ellipse	91.1546
Area of tumor (px)	100.0000
Ratio of difference in area	0.9075
Eccentricity	0.7787
Shape conformity (pxw)	0.7524
R-squared	0.1989

Figure 5.4: Output data from the ECCF system for tumor shape 3. Tumor 3 is also close to round. Its radial distribution is almost flat- similar to tumor 2. It is however slightly smaller than tumor 2 with a radius of roughly 5.5 pixels. The shortest radii in the $r(\theta)$ is shown to be 4.26 pixels in c). The shape conformity value reflects the agreement shown in d), e) and f).

The purpose of the ECCF system is not intended to test for perfect roundness, as perfectly elliptical shapes are not of interest. It is intended to extract objective numerical information from arbitrary shapes for the purpose of data mining. The quantities for

tumor area and eccentricity will always be approximate. R^2 and the shape conformity value can be used to indicate the amount a shape deviates from that of a fitted ellipse. We suspect that tumors that fall within some volume, or volumes, in the space spanned by the quantities we calculate and that the ECCF system may be trained to determine such volumes given data sets of significant size consisting of tumors with labeled pathologies. Future work would entail the construction and analysis of such data sets.

REFERENCES

[1] NIH Publication No. 05-1556

[2] J. Ferlay, F. Bray, P. Pisani and D.M. Parkin. GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide. Version 1.0. IARC CancerBase No. 5. Lyon, IARCPress, 2001.

[3] American Cancer Society. Breast Cancer Facts and Figures 2005-2006. Atlanta: American Cancer Socety, Inc.

[4] American Cancer Society. Cancer Facts and Figures 2006. Atlanta: Amerian Cancer Society; 2006.

[5] Hofvind SS, Wang H, Thoresen S., The Norwegian Breast Cancer Screening Program: re-attendance related to the women's experiences, intentions and previous screening result. The Cancer Registry of Norway, Oslo, Norway. May 2003

[6] Joseph, Chris K. 2004. R2 Technology and University of Chicago Announce License Agreement to Collaborate on Computer Aided Detection (CAD) Workstation. R2 Technology. Retrieved September 2006. (http://www.latimes.com/news/nation/updates/lat_vieques000505.htm).

[7] Oxford e-Science Centre. 2002. Oxford University, IBM and UK Government to Build Massive Computing Grid for Breast Cancer Screening and Diagnosis. Retrieved September 2006. (http://e-science.ox.ac.uk/pressreleases/ediamond.xml)

[8] Moores University of California San Diego Cancer Center. 2005. "UCSD Brings Digital Mammography to San Diego" San Diego, California. Retrieved September 2006. (http://cancer.ucsd.edu/Aboutus/News/stories/Digital_Mammo.asp)

[9] Sample, John. 2003. "Computer Assisted Screening of Digital Mammogram Images." Ph.D. dissertation, Department of Computer Science, Louisiana State University, Baton Rouge, LA.

APPENDIX A - LJPEG-TO-TIFF CONVERSION

The images we use from the Digital Database for Screening Mammography (DDSM) at the University of South Florida are stored in a Lossless JPEG image format-LJPEG. This document is a guide for converting LJPEG images from the DDSM at USF into the TIFF image format readable by MATLAB. JPEG is an image compression format commonly used to reduce the file sizes of images typically for use on the web. The JPEG compression scheme is lossy, that is, information is discarded from images during compression in order to obtain smaller image file sizes. The result is a significant reduction in file size accompanied by a reduction in image quality, which is however hardly noticeable to the human eye depending on the amount of compression performed on the image.

The LJPEG format is a lossless compression scheme as it results in no loss of image information. It is typically used in medical, military and space imaging, high-end film, professional studio-quality photography, and industrial machine vision systems.

This documentation is written for the specific purpose of converting LJPEG images from the Digital Database for Screening Mammography (DDSM), at the University of South Florida, to TIFF for use in MATLAB. It's detailed step-by-step to supplement Micheal Heath's documentation, so that one can get up and running quickly and easily.

A table summarizing the contents of the database can be found at

http://marathon.csee.usf.edu/Mammography/Database.html

Installation/Setup

The code is stated to be portable, but you're probably less likely to encounter compilation errors if you're running on a Unix platform. A Sun machine was used in this case.

Make a sub-directory in your home directory and change-directory to it

mkdir ~/heath

cd heath

Use ftp to download heathusf_v1.1.0.tar.gz into the heath subdirectory

ftp marathon.csee.usf.edu

login: "anonymous" (without the " ")

password: (just hit enter)

cd pub/heathusf

bin (make sure you're in binary transfer mode)

get heathusf_v1.1.0.tar.gz

bye

Uncompress and untar the software and compile the programs by using the following commands:

cd ~/heath

gunzip -c heathusf_v1.1.0.tar.gz | tar -xvf -

cd code

build_heathusf

MATLAB can import TIFF images. So we will use Heath's code to convert the LJPEG files to TIFF. Three executables will be required for this process, "case_decompress", "mkimage", "jpeg".

You should find the "case_decompress" executable all ready compiled in the subdirectory ".../heath/code/compression". "jpeg" should be in .../heath/code/compression/JPEG/PVRG/jpegdir. If it's not do the following:

```
cd ~/heath/code/compression/JPEG/PVRG/jpegdir
make
```

You will have to compile "mkimage". A file called "libmammo.a" is required. You will have to create this file in the heath/code/common subdirectory running an executable called build_library.

```
cd ~/heath/code/common
build_library
```

Compile the mkimage using its makefile

```
cd ~/heath/code/image
make
```

Since you'll need to call these executables from various sub-directories you'll want to set paths to the directories of these files. Edit your .login file to include the following paths to these executables

```
setenv PATH .../heath/code/image:$PATH
setenv PATH .../heath/code/compression:$PATH
setenv PATH .../heath/code/compression/JPEG/PVRG/jpegdir:$PATH
```

NOTE: You should replace the "..." with your home directory

Finally, copy “jpeg” into the same sub-directory as “case_decompress”.

cd ~/heath/code/compression

cp ~/heath/code/compression/JPEG/PVRG/jpegdir/jpeg .

NOTE: “case_decompress” calls “jpeg” when it runs. The paths we included in the .login file allow “jpeg” to be found and executed from any subdirectory when typed in at the command prompt. However “case_decompress” will only be able to find “jpeg” if it’s in the same directory with it.

Running the Programs

Change directory to a directory where you have a case of images. Run case_decompress to decompress the files

case_decompress

A LJPEG.1 file will be decompressed from each of the four LPJEG files. Convert these files to TIFF format using “mkimage” replacing <FILENAME> with the name of the .ics file you’ll find with each the case of images (example: A-0003-1.ics, A-0565-1.ics)

mkimage –ics <FILENAME> –view LEFT_CC -tif

mkimage –ics <FILENAME> –view RIGHT_CC -tif

mkimage –ics <FILENAME> –view LEFT_MLO -tif

mkimage –ics <FILENAME> –view RIGHT_MLO -tif

You should now have 4 TIFF image files for viewing in MATLAB

Handling TIFF Images in MATLAB

TIFF images of can be viewed and manipulated in MATLAB.

Start Matlab.

Go to File -> Import Data.

Locate the TIFF file you wish to view.

The Import Wizard window will appear, just click the “finish” button at the bottom.

You should see the TIFF image as matrix in the workspace window. (You might consider shortening the rather long file name (e.g. A_0002_1.LEFT_CC.LJPEG.1.image.tif). For this example the TIFF image matrix will be renamed “X”.) To view the TIFF image enter the following commands into MATLAB’s command window.

```
colormap('gray')
imagesc(X)
```

Imagesc() simply displays the matrix as pixel map. colormap() imposes a grayscale colormapping scheme on the image.

A “figure No.” window will appear with the TIFF image inside. The image however will more than likely be skewed as the aspect ratio of the image is not automatically adjusted when it is imported into MATLAB. To correct this, go to the drop down menu figure window and do the following.

Go to “Edit”

Under edit select “Figure properties”

A “Property editor – Figure” window will appear. In that widow:

Go to “Edit Properties for:” drop down menu at the top and click once.

It will open down and you’ll see “Root”, “Figure”, “Axes” & “Image” options

Select “Image”

Then click the button that says “Fix axis to image”

You should now see the TIFF image properly displayed in the “Figure No.1” window.

APPENDIX B

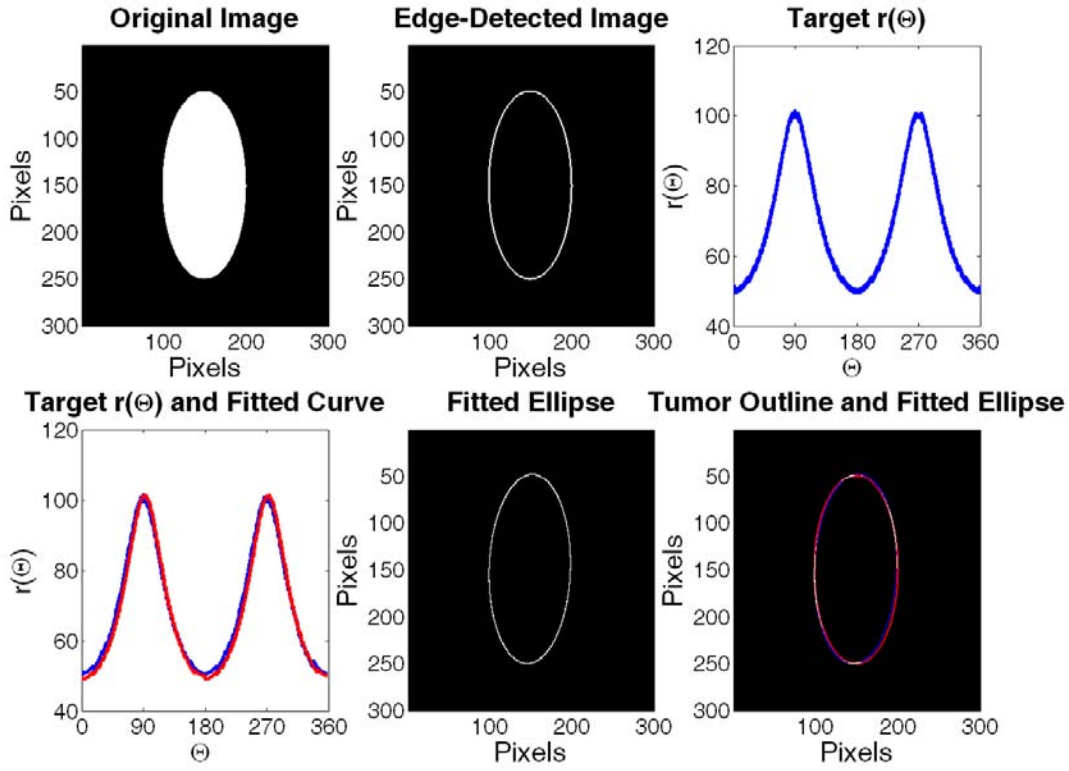
ELLIPSE TEST SETS

SET 1

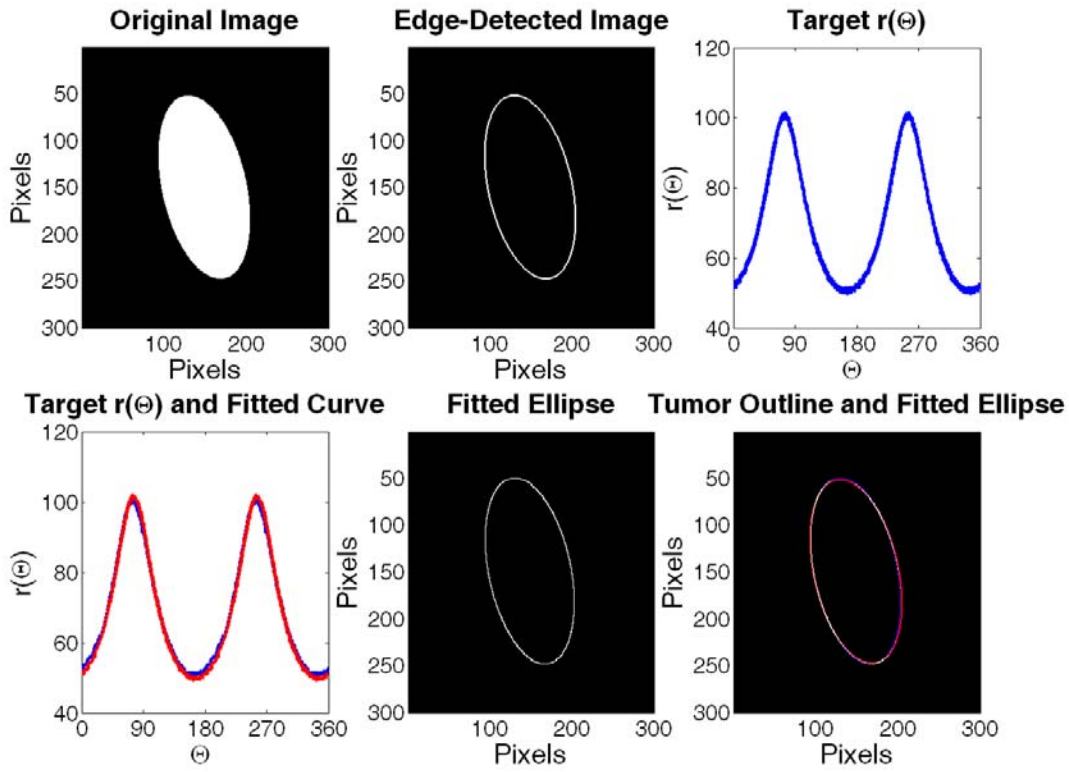
Dimensions

pxw – pixel width (unit of length)
px – pixels (unit of area)

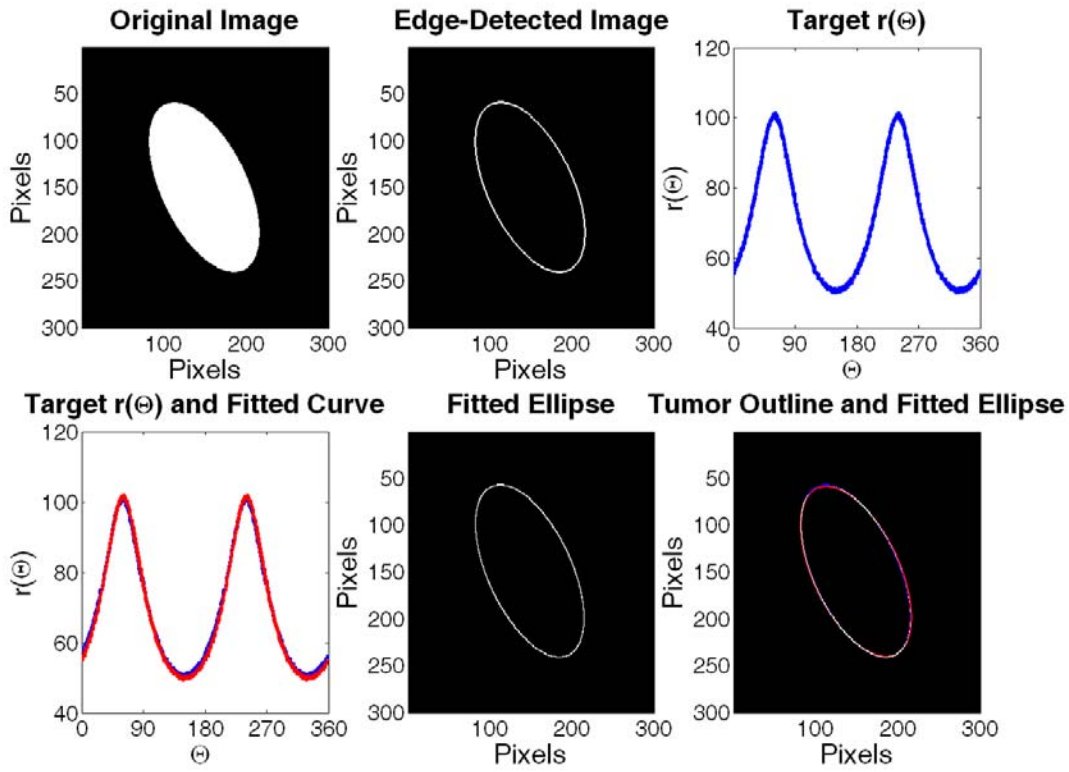
Angles test



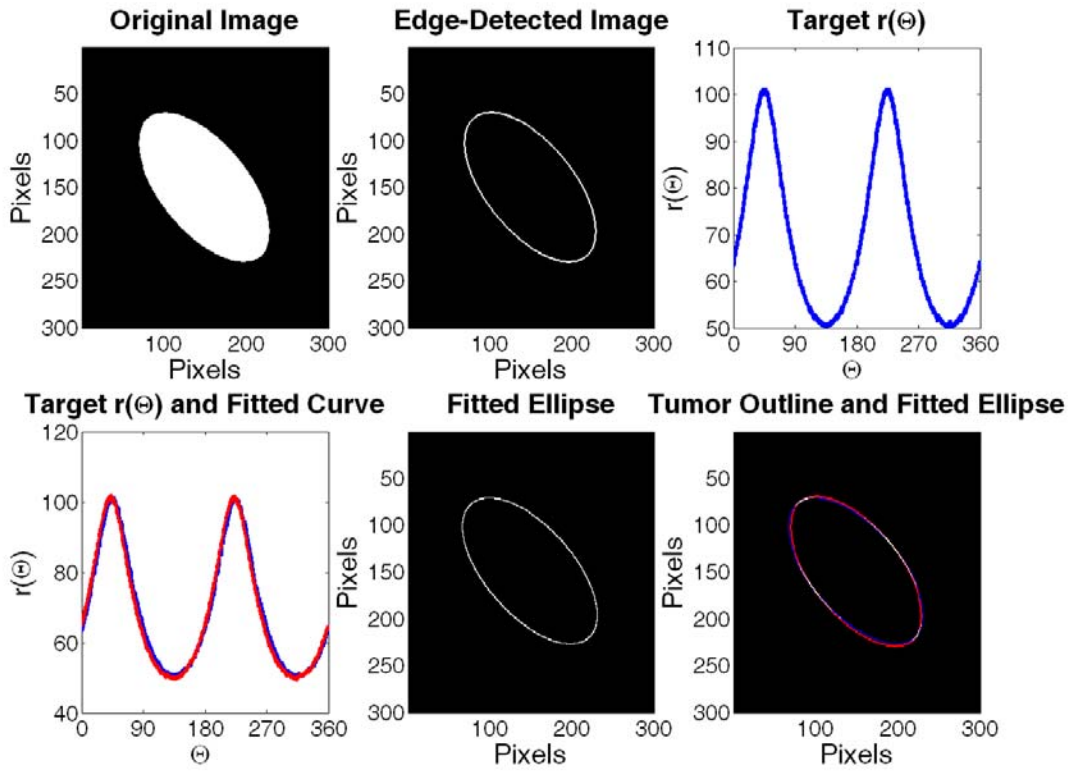
Calculated quantities - ellipse_angle_at_0.tif	
Semi-major axis (pxw)	101.0792
Semi-minor axis (pxw)	49.5092
Angle	87.7320
Area of Fitted Ellipse	1.572162e+004
Area of tumor (px)	16000.0000
Ratio of difference in area	0.9824
Eccentricity	0.8718
Shape conformity (pxw)	1.491547e+000
R-squared	0.9942



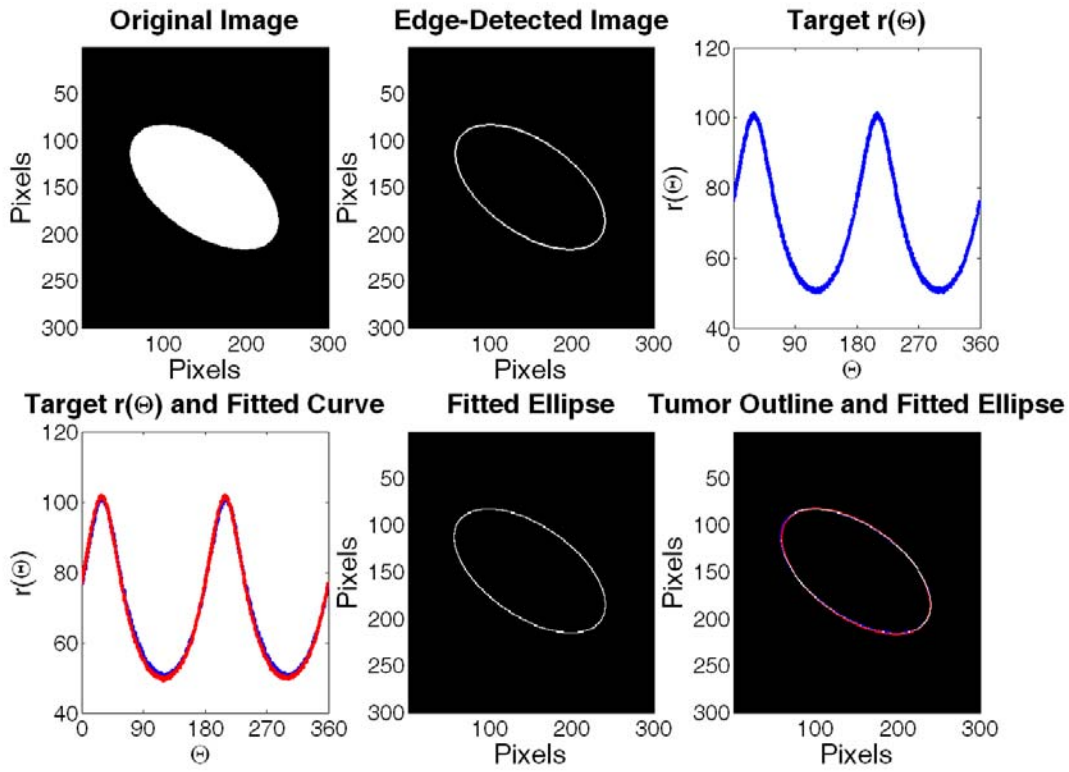
Calculated quantities - ellipse_angle_at_15.tif	
Semi-major axis (pxw)	101.5012
Semi-minor axis (pxw)	49.8448
Angle	103.9678
Area of Fitted Ellipse	1.589427e+004
Area of tumor (px)	16014.0000
Ratio of difference in area	0.9925
Eccentricity	0.8711
Shape conformity (pxw)	9.596332e-001
R-squared	0.9985



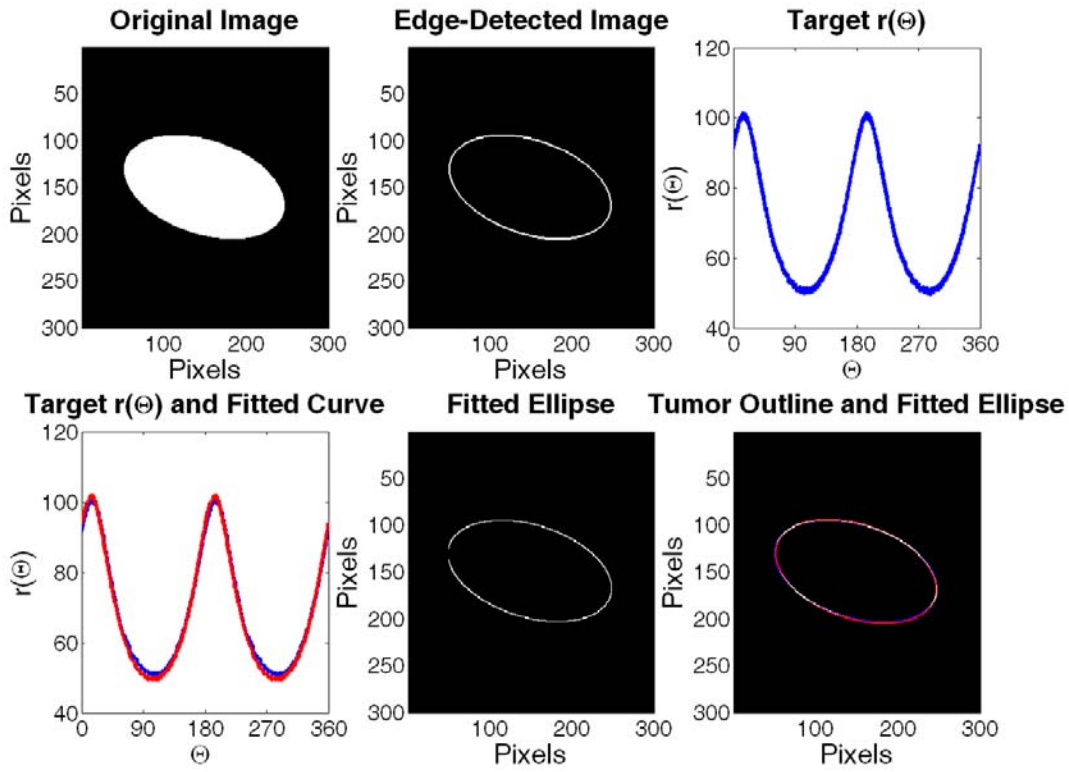
Calculated quantities - ellipse_angle_at_30.tif	
Semi-major axis (pxw)	101.4027
Semi-minor axis (pxw)	49.9270
Angle	119.2192
Area of Fitted Ellipse	1.590502e+004
Area of tumor (px)	16037.0000
Ratio of difference in area	0.9917
Eccentricity	0.8704
Shape conformity (pxw)	8.773067e-001
R-squared	0.9987



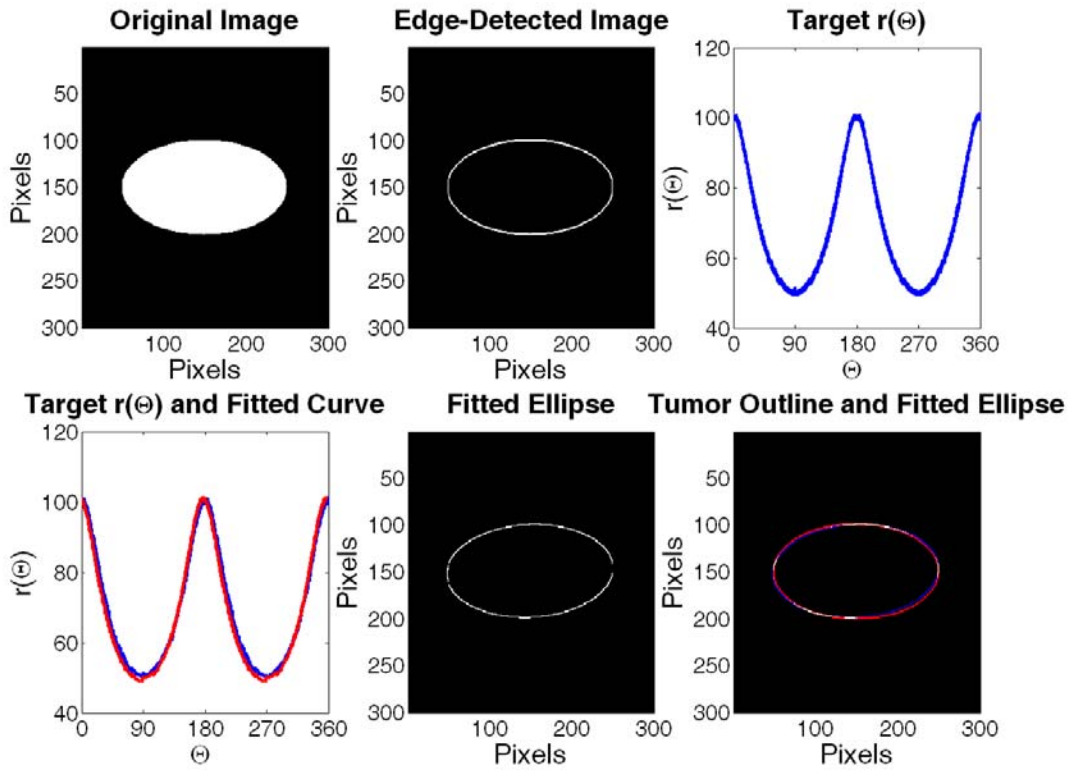
Calculated quantities - ellipse_angle_at_45.tif	
Semi-major axis (pxw)	101.2052
Semi-minor axis (pxw)	50.0843
Angle	137.4026
Area of Fitted Ellipse	1.592407e+004
Area of tumor (px)	16043.0000
Ratio of difference in area	0.9926
Eccentricity	0.8690
Shape conformity (pxw)	1.539821e+000
R-squared	0.9939



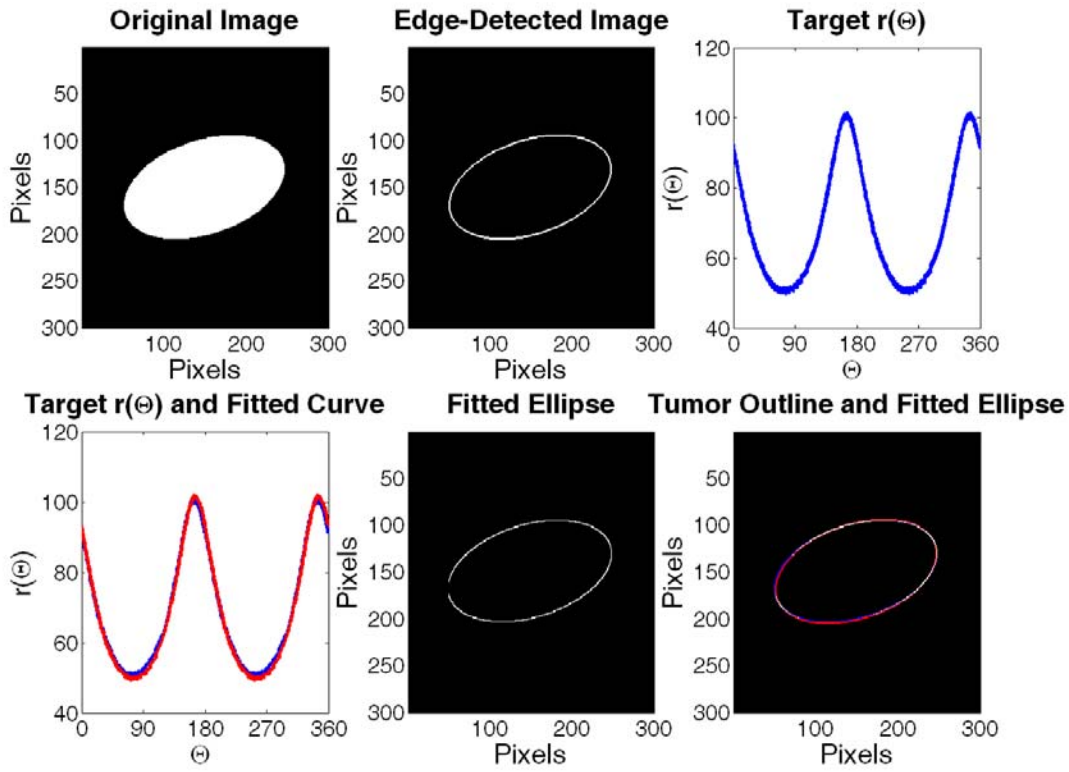
Calculated quantities - ellispe_angle_at_60.tif	
Semi-major axis (pxw)	101.4027
Semi-minor axis (pxw)	49.9229
Angle	150.7808
Area of Fitted Ellipse	1.590374e+004
Area of tumor (px)	16037.0000
Ratio of difference in area	0.9917
Eccentricity	0.8704
Shape conformity (pxw)	8.779854e-001
R-squared	0.9987



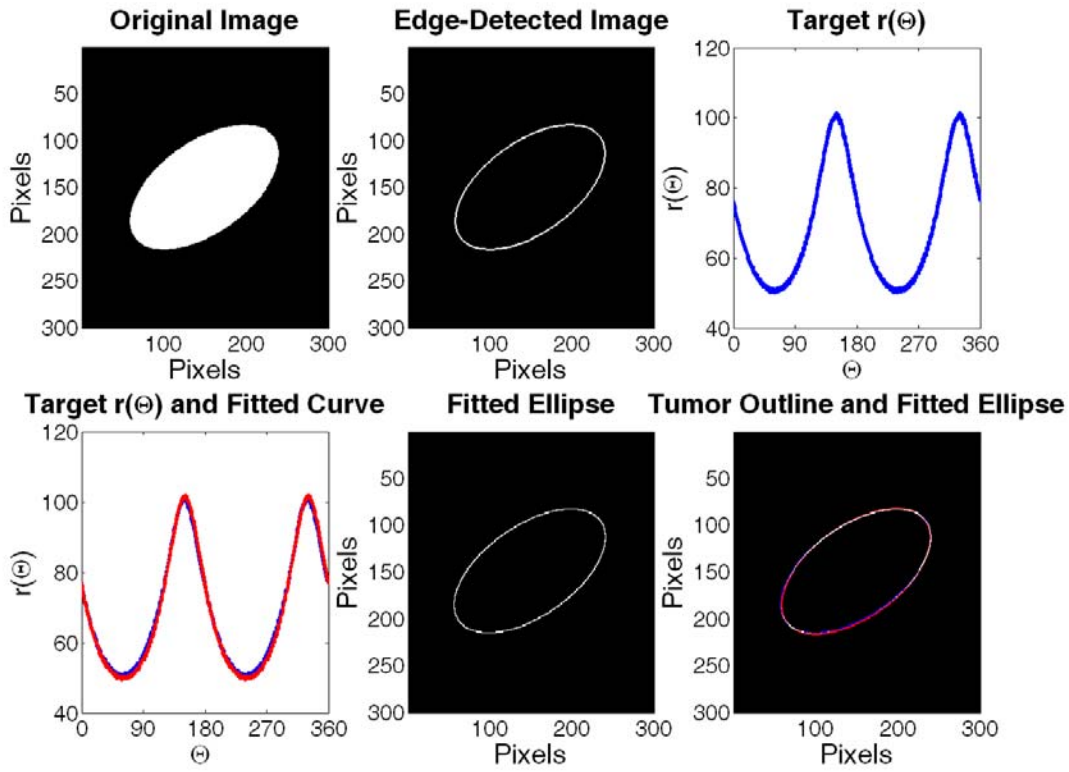
Calculated quantities - ellispe_angle_at_75.tif	
Semi-major axis (pxw)	101.5012
Semi-minor axis (pxw)	49.6660
Angle	166.0322
Area of Fitted Ellipse	1.583728e+004
Area of tumor (px)	16013.0000
Ratio of difference in area	0.9890
Eccentricity	0.8721
Shape conformity (pxw)	1.017578e+000
R-squared	0.9984



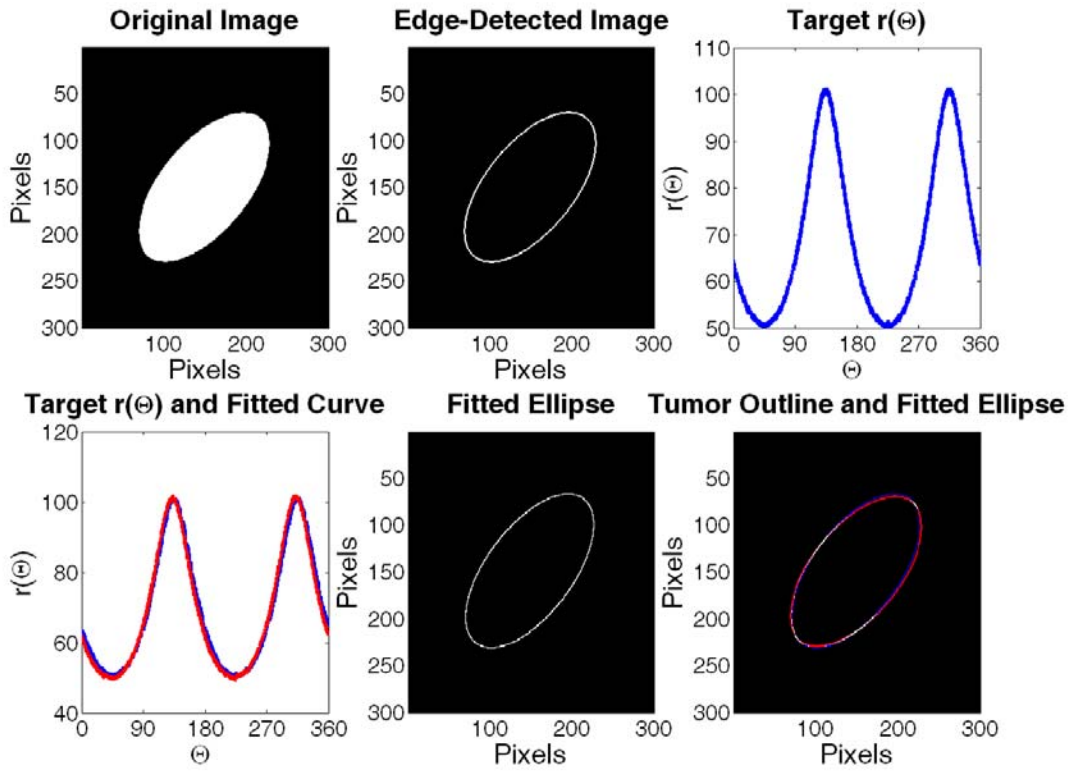
Calculated quantities - ellispe_angle_at_90.tif	
Semi-major axis (pxw)	101.0792
Semi-minor axis (pxw)	49.5003
Angle	2.2680
Area of Fitted Ellipse	1.571879e+004
Area of tumor (px)	15998.0000
Ratio of difference in area	0.9824
Eccentricity	0.8719
Shape conformity (pxw)	1.490115e+000
R-squared	0.9944



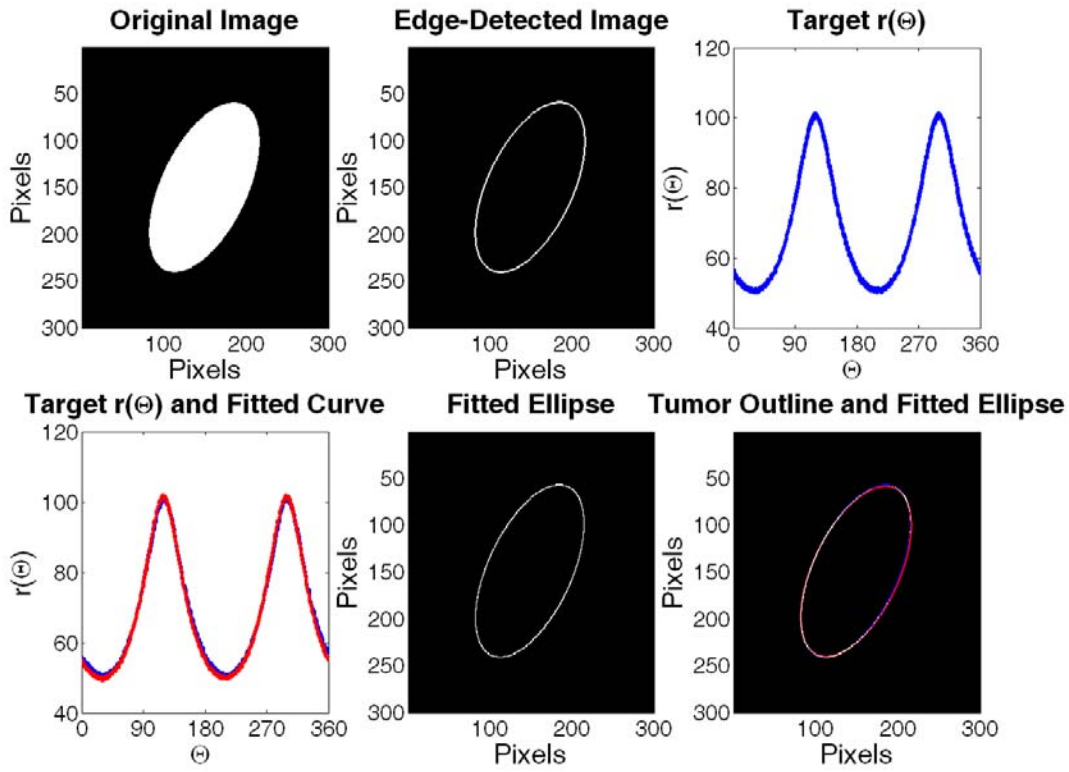
Calculated quantities - ellispe_angle_at_105.tif	
Semi-major axis (pxw)	101.5012
Semi-minor axis (pxw)	49.8448
Angle	13.9678
Area of Fitted Ellipse	1.589427e+004
Area of tumor (px)	16014.0000
Ratio of difference in area	0.9925
Eccentricity	0.8711
Shape conformity (pxw)	9.596332e-001
R-squared	0.9985



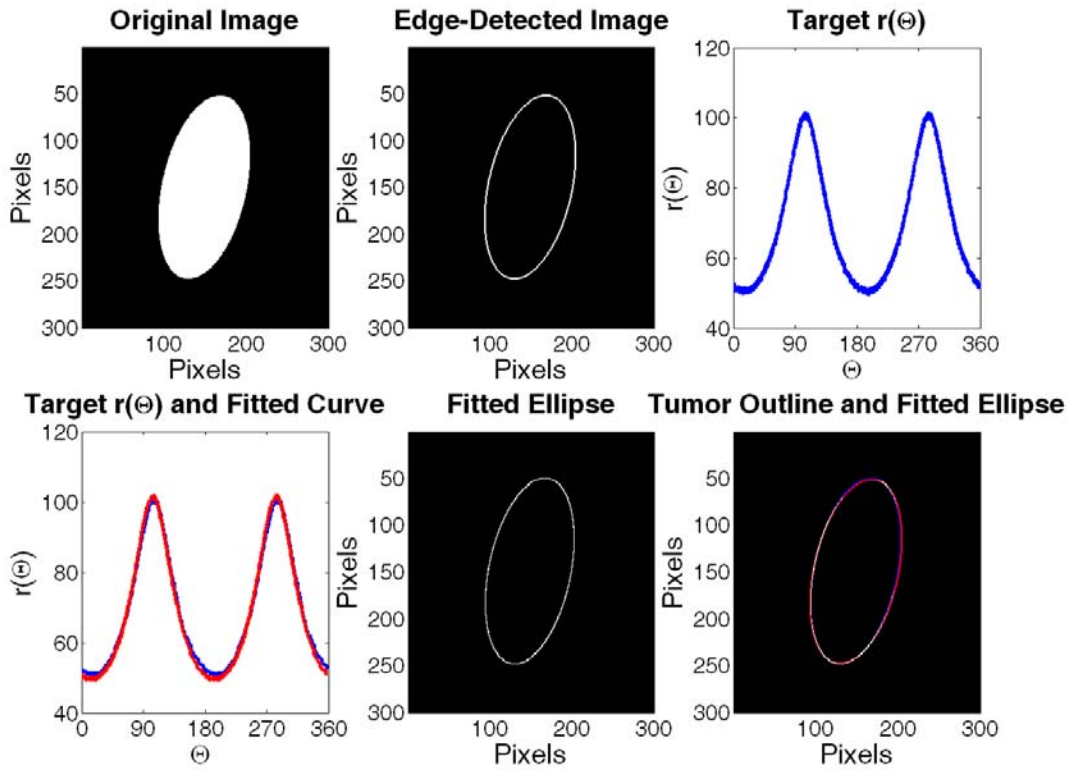
Calculated quantities - `ellipse_angle_at_120.tif`	
Semi-major axis (pxw)	101.4027
Semi-minor axis (pxw)	49.9270
Angle	29.2192
Area of Fitted Ellipse	1.590502e+004
Area of tumor (px)	16037.0000
Ratio of difference in area	0.9917
Eccentricity	0.8704
Shape conformity (pxw)	8.773067e-001
R-squared	0.9987



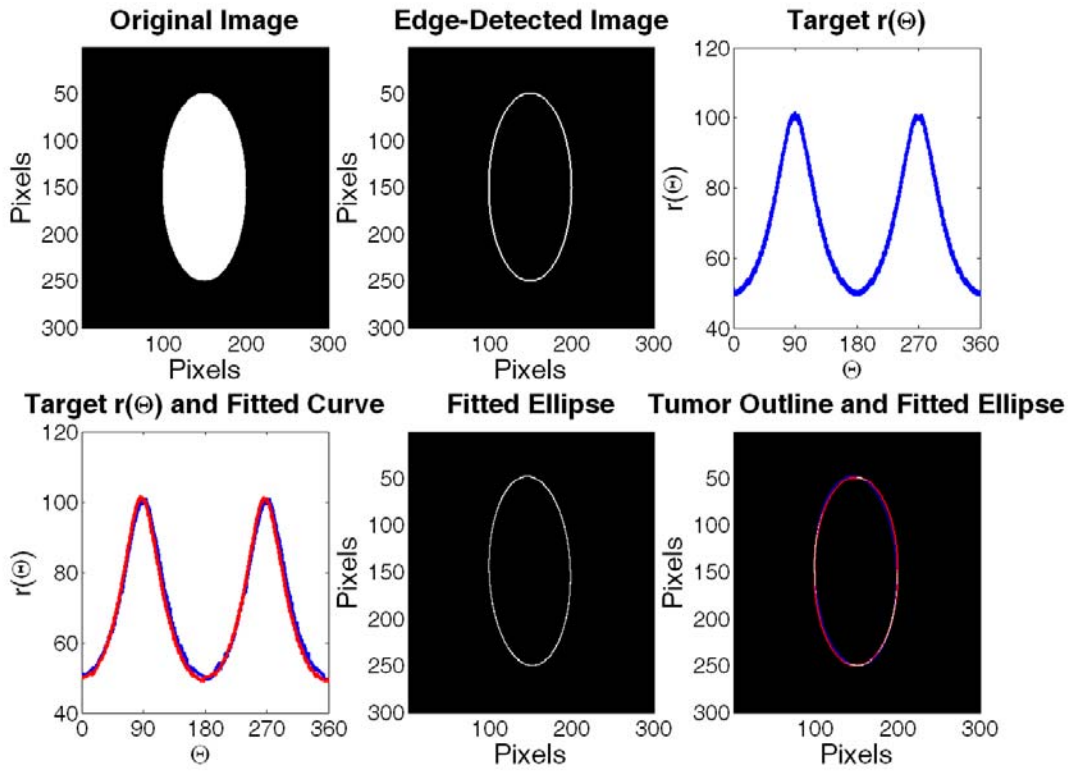
Calculated quantities - ellispe_angle_at_135.tif	
Semi-major axis (pxw)	101.2052
Semi-minor axis (pxw)	50.0843
Angle	47.4026
Area of Fitted Ellipse	1.592407e+004
Area of tumor (px)	16043.0000
Ratio of difference in area	0.9926
Eccentricity	0.8690
Shape conformity (pxw)	1.539821e+000
R-squared	0.9940



Calculated quantities - ellispe_angle_at_150.tif	
Semi-major axis (pxw)	101.4027
Semi-minor axis (pxw)	49.9229
Angle	60.7808
Area of Fitted Ellipse	1.590374e+004
Area of tumor (px)	16037.0000
Ratio of difference in area	0.9917
Eccentricity	0.8704
Shape conformity (pxw)	8.779854e-001
R-squared	0.9987



Calculated quantities - ellispe_angle_at_165.tif	
Semi-major axis (pxw)	101.5012
Semi-minor axis (pxw)	49.6660
Angle	76.0322
Area of Fitted Ellipse	1.583728e+004
Area of tumor (px)	16013.0000
Ratio of difference in area	0.9890
Eccentricity	0.8721
Shape conformity (pxw)	1.017578e+000
R-squared	0.9984



Calculated quantities - ellispe_angle_at_180.tif	
Semi-major axis (pxw)	101.0792
Semi-minor axis (pxw)	49.5003
Angle	92.2680
Area of Fitted Ellipse	1.571881e+004
Area of tumor (px)	15998.0000
Ratio of difference in area	0.9824
Eccentricity	0.8719
Shape conformity (pxw)	1.490109e+000
R-squared	0.9941

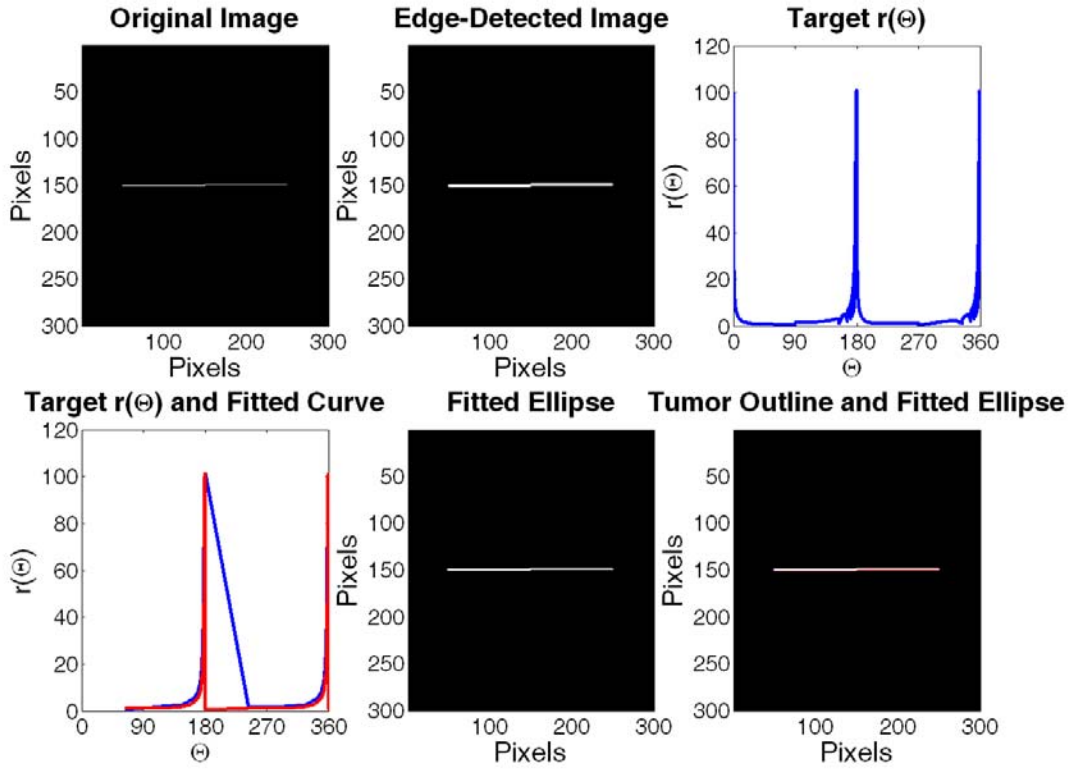
APPENDIX B

SET 2

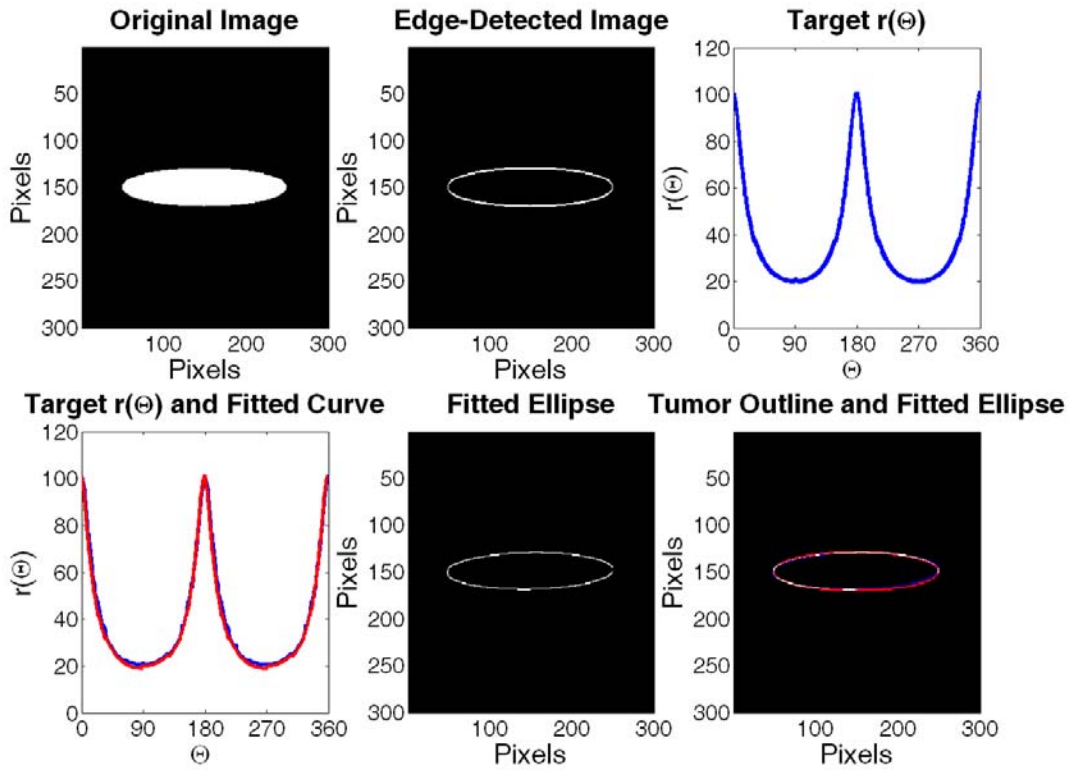
Dimensions

pxw – pixel width (unit of length)
px – pixels (unit of area)

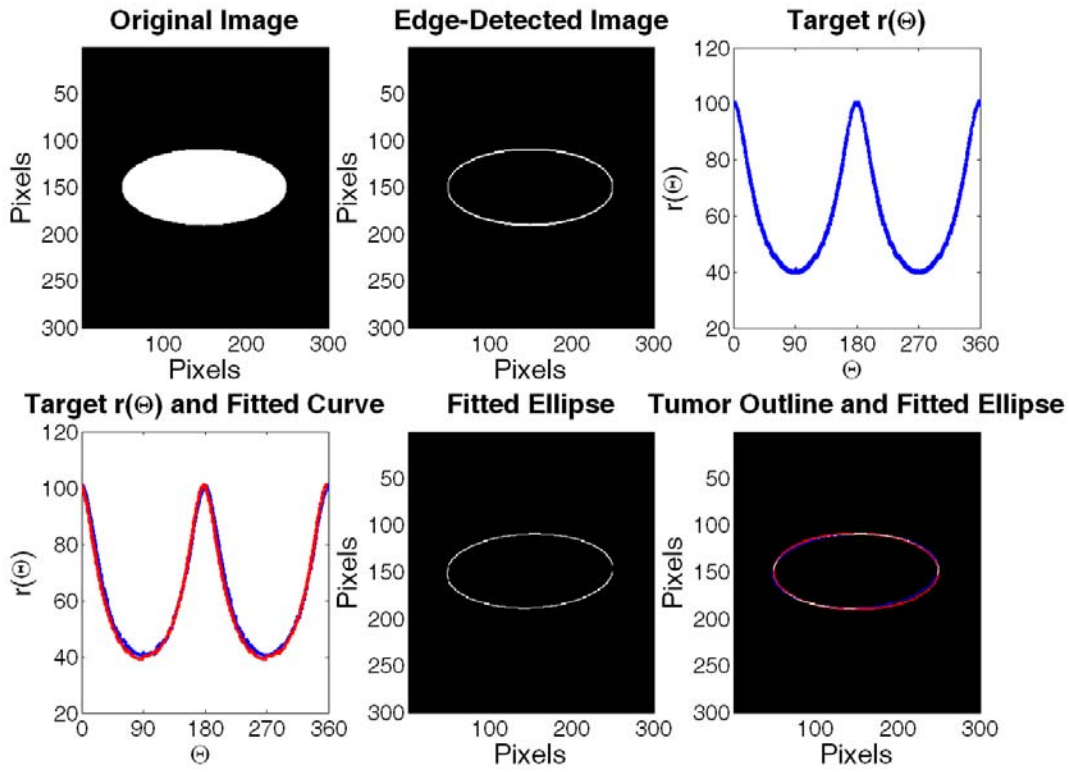
Eccentricity examples for horizontal ellipses



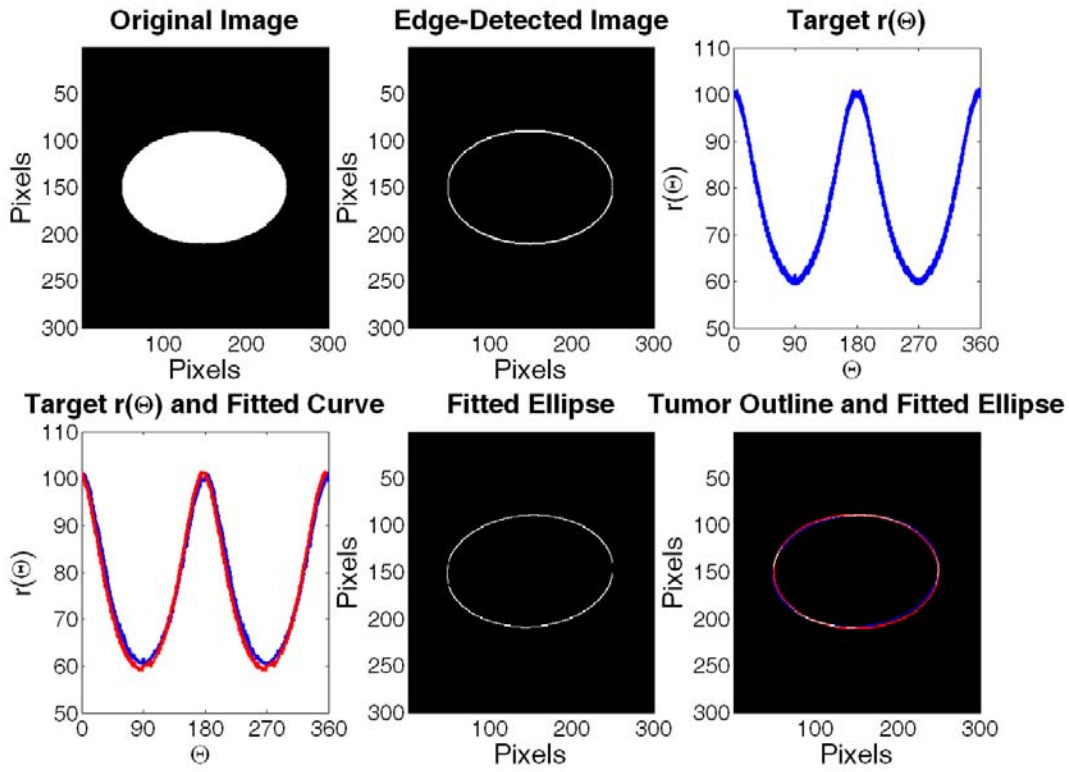
Calculated quantities - `Ellispe_ecc_at_0.tif`	
Semi-major axis (pxw)	`101.0012`
Semi-minor axis (pxw)	`0.4951`
Angle	`90.2836`
Area of Fitted Ellipse	`1.570824e+002`
Area of tumor (px)	`202.0000`
Ratio of difference in area	`0.7498`
Eccentricity	`1.0000`
Shape conformity (pxw)	`6.378707e-001`
R-squared	`0.1733`



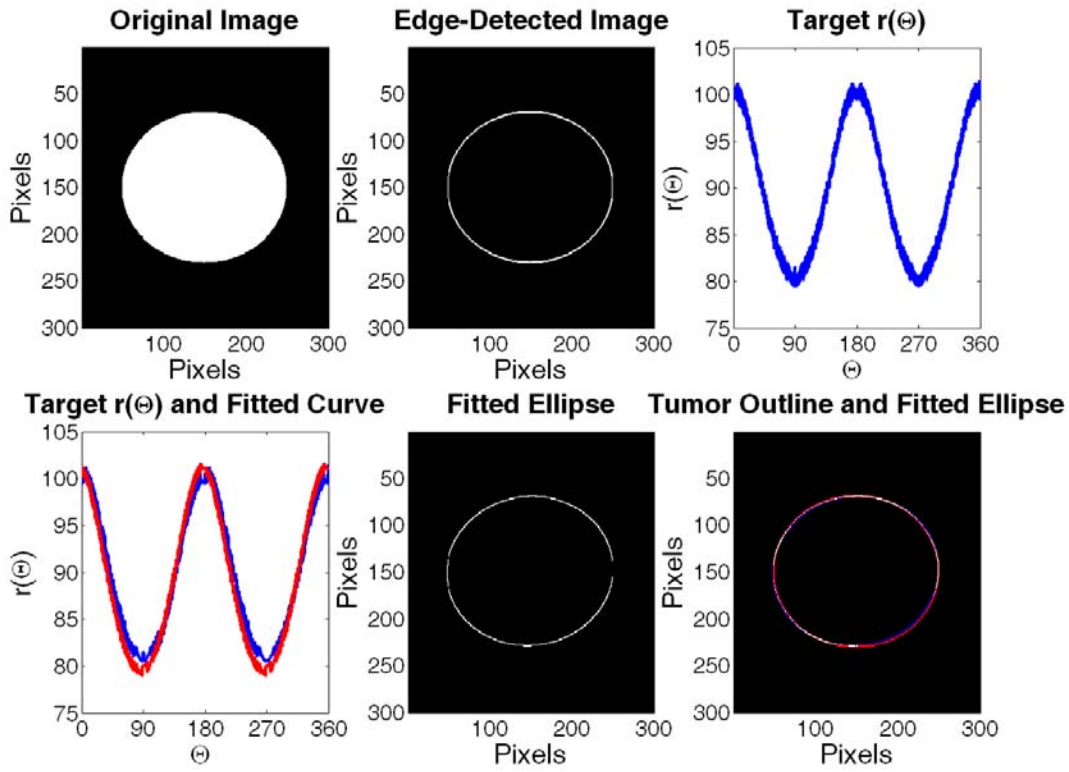
Calculated quantities - Ellispe_ecc_at_20.tif	
Semi-major axis (pxw)	101.0111
Semi-minor axis (pxw)	19.5040
Angle	90.8509
Area of Fitted Ellipse	6.189329e+003
Area of tumor (px)	6494.0000
Ratio of difference in area	0.9520
Eccentricity	0.9812
Shape conformity (pxw)	1.135740e+000
R-squared	0.9959



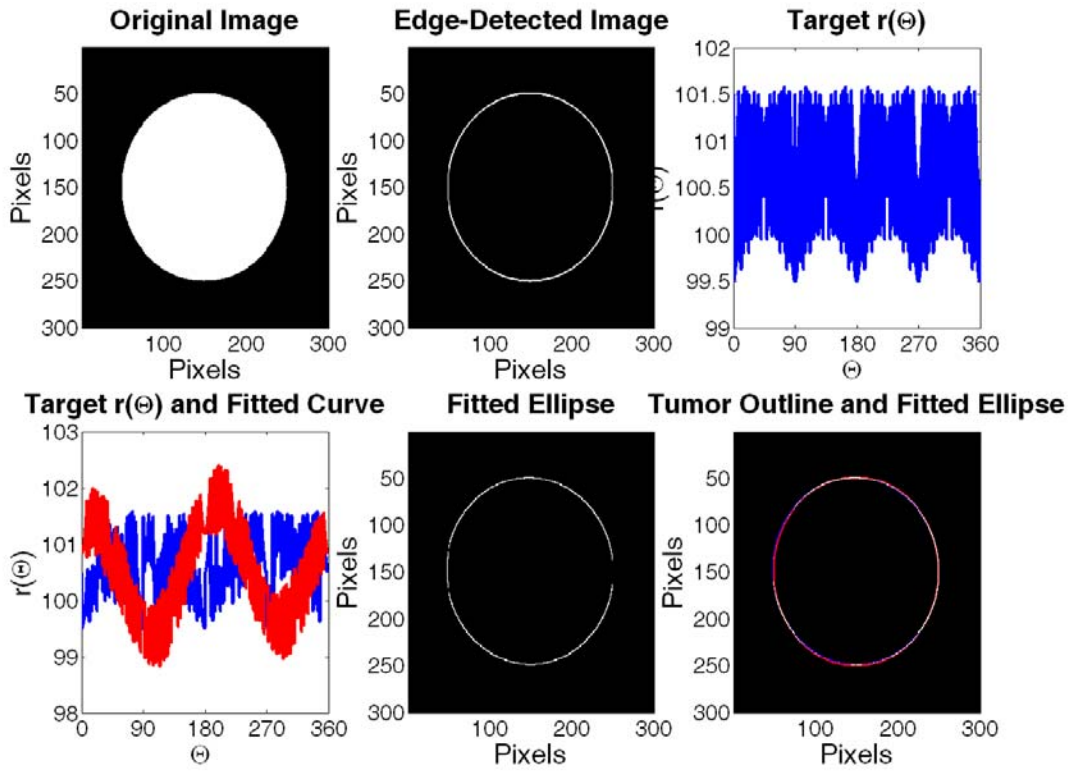
Calculated quantities - `ellispe_ecc_at_40.tif`	
Semi-major axis (pxw)	`101.0445`
Semi-minor axis (pxw)	`39.5009`
Angle	`91.7014`
Area of Fitted Ellipse	`1.253919e+004`
Area of tumor (px)	`12830.0000`
Ratio of difference in area	`0.9771`
Eccentricity	`0.9204`
Shape conformity (pxw)	`1.391722e+000`
R-squared	`0.9957`



Calculated quantities - ellispe_ecc_at_60.tif	
Semi-major axis (pxw)	101.1002
Semi-minor axis (pxw)	59.5000
Angle	87.4489
Area of Fitted Ellipse	1.889813e+004
Area of tumor (px)	19146.0000
Ratio of difference in area	0.9870
Eccentricity	0.8085
Shape conformity (pxw)	1.398127e+000
R-squared	0.9938



Calculated quantities - `ellipse_ecc_at_80.tif`	
Semi-major axis (pxw)	`101.1781`
Semi-minor axis (pxw)	`79.4997`
Angle	`86.6003`
Area of Fitted Ellipse	`2.526978e+004`
Area of tumor (px)	`25478.0000`
Ratio of difference in area	`0.9918`
Eccentricity	`0.6186`
Shape conformity (pxw)	`1.161083e+000`
R-squared	`0.9871`



Calculated quantities - ellispe_ecc_at_100.tif	
Semi-major axis (pxw)	101.5800
Semi-minor axis (pxw)	99.4995
Angle	163.7060
Area of Fitted Ellipse	3.175258e+004
Area of tumor (px)	31814.0000
Ratio of difference in area	0.9981
Eccentricity	0.2014
Shape conformity (pxw)	9.518279e-001
R-squared	-0.1608

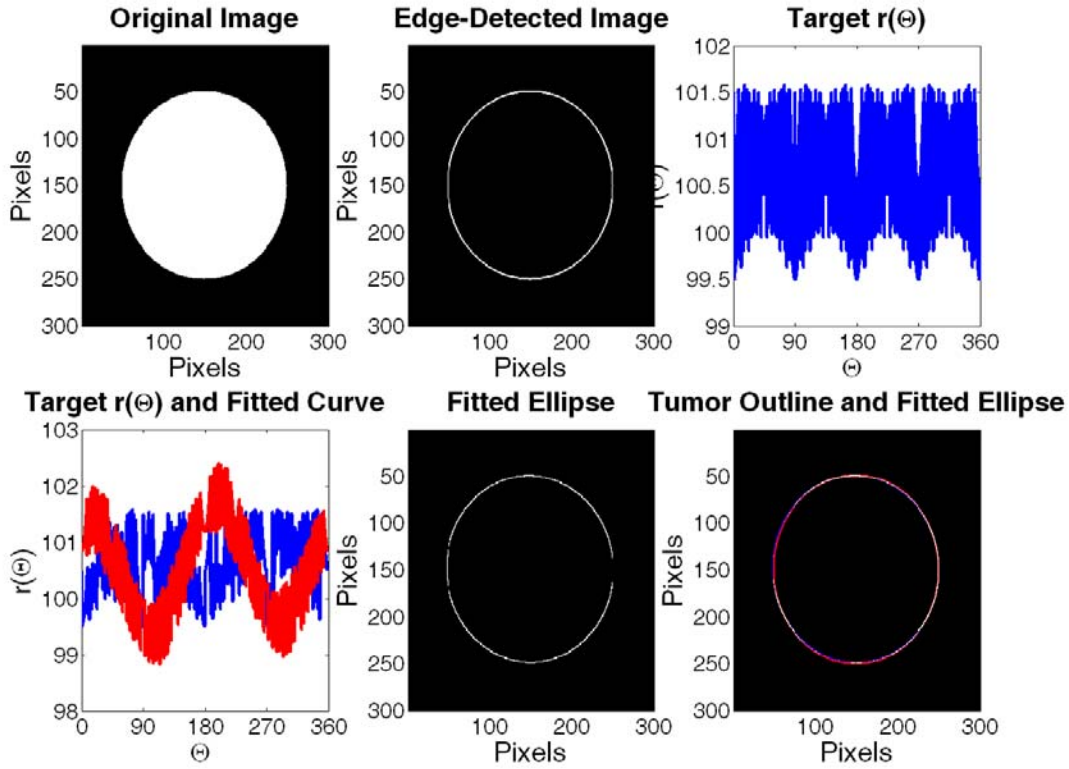
APPENDIX B

SET 3

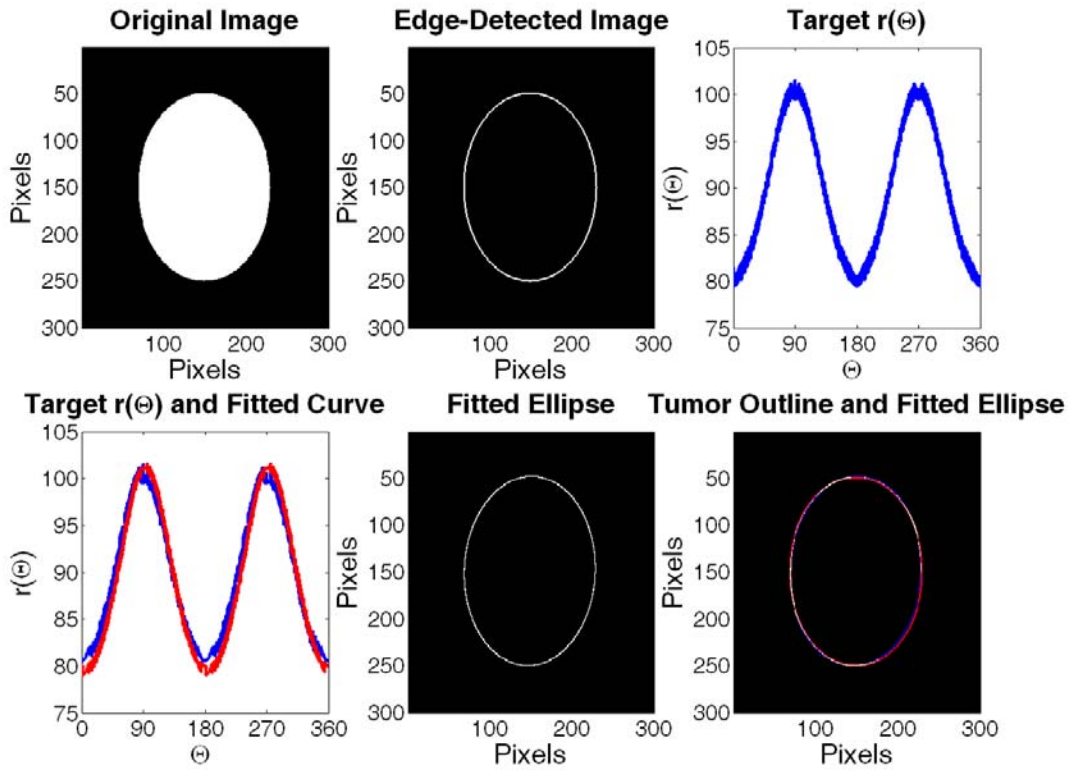
Dimensions

pxw – pixel width (unit of length)
px – pixels (unit of area)

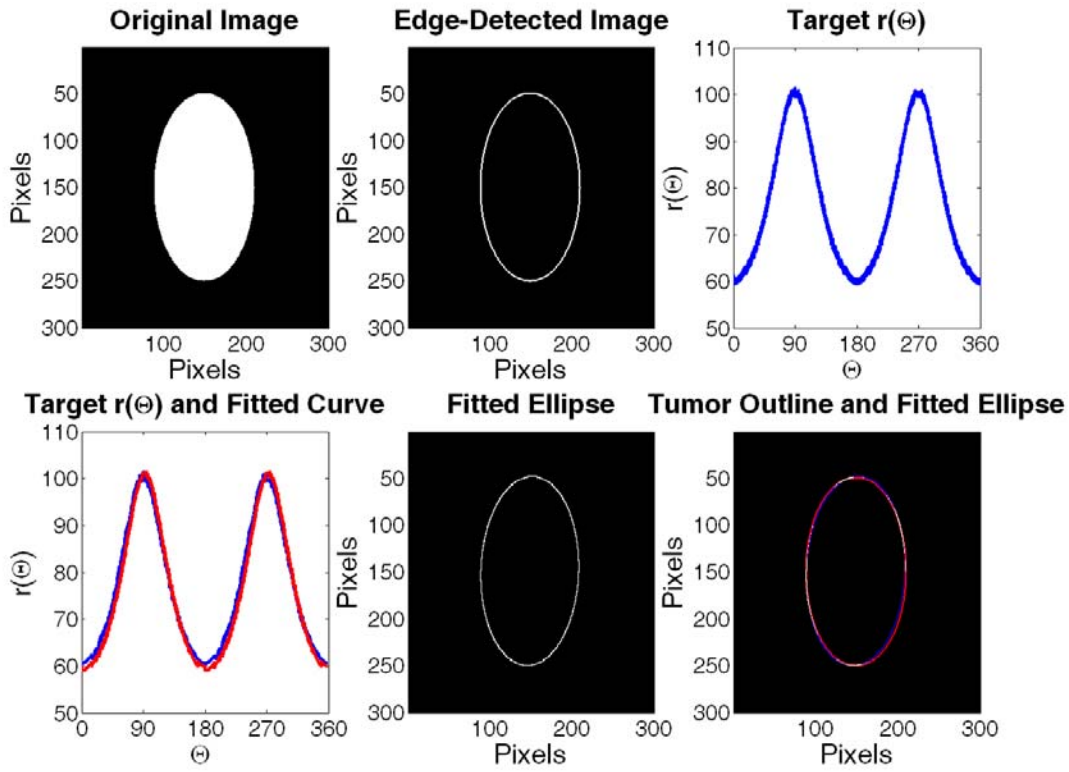
Eccentricity examples for vertical ellipses



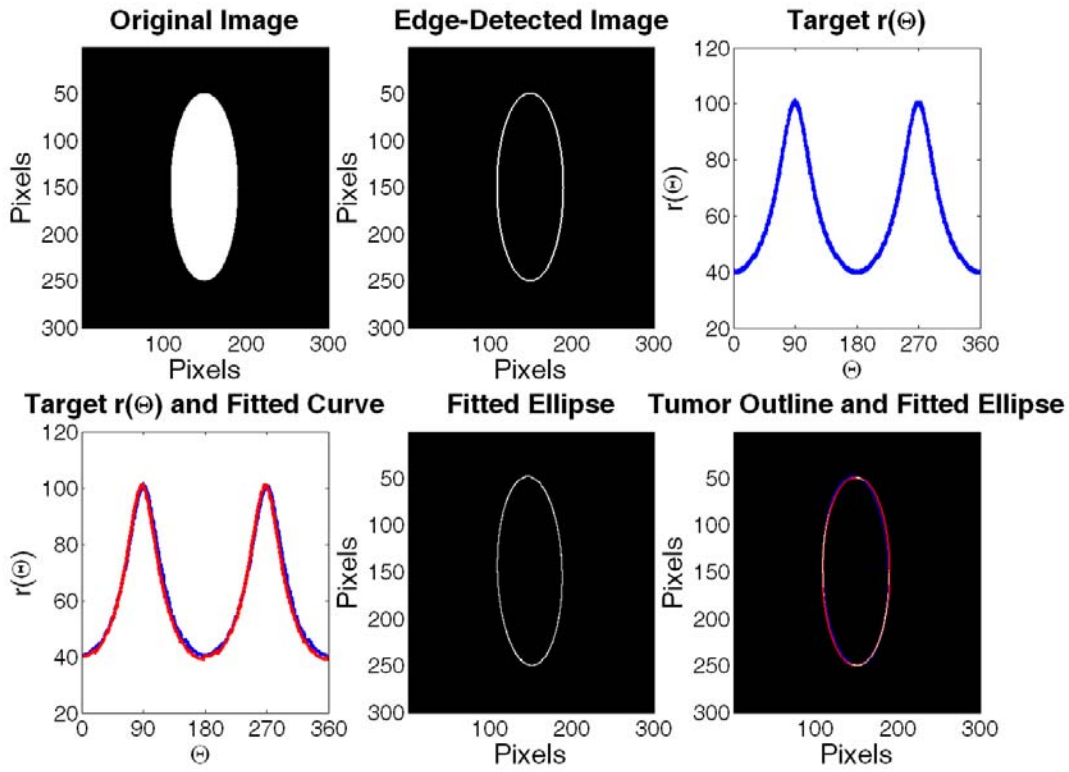
Calculated quantities – `ellipse_ecc_at_100.tif`	
Semi-major axis (pxw)	101.5800
Semi-minor axis (pxw)	99.4995
Angle	163.7060
Area of Fitted Ellipse	3.175258e+004
Area of tumor (px)	31814.0000
Ratio of difference in area	0.9981
Eccentricity	0.2014
Shape conformity (pxw)	9.518279e-001
R-squared	-0.1608



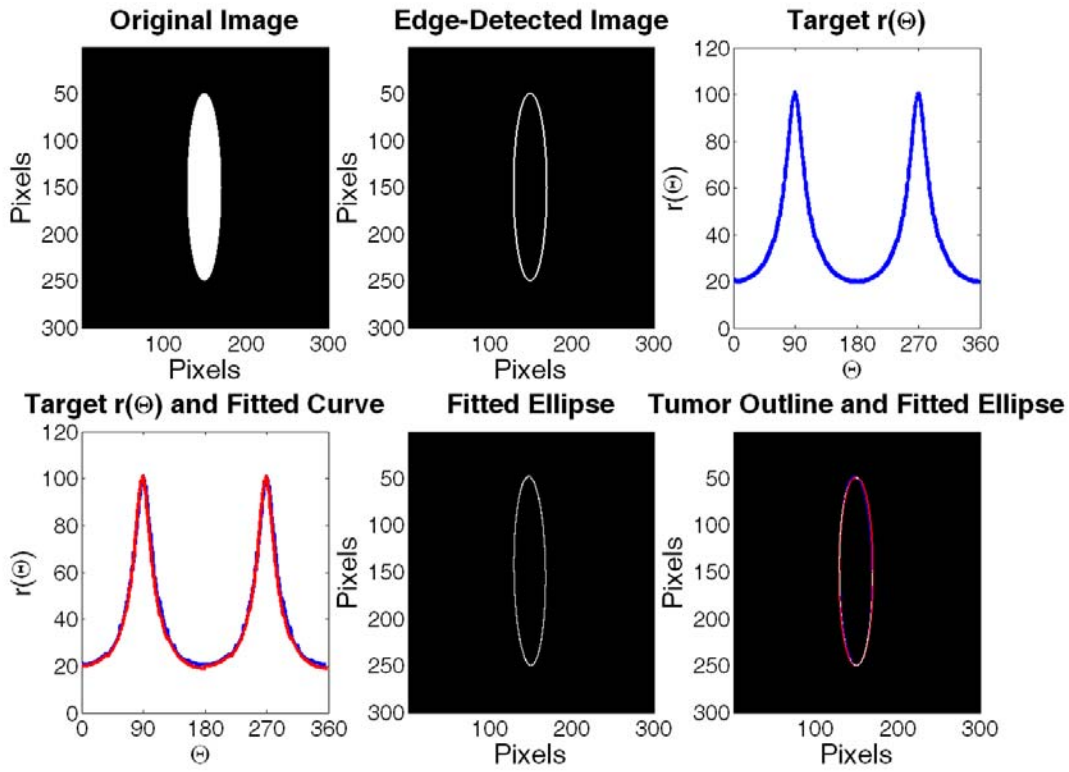
Calculated quantities - `ellipse_ecc_at_80.tif`	
Semi-major axis (pxw)	101.1781
Semi-minor axis (pxw)	79.4997
Angle	86.6003
Area of Fitted Ellipse	2.526978e+004
Area of tumor (px)	25478.0000
Ratio of difference in area	0.9918
Eccentricity	0.6186
Shape conformity (pxw)	1.161083e+000
R-squared	0.9871



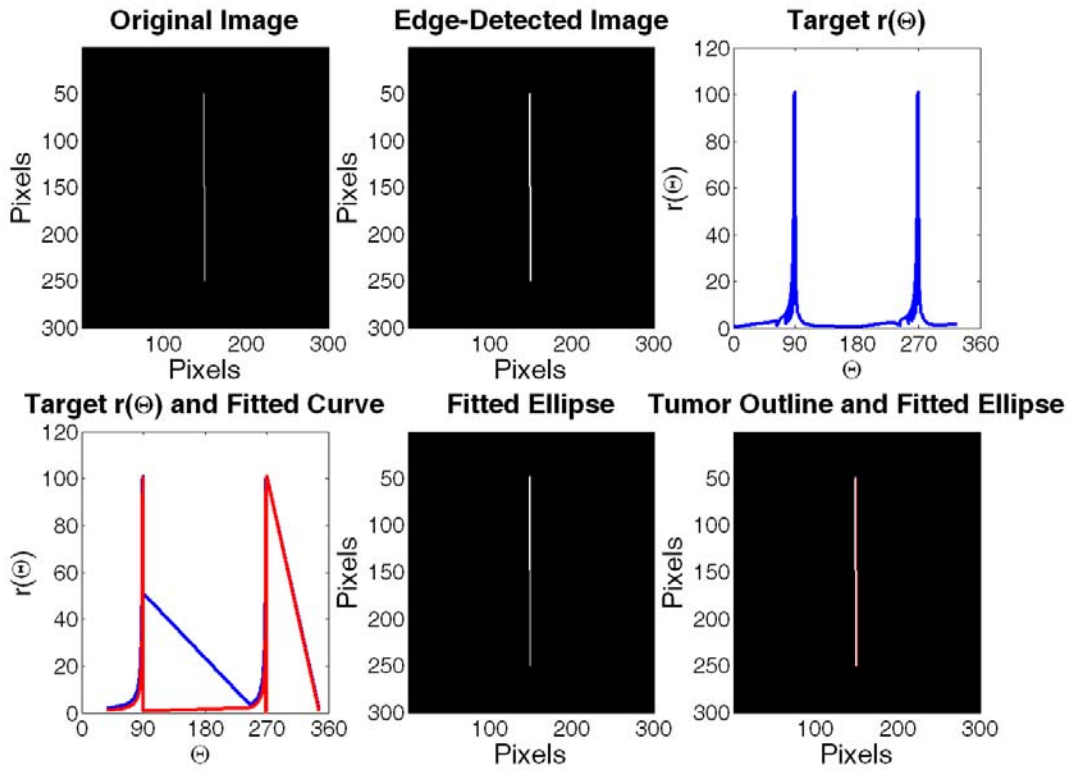
Calculated quantities - `ellipse_ecc_at_60.tif`	
Semi-major axis (pxw)	101.1002
Semi-minor axis (pxw)	59.5000
Angle	87.4489
Area of Fitted Ellipse	1.889813e+004
Area of tumor (px)	19146.0000
Ratio of difference in area	0.9870
Eccentricity	0.8085
Shape conformity (pxw)	1.398127e+000
R-squared	0.9938



Calculated quantities - `ellipse_ecc_at_40.tif`	
Semi-major axis (pxw)	101.0445
Semi-minor axis (pxw)	39.5009
Angle	91.7014
Area of Fitted Ellipse	1.253919e+004
Area of tumor (px)	12830.0000
Ratio of difference in area	0.9771
Eccentricity	0.9204
Shape conformity (pxw)	1.391722e+000
R-squared	0.9957



Calculated quantities - `ellipse_ecc_at_20.tif`	
Semi-major axis (pxw)	101.0111
Semi-minor axis (pxw)	19.5040
Angle	90.8509
Area of Fitted Ellipse	6.189329e+003
Area of tumor (px)	6494.0000
Ratio of difference in area	0.9520
Eccentricity	0.9812
Shape conformity (pxw)	1.135740e+000
R-squared	0.9959



Calculated quantities - `ellipse_ecc_at_0.tif`	
Semi-major axis (pxw)	101.0012
Semi-minor axis (pxw)	0.4951
Angle	90.2836
Area of Fitted Ellipse	1.570824e+002
Area of tumor (px)	202.0000
Ratio of difference in area	0.7498
Eccentricity	1.0000
Shape conformity (pxw)	6.378707e-001
R-squared	0.1733

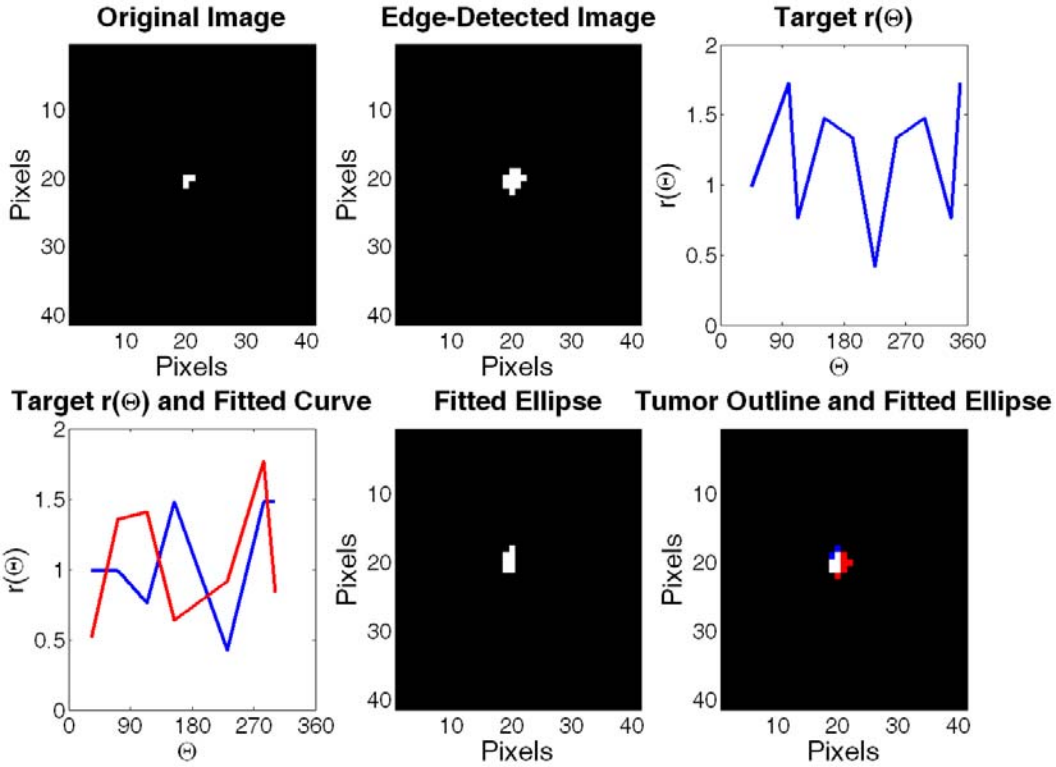
APPENDIX B

SET 4

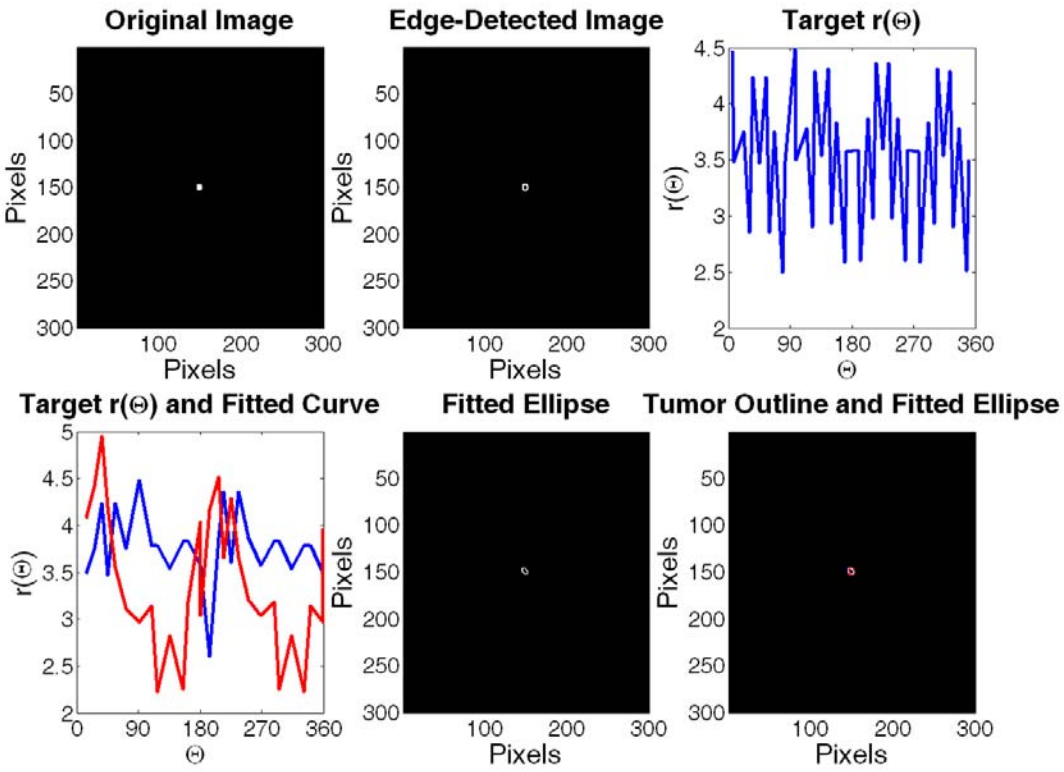
Dimensions

pxw – pixel width (unit of length)
px – pixels (unit of area)

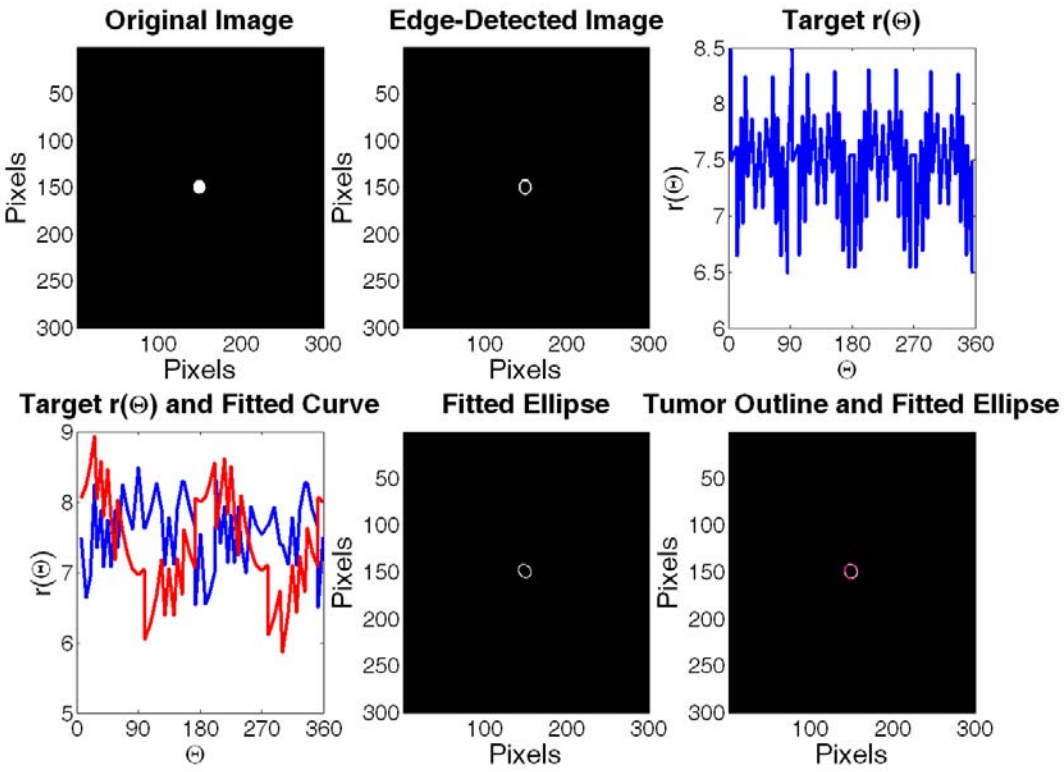
Size examples



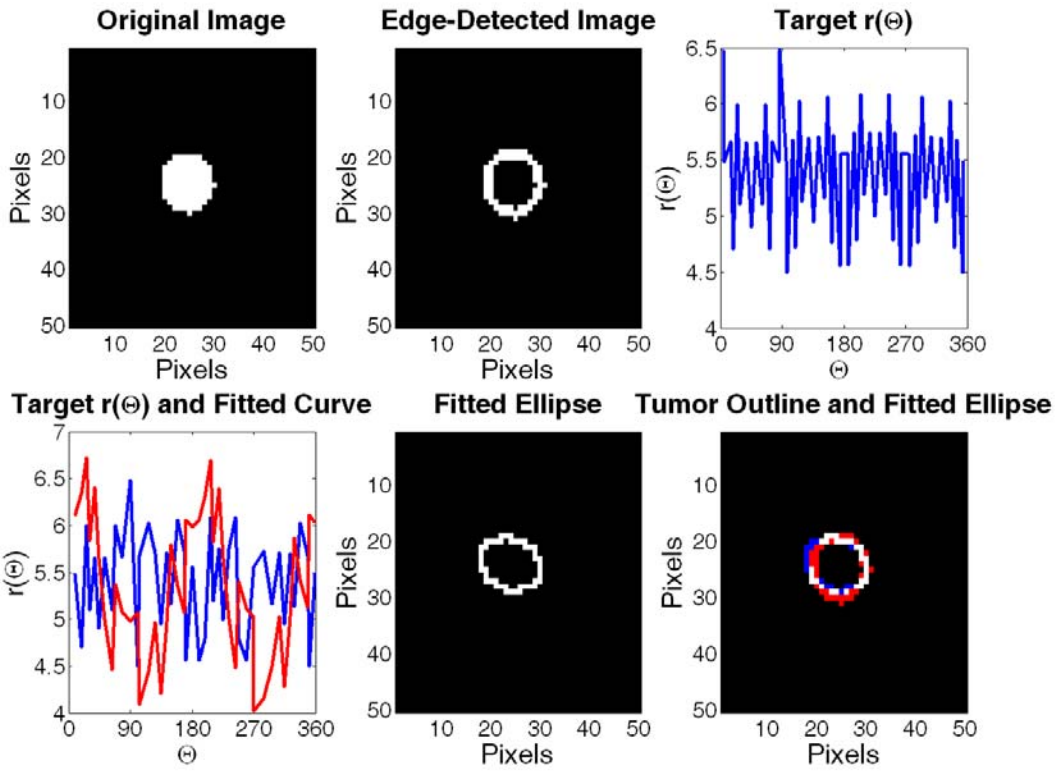
Calculated quantities - FittedEllipse1.tif	
Semi-major axis (pxw)	1.5811
Semi-minor axis (pxw)	0.4243
Angle	71.5651
Area of Fitted Ellipse	2.107444e+000
Area of tumor (px)	3.0000
Ratio of difference in area	0.6505
Eccentricity	0.9633
Shape conformity (pxw)	0.4417
R-squared	0.0270



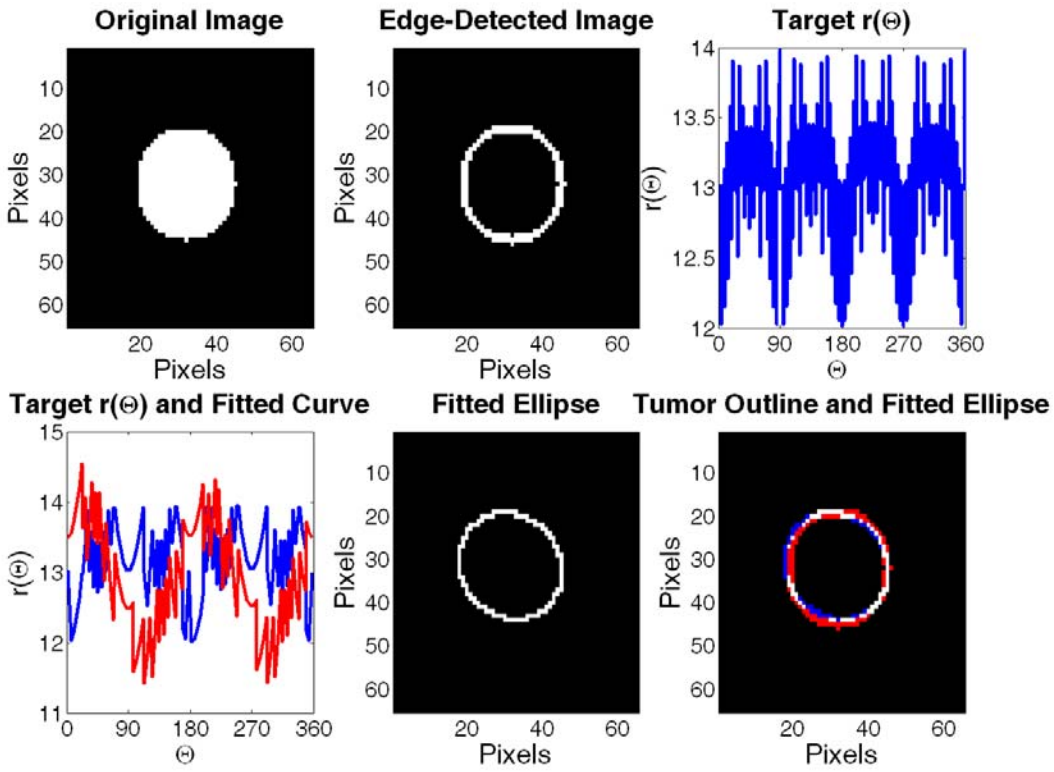
Calculated quantities - ellispe_siz_at_3.tif	
Semi-major axis (pxw)	4.3012
Semi-minor axis (pxw)	2.4963
Angle	144.4623
Area of Fitted Ellipse	3.373096e+001
Area of tumor (px)	38.0000
Ratio of difference in area	0.8810
Eccentricity	0.8144
Shape conformity (pxw)	7.037949e-001
R-squared	-0.0778



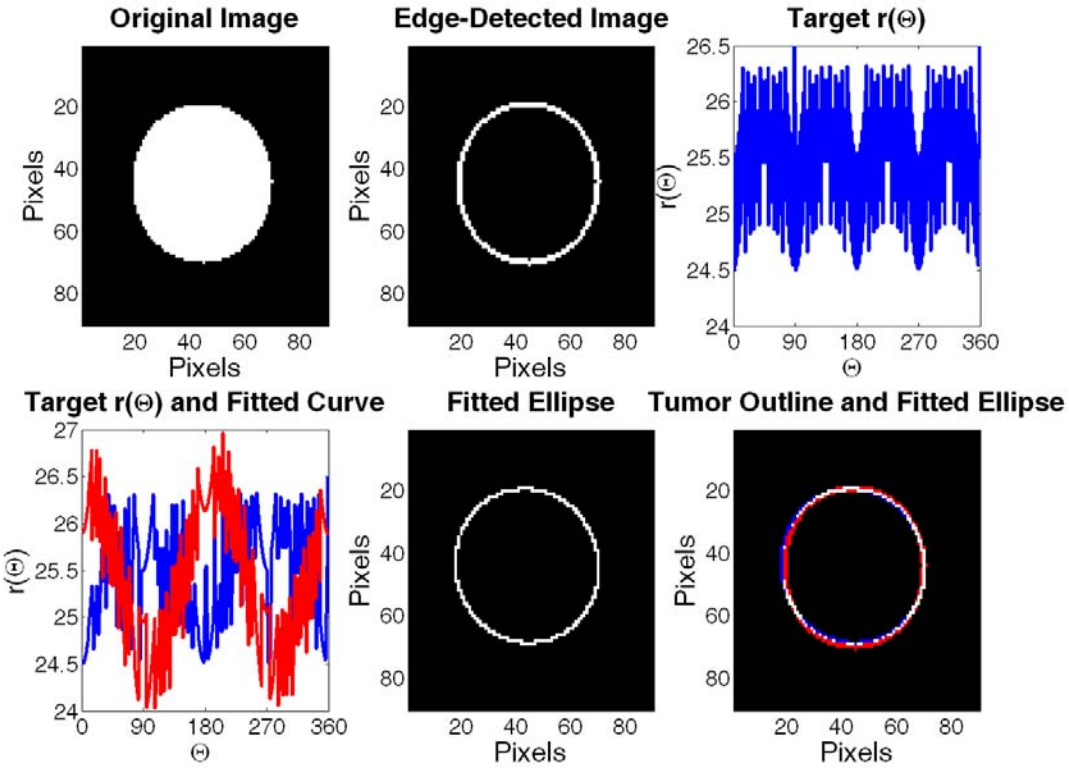
Calculated quantities - `ellispe_siz_at_5.tif`	
Semi-major axis (pxw)	6.5192
Semi-minor axis (pxw)	4.4953
Angle	147.5288
Area of Fitted Ellipse	9.206654e+001
Area of tumor (px)	98.0000
Ratio of difference in area	0.9376
Eccentricity	0.7242
Shape conformity (pxw)	7.635382e-001
R-squared	0.0104



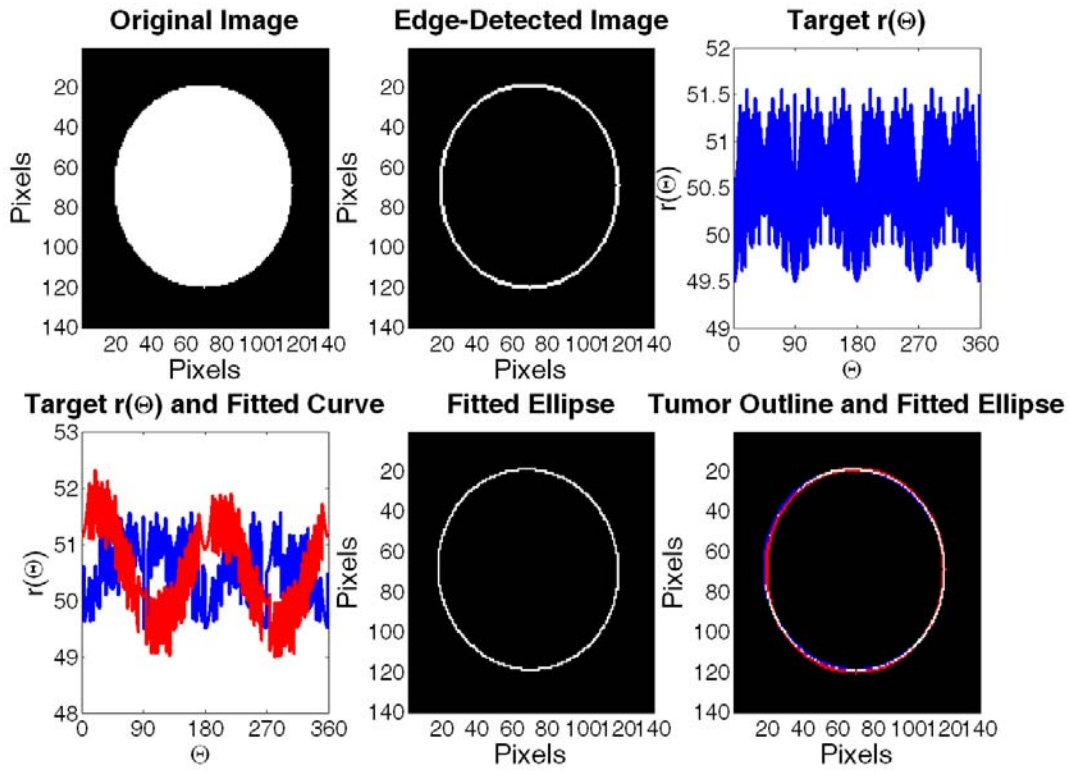
Calculated quantities - FittedEllipse10.tif	
Semi-major axis (pxw)	6.1847
Semi-minor axis (pxw)	4.4984
Angle	165.9638
Area of Fitted Ellipse	8.740236e+001
Area of tumor (px)	90.0000
Ratio of difference in area	0.9707
Eccentricity	0.6863
Shape conformity (pxw)	6.314390e-001
R-squared	-0.1700



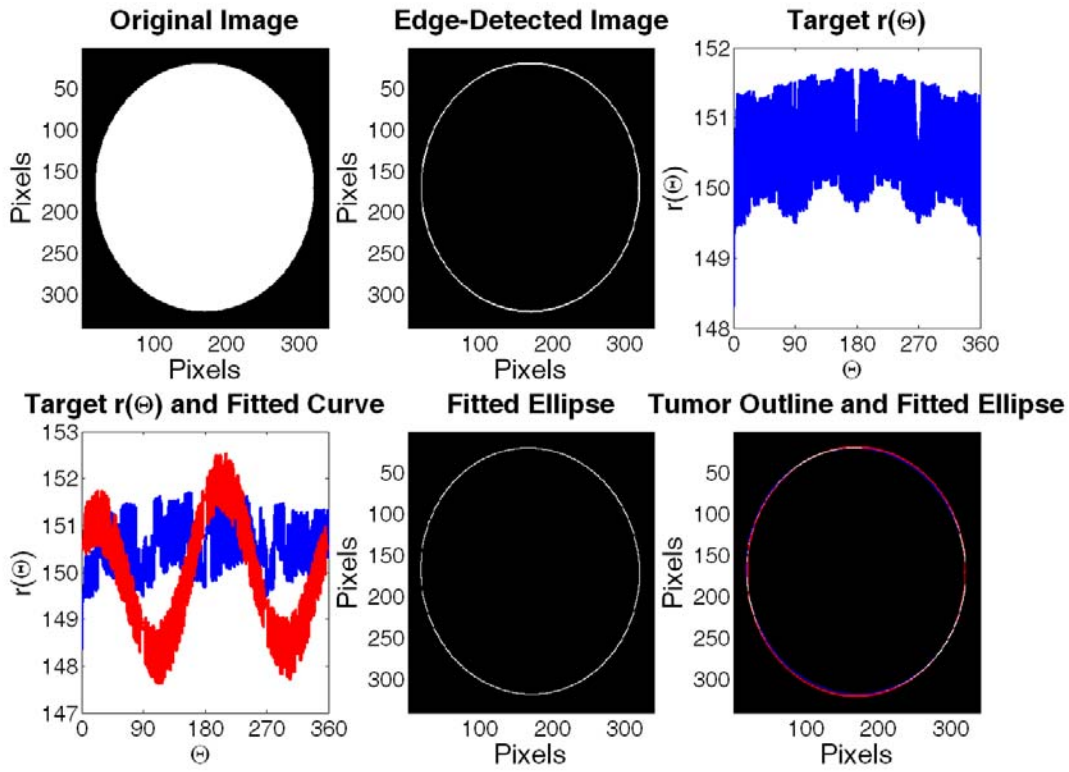
Calculated quantities - `FittedEllipse25.tif`	
Semi-major axis (pxw)	13.9284
Semi-minor axis (pxw)	12.0135
Angle	158.9625
Area of Fitted Ellipse	535.0000
Area of tumor (px)	0.9824
Ratio of difference in area	0.5060
Eccentricity	0.6863
Shape conformity (pxw)	7.374562e-001
R-squared	-0.1978



Calculated quantities - `FittedEllipse50.tif`	
Semi-major axis (pxw)	26.3154
Semi-minor axis (pxw)	24.4981
Angle	165.6997
Area of Fitted Ellipse	2.025314e+003
Area of tumor (px)	2042.0000
Ratio of difference in area	0.9918
Eccentricity	0.3652
Shape conformity (pxw)	7.266679e-001
R-squared	-0.3214



Calculated quantities - `FittedEllipse100.tif`	
Semi-major axis (pxw)	51.5606
Semi-minor axis (pxw)	49.4990
Angle	160.1593
Area of Fitted Ellipse	8.017973e+003
Area of tumor (px)	8026.0000
Ratio of difference in area	0.9990
Eccentricity	0.2799
Shape conformity (pxw)	8.175284e-001
R-squared	-0.193



Calculated quantities – `FittedEllipse300.tif`	
Semi-major axis (pxw)	151.5404
Semi-minor axis (pxw)	148.3266
Angle	156.0567
Area of Fitted Ellipse	7.061506e+004
Area of tumor (px)	71260.0000
Ratio of difference in area	0.9909
Eccentricity	0.2049
Shape conformity (pxw)	1.462284e+000
R-squared	0.0466

APPENDIX C – THE CIRCULAR PERMUTATION MATRIX S

The circular permutation matrix S is a matrix that when multiplied with a matrix M results in a circular permutation of the elements in M. The direction in which the elements in M are shifted depends on whether M is left multiplied or right multiplied by S or its transpose S^T. For example,

MS = shifts elements in M one column to the right.
MS^T = shifts elements in M one column to the left.
SM = shifts elements in M up one row.
$S^T M$ = shifts elements in M down one row.

Elements shifted out of the bounds of the matrix M are re-introduced on the opposite side. In other words elements leaving out right most column of M reappear in the 1st column of M. Elements shifted out of bottom row of M become the top row of $S^T M$.

The size of S must be commensurate with the size of M. Regardless of the size of M, S is always a square matrix of size N with ones along the upper off diagonal and a one in the 1st column on the bottom row of S. All other elements in S are zero. Concisely, the elements of S are,

$s_{i,i+1} = 1$ for i = 1,N-1
$s_{N,1} = 1$

For a square matrix M with dimensions 4×4, S is constructed as,

$$S = \begin{matrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{matrix}$$

, for M 5×5,

$$S = \begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{matrix}$$

, for M 6×6,

$$S = \begin{matrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

...and so forth.

In general, M is an m-by-n matrix. When M has rectangular dimensions two different circular shift matrices, S_m and S_n, of size m and n respectively are required. S_m and S_n are applied as follows,

$MS_n = \text{shifts elements in M right.}$
$MS_n^T = \text{shifts elements in M left.}$
$S_m M = \text{shifts elements in M up.}$
$S_m^T M = \text{shifts elements in M down.}$

, where S_m is a square matrix of size m and S_n is a square matrix of size n.

The matrix S is useful for taking image derivatives.

$$\Delta I_{left} = I - I \cdot S^T$$
$$\Delta I_{right} = I - I \cdot S$$
$$\Delta I_{up} = I - S \cdot I$$
$$\Delta I_{down} = I - S^T \cdot I$$

VITA

Nicholas Jabari Ouma Lee was born on August 2, 1973, in Jackson, Mississippi, to Verna and Nathaniel Lee. Jabari graduated from the Callaway High School and the Academic and Performing Arts Complex in Jackson, Mississippi, in June 1991 and enrolled at Jackson State University where he majored in physics. Jabari was privileged to work under the mentorship of Dr. Kunal Ghosh on describing frequency-dependent behavior of dielectric properties of porous media. He graduated from Jackson State University with a Bachelor of Science degree from the Department of Physics and Atmospheric Sciences and was awarded a fellowship from the David and Lucille Packard Foundation in 1995.

In August 1997, Jabari entered the doctoral program in physics at Louisiana State University where he was awarded a National Science Foundation Graduate Research Traineeship. He joined the Concurrent Computing Laboratory for Materials Simulations at that time. He was awarded the degree of Doctor of Philosophy in physics from Louisiana State University in December 2005. He currently lives in Los Angeles, California. Upon graduation, his plans are to pursue a career in computational science.