HIERARCHICAL SEGMENTATION OF MAMMOGRAMS
BASED ON PIXEL INTENSITY

Martin Masek

This thesis is presented for the degree of Doctor of Philosophy of
The University of Western Australia

Centre for Intelligent Information Processing Systems
School of Electrical, Electronic, and Computer Engineering
The University of Western Australia
Crawley, WA, 6009

June 2003
ABSTRACT

Mammography is currently used to screen women in targeted risk classes for breast cancer. Computer assisted diagnosis of mammograms attempts to lower the workload on radiologists by either automating some of their tasks or acting as a second reader.

The task of mammogram segmentation based on pixel intensity is addressed in this thesis. The mammographic process leads to images where intensity in the image is related to the composition of tissue in the breast; it is therefore possible to segment a mammogram into several regions using a combination of global thresholds, local thresholds and higher-level information based on the intensity histogram. A hierarchical view is taken of the segmentation process, with a series of steps that feed into each other. Methods are presented for segmentation of: 1. image background regions; 2. skin-air interface; 3. pectoral muscle; and 4. segmentation of the database by classification of mammograms into tissue types and determining a similarity measure between mammograms. All methods are automatic.

After a detailed analysis of minimum cross-entropy thresholding, multi-level thresholding is used to segment the main breast tissue from the background. Scanning artefacts and high intensity noise are separated from the breast tissue using binary image operations, rectangular labels are identified from the binary image by their shape, the Radon transform is used to locate the edges of tape artefacts, and a filter is used to locate vertical running roller scratching.

Orientation of the image is determined using the shape of the breast and properties of the breast tissue near the breast edge. Unlike most existing orientation algorithms, which only distinguish between left facing or right facing breasts, the algorithm developed determines orientation for images flipped upside down or rotated onto their side and works successfully on all
images of the testing database. Orientation is an integral part of the segmentation process, as skin-air interface and pectoral muscle extraction rely on it.

A novel way to view the skin-line on the mammogram is as two sets of functions, one set with the x-axis along the rows, and the other with the x-axis along the columns. Using this view, a local thresholding algorithm, and a more sophisticated optimisation based algorithm are presented. Using fitted polynomials along the skin-air interface, the error between polynomial and breast boundary extracted by a threshold is minimised by optimising the threshold and the degree of the polynomial. The final fitted line exhibits the inherent smoothness of the polynomial and provides a more accurate estimate of the skin-line when compared to another established technique.

The edge of the pectoral muscle is a boundary between two relatively homogenous regions. A new algorithm is developed to obtain a threshold to separate adjacent regions distinguishable by intensity. Taking several local windows containing different proportions of the two regions, the threshold is found by examining the behaviour of either the median intensity or a modified cross-entropy intensity as the proportion changes. Image orientation is used to anchor the window corner in the pectoral muscle corner of the image and straight-line fitting is used to generate a more accurate result from the final threshold. An algorithm is also presented to evaluate the accuracy of different pectoral edge estimates.

Identification of the image background and the pectoral muscle allows the breast tissue to be isolated in the mammogram. The density and pattern of the breast tissue is correlated with 1. Breast cancer risk, and 2. Difficulty of reading for the radiologist. Computerised density assessment methods have in the past been feature-based, a number of features extracted from the tissue or its histogram and used as input into a classifier. Here, histogram distance measures have been used to classify mammograms into density types, and
also to order the image database according to image similarity. The advantage of histogram distance measures is that they are less reliant on the accuracy of segmentation and the quality of extracted features, as the whole histogram is used to determine distance, rather than quantifying it into a set of features. Existing histogram distance measures have been applied, and a new histogram distance presented, showing higher accuracy than other such measures, and also better performance than an established feature-based technique.
# TABLE OF CONTENTS

- **ABSTRACT** .................................................................................................................. I
- **TABLE OF CONTENTS** ................................................................................................. V
- **ACKNOWLEDGMENTS** ................................................................................................. IX
- **LIST OF PUBLICATIONS** ........................................................................................... XI
- **1. INTRODUCTION** ....................................................................................................... 1
  - **THESIS OUTLINE** ....................................................................................................... 2
- **2. ANALYSIS OF THE MAMMOGRAM** ....................................................................... 7
  - **BREAST POSITIONING IN MAMMOGRAPHY** ......................................................... 9
  - **BREAST TISSUE TYPES** .......................................................................................... 13
  - **CONCLUSION** ........................................................................................................... 14
- **3. COMPUTER ASSISTED DIAGNOSIS OF MAMMOGRAMS** ............................ 15
  - **INTRODUCTION** ....................................................................................................... 15
  - **ABNORMALITY DETECTION THROUGH MULTIPLE IMAGE COMPARISON** ..... 15
  - **DETECTION AND CLASSIFICATION OF LESIONS AND MICROCALCIFICATIONS** ......................................................................................................................... 17
  - **RISK ASSESSMENT** .................................................................................................... 21
  - **COMPUTER ASSISTED MAMMOGRAM DIAGNOSIS AND THE ROLE OF THIS THESIS** ......................................................................................................................... 22
- **4. NOTATIONAL AND EXPERIMENTAL FRAMEWORK** ............................... 23
  - **MIAS MAMMOGRAM DATABASE** ........................................................................... 23
  - **IMAGE REDUCTION** .................................................................................................. 24
  - **IMAGE DISPLAY** ....................................................................................................... 24
    - **Histogram Equalisation** .......................................................................................... 24
    - **Other Presentation Techniques Used** ..................................................................... 26
  - **COMPUTATIONAL PLATFORM** .................................................................................. 26
- **5. MINIMUM CROSS-ENTROPY THRESHOLDING** ............................................. 27
  - **INTRODUCTION** ....................................................................................................... 27
  - **THRESHOLDING** ......................................................................................................... 28
  - **MINIMUM CROSS-ENTROPY FOR INTENSITY THRESHOLDING** ......................... 29
  - **THE LLBP APPROACH** ............................................................................................ 33
  - **PAL’S APPROACH** ................................................................................................. 37
  - **MODIFIED LLBP APPROACH** ................................................................................ 40
  - **ALGORITHM EVALUATION** ...................................................................................... 43
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic Histogram Evaluation</td>
<td>44</td>
</tr>
<tr>
<td>Varying Gaussian Ratio</td>
<td>44</td>
</tr>
<tr>
<td>Varying Gaussian Mean Spacing</td>
<td>48</td>
</tr>
<tr>
<td>Examining Changes in Standard Deviation</td>
<td>60</td>
</tr>
<tr>
<td>Performance in Terms of Speed</td>
<td>74</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>75</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>76</td>
</tr>
<tr>
<td>6. MULTI LEVEL THRESHOLDING FOR SYSTEMATIC MAMMOGRAM DECOMPOSITION</td>
<td>79</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>79</td>
</tr>
<tr>
<td>METHOD</td>
<td>81</td>
</tr>
<tr>
<td>USAGE GUIDELINES FOR MINIMUM CROSS-ENTROPY MULTI LEVEL THRESHOLDING</td>
<td>84</td>
</tr>
<tr>
<td>Histogram Analysis</td>
<td>85</td>
</tr>
<tr>
<td>INTENSITY BASED SEGMENTATION OF MEDIOLATERAL-OBLIQUE VIEW MAMMOGRAMS</td>
<td>89</td>
</tr>
<tr>
<td>RESULTS</td>
<td>92</td>
</tr>
<tr>
<td>Initial Threshold – Near-skin tissue</td>
<td>92</td>
</tr>
<tr>
<td>Below Initial Threshold – Skin-air interface</td>
<td>94</td>
</tr>
<tr>
<td>Above Initial Threshold – glandular tissue, pectoral muscle</td>
<td>96</td>
</tr>
<tr>
<td>Performance Across the Image Database</td>
<td>99</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>102</td>
</tr>
<tr>
<td>7. SEGMENTING IMAGE BACKGROUND AND DETERMINING ORIENTATION</td>
<td>103</td>
</tr>
<tr>
<td>REMOVING NOISE FROM THE IMAGE BACKGROUND</td>
<td>104</td>
</tr>
<tr>
<td>SIMPLE BACKGROUND NOISE REMOVAL FROM BINARY MASKS</td>
<td>105</td>
</tr>
<tr>
<td>ELIMINATING HIGH INTENSITY REGIONS FROM THE IMAGE BACKGROUND</td>
<td>106</td>
</tr>
<tr>
<td>METHOD</td>
<td>107</td>
</tr>
<tr>
<td>RESULTS</td>
<td>108</td>
</tr>
<tr>
<td>LOCATING RECTANGULAR LABELS</td>
<td>109</td>
</tr>
<tr>
<td>Method</td>
<td>110</td>
</tr>
<tr>
<td>Constraints</td>
<td>110</td>
</tr>
<tr>
<td>Results</td>
<td>112</td>
</tr>
<tr>
<td>Summary of Background Noise Removal</td>
<td>113</td>
</tr>
<tr>
<td>TAPE ARTEFACT LOCATION</td>
<td>114</td>
</tr>
<tr>
<td>Tape Location Results</td>
<td>117</td>
</tr>
<tr>
<td>SCRATCH REMOVAL</td>
<td>117</td>
</tr>
<tr>
<td>Method</td>
<td>118</td>
</tr>
<tr>
<td>Testing Data Set</td>
<td>121</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>122</td>
</tr>
<tr>
<td>Conclusion</td>
<td>123</td>
</tr>
<tr>
<td>MAMMOGRAM ORIENTATION</td>
<td>124</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank all the people without who this thesis would not have been possible. First to my supervisors, Yianni Attikiouzel and Christopher deSilva for providing support and guidance for my research along the way, but also to other members of staff who took their valuable time to offer advice, particularly Ramachandran Chandrasekhar, and Poh Lian Choong.

Next I would like to thank my fellow students, who have provided entertainment and advice throughout my time here, Thomas Hanselmann, James Young, Sze Man Kwok, Brad Finch, Fangwei Zhao, Frederick Chee, Ashbolah Aziz, Tony Jan, Stephen Phillipson, and Teresa Wong.

Finally, and most importantly, I would like to thank my family.
The work presented in this thesis has resulted in the following publications:


Chapter 1

INTRODUCTION

In Australia, breast cancer is the most common cause of death due to cancer in women [1]. Early detection is essential for treatment and many countries use mammography, which is X-ray imaging of the breast, to screen women who are at risk through factors such as family history and age. The role of screening programs is to provide a preliminary diagnosis, with abnormal cases referred for further assessment, for example by biopsy.

As a result of mass screening, a large number of images is generated. The screening program, Breast Screen Australia, exists to target women between 50 and 69 year of age (although the scheme is available for all women over 40) at a recommended screening interval of 24 months [1]. The load on radiologists is expected to increase in the future, both due to the aging population and the increasing proportion of women using the service. In the years 1997 and 1998, 54.3% of women in the target age group were screened. The target of Breast Screen Australia is 70% [1].

With current film-screen mammograms, the film is developed and examined by a radiologist using a light box. The future promises a fully digital, filmless, solution. Several manufacturers are working on systems where mammograms are acquired digitally and displayed on a computer screen for ‘soft copy’ reading. As an indication, the U.S. Food and Drug Administration (FDA) [2] has approved four digital mammography systems from three manufacturers, General Electric Medical Systems (Senographe 2000D approved 28/1/2000) [3], Fischer Imaging (SenoScan approved 19/11/2001) [4], and Lorad (Lorad Digital Breast Imager approved 15/3/2002 [5] and Lorad Hologic Selenia Full Field Digital Mammography System approved 2/10/2002, details not posted at time of writing).

The involvement of a computer in the mammographic process introduces new flexibility into the system, with the possibility of the computer assisting the radiologist in diagnosis through such tasks as determining adequacy of the image or highlighting suspicious regions. Furthermore, while such computerised systems are possible for digital mammography, they
could also be provided as add-ons to legacy film-screen machines where the film is scanned
in and processed through the computer. An example of such a device is the Image Checker
by R2 technology, approved by the U.S. FDA in June 1998 [6]. Here the film is scanned and
suspicious areas highlighted on a computer display for the radiologist to check if any have
been missed [6]. Such systems are also easily integrated into digital mammography systems.
In April 2002 the Image Checker system was given clearance by the US FDA to be integrated
into the General Electric Senographe 2000D [7].

Thesis Outline
Segmentation of the mammogram image into its various distinct regions is an important first
step for further processing to occur. In this thesis, mammogram segmentation is approached
in a number of different ways to separate and identify:

1. Non-breast image regions: image background, labels, scanning artefacts, tape
   artefacts and scratches.

2. The breast: orientation of the breast and both an accurate breast outline and a fast
   segmentation of the ‘interesting’ breast tissue.

3. Pectoral muscle edge.

4. Mammograms similar/different to an example.

5. Mammograms that fall into a distinct tissue density class.

These are presented in the order that they must be performed as each segmentation chapter
uses at least some of the regions extracted in chapters preceding it. The first three points in
the list above refer to segmentation of individual images and are essential for several further
image-processing operations. The fourth and fifth, to keep with the theme of segmentation,
can be thought of as segmentation of an image database. A flowchart of the process is
shown in Figure 1.1.
This thesis is divided into eleven chapters, the first being this introduction. Chapter 2 consists of an overview of mammography and serves to identify some of the terminology used. A brief introduction to the field of computer assisted diagnosis in mammography follows in Chapter 3. This overview serves to highlight the scope and breadth of research topics currently being pursued rather than act as a detailed literature review. More detailed reviews of relevant literature are presented in the chapters that address similar problems. Chapter 4 is a summary of the experimental framework, presenting the database of images.
used as well as image reduction and presentation schemes and the hardware and software tools used to generate the results presented in this thesis.

Chapters 5 to 10 present the main body of work in this thesis and form a progression of segmentation topics.

**Chapter 5 – Minimum Cross-Entropy Thresholding**
The automatic setting of an intensity threshold is a key to many of the segmentation algorithms presented in this thesis. This chapter serves as an overview of the concept of cross-entropy as applied to image thresholding for the case of an image with two regions, or bimodal histogram. Several interpretations of this concept exist and this chapter takes a critical approach to evaluating the different algorithms and resolving some arguments present in the literature. Existing algorithms are reviewed and a new algorithm based on the concept of cross-entropy is proposed. All algorithms are evaluated on a bi-modal image model and their individual properties are noted as the parameters of the model vary.

**Chapter 6 – Multilevel Thresholding for Systematic Mammogram Decomposition**
This chapter extends the concept of minimum cross-entropy thresholding from being used to segment a two-region image to segmenting an image with many intensity regions. The thresholding algorithm is coupled with a systematic scheme for iterating the thresholding process and dividing thresholded images into sub-images for further thresholding applications. The chapter introduces an interactive program for evaluating initial suitability of the algorithm on an image and documents the regions extracted from a mammogram at various steps – the most important are an approximation to the breast boundary and separation of background noise.

**Chapter 7 – Segmenting Image Background and Determining Orientation**
The thresholds obtained from the algorithm in Chapter 6 are combined with properties of the image regions to locate noise in the background of the image, including labels, scanning artefacts and general noise. The orientation of the image is determined using properties of the breast near its edge and the characteristic asymmetry found in the mediolateral oblique view. The orientation algorithm successfully identifies images facing left or right, or images that are upside down or rotated onto their side. Further noise location not dependent on the
previous thresholding results is also performed using a filter used to locate scratches running vertically down the image and the Radon transform is used to locate the edges of tapes.

Chapter 8 – Extracting the Skin-Air Interface
The global thresholding algorithm in Chapter 6 often results in a rough breast boundary due to its changing intensity along the image. This chapter refines the boundary by viewing the breast boundary as a combination of a vertical and horizontal boundary functions. Two algorithms are proposed, one to take local thresholds along the estimated boundary, and the other to fit polynomials along the estimated boundary. The polynomial-fitting algorithm is evaluated and shown to be more accurate when compared with an established algorithm for locating the breast boundary.

Chapter 9 – Expanding Window Thresholding For Delineating the Pectoral Muscle
A novel image processing algorithm is introduced and used to segment the pectoral muscle from the image. The behaviour of the median value or a modified cross-entropy value in a neighbourhood containing two regions changes in a predictable way as different proportions of each region are included in the neighbourhood. This property is used to automatically determine a threshold to separate the pectoral muscle from its neighbouring region and the fitting of a straight line to the pectoral edge is used to minimise remaining noise. It is then shown that the algorithm can also be applied to segment other image regions interactively.

Chapter 10 – Segmenting the Mammogram Database, Classification of Breast Tissue
Segmentation of the database is a term used to describe the classification of its images. The image background and pectoral muscle are removed from the image leaving the breast tissue. The appearance of the normal breast tissue is linked to both risk of developing breast cancer and also the difficulty associated with reading that mammogram. Current methods classify mammograms based on some features extracted from the mammogram into usually 3 or 4 classes associated with risk. The approach taken in this chapter is non-feature based, using the entire breast tissue region and histogram closeness measures as a type of classifier to enable:

1. Classification of images according to their closeness to an average ‘risk class’ image;

2. Ordering of the database images in a continuous spectrum based on distance from some reference image.
The first application has the same function as the existing feature-based mammogram classifiers, the second enables flexibility in computer assisted mammogram screening, allowing the radiologist to retrieve images from a database based on the difficulty of reading.

To perform the histogram-based image comparison existing histogram closeness measures are examined according to their sophistication. A new method is introduced, the closest point distance, that shows superior performance in the case of the mammogram database to both the current histogram closeness measures and feature-based classification of the same mammogram database reported in the literature.

Following these chapters, the thesis is concluded in Chapter 11.
ANALYSIS OF THE MAMMOGRAM

The breast is a sweat gland that has evolved to produce and secrete milk during lactation [8]. Over a lifetime, the female breasts undergo a number of changes which lead to a wide spectrum of appearances on a mammogram. The appearance is affected by factors such as age [9], the menstrual cycle [8], pregnancy [10], and hormone replacement therapy [11].

The breasts can be roughly divided into fatty tissue and glandular tissue, together with parts of the lymphatic system and blood vessels [8]. The glandular tissue is considered ‘active’ tissue [9], being composed of milk-producing lobules [9] along with a ductal system leading from the lobules towards the nipple [8]. The glandular tissue distribution is essentially the same in both breasts [8], although normal variations in breast development can lead to some degree of asymmetry [8], [9]. Young women typically have a large proportion of glandular tissue, this is gradually replaced by fatty tissue as the woman gets older in a process known as involution [9]. It is agreed that involution is related to age, however its relation to menopause is unclear in the literature. Some authors [8], [10] use involution and menopause to mean the same, although Roebuck [9] separates the two as separate processes, differentiating between involution and atrophy due to menopause.

An interesting account of the history of breast cancer is given by Gruhn, Staren, and Wolter [12]. Their book documents milestones in the diagnosis and treatment from one of the oldest known diagnosis and treatment records, written between 3000-2500 B.C., to more recent knowledge of the breast and its abnormalities. Breast abnormalities occur mainly in the active tissue of the breast [8] and can be divided into malignant and benign breast diseases. There is a wide variety of breast diseases and their associated signs, the details of which are described in atlases of breast disease, such as is included in [11], but are outside the scope of this thesis.

Mammography is an accepted and often preferred screening modality to detect abnormalities in the breast [11]. An X-ray passes through the breast, being absorbed selectively by different
tissue types and emerges to be recorded onto a film or plate [9]. Film-Screen mammography is generally the most common form today, where the X-rays strike a screen which emits photons that expose a photographic film [13]. Over the years the radiation dose to the patient has decreased. Radiation exposure is measured either in roentgen (R), or coulomb per kilogram (C/kg), where one $R = 2.58 \times 10^{-4} \text{C/kg}$. The first dedicated mammography unit introduced commercially in 1969 typically delivered an 8-12R patient dose [8]. By 1976 a film-screen system was introduced which lowered this dose to approximately 0.08R [8]. At the same time as radiation dose decreased, image quality increased, through such technology as anti-scatter grids that absorb scattered X-rays and increase contrast of the image [8], improved compression systems and automatic exposure [11].

There are alternatives to the Film-Screen receptor combination. Xeromammography, a now largely obsolete technique, became established when non-screen films were being used [9]. Instead of film, a sheet of aluminium coated with selenium is used [10]. After X-ray exposure, a charge pattern is left on the plate which is transferred onto paper using a coloured powder [10]. The process enhances fine detail, producing an edge enhanced image, although it has poor large area contrast [13]. When introduced, Xeromammography offered a higher quality image with a lower radiation dose [10]. However, due to the development of the anti-scatter grid, and improvements in film-screen technology this is no longer the case, so this along with its higher costs and maintenance have contributed to the Xeromammography process being displaced by Film-Screen mammography [13], [9].

Just as the Film-Screen combination displaced Xeromammography, digital detectors recently on the market have the potential to become the dominant technology due to the advantage of generating digital images. The Full Field Digital Mammography systems work on the same principle as Film-Screen, except the Film-Screen detector is replaced by a digital detector. This is currently similar to the film-screen process, a layer of cesium iodide scintillator used as the screen, coupled to either Charge Coupled Devices (CCDs) in the case of the Lorad and Fischer Imaging systems [14], [15], or a layer of amorphous silicon in the case of the General Electric system [16]. While these may be the way of the future, at the time of writing they are a new technology and used in a small number of facilities [17], although studies have shown no decrease in performance [18]. Since no public databases of Full Field Digital mammogram images are yet available for research, this thesis used digitised images obtained from the film-screen process.
Breast Positioning in Mammography

In film-screen as well as full field digital mammography, the breast is compressed between two parallel plates to spread tissue and make the breast a block of uniform *physical thickness* for the X-rays to pass through. This compression can be performed at different angles to generate different orientations of the breast. The two standard views used in screening mammography are cranio-caudal, generating a top to bottom view of the breast, and a mediolateral oblique, a side-on view at approximately 45%. Examples of each view are shown in Figure 2.1 with the craniocaudal view shown in (a) and mediolateral oblique view in (b).

![Figure 2.1: The two standard views of the breast used in screening mammography. (a) shows the craniocaudal view, from the head down and (b) shows a mediolateral oblique view, with the breast viewed from the side.](image)

The mammogram segmentation algorithms in this thesis have been tested on images from the database of the Mammographic Image Analysis Society (MIAS) [19]. This database consists of only mediolateral oblique view mammograms and so only the appearance of the breast in this view will be discussed. The craniocaudal view image in Figure 2.1(a) was taken
from the Digital Database for Screening Mammography (DDSM) from the University of South Florida.

While the breast is compressed to a uniform physical thickness during mammography, the radiographic density of each tissue type present in the breast determines the appearance of the mammogram. Radiographic density is the term used to describe the level of attenuation that the X-rays experience from the source to the detector. The higher the density in-between, the less developed the film, resulting in appearance from fully exposed (black) to unexposed (white) depending on tissue type.

The fat in the breast has a low density, allowing the X-rays to pass through easily to expose the film – hence fatty areas of the mammogram are dark, in some places almost as dark as portions of the image where there is no tissue (the image background).

The glandular tissue in the breast has a somewhat higher density, resulting in brighter areas, as does the tissue of the pectoral muscle. Some lesions have high-density characteristics, and microcalcifications are very high in density.

As the X-rays pass through a three-dimensional breast to create a two-dimensional image, the brightness of a pixel on the mammogram often represents the superposition of a number of tissue types that the X-ray has passed through on its way. Superposition presents several problems for segmentation of the breast tissue. It also presents problems with diagnosis. Microcalcifications are of much higher density than a comparable volume of glandular tissue. However, when superimposed on a large amount of glandular tissue, the ‘bright’ microcalcification appears less significant. This is the main reason why mammography is better suited to older women with involuted breasts, as the high percentage of active glandular tissue in younger women makes mammography of “limited value and occasionally verges on the useless” [9] – ultrasound being the preferred modality in this case. The exact definition of ‘older’ in this context is however unclear, Roebuck’s assertion that involution can proceed before menopause indicates that mammography may be suitable for some pre-menopausal women, Roebuck giving an example stating a woman may appear radiologically involuted at age of 35 years [9].

Regions present in mammograms of the MIAS database are now identified to give the reader a chance to understand what part of the image is being discussed when certain names are
used. The simplest distinction is to differentiate the image between breast and non-breast regions. The non-breast regions include the image background, labels, scanning artefacts and tapes, which might be superimposed over the breast and are discussed in greater detail in Chapter 7.

While non-breast regions are straightforward to identify, the division between regions inside the breast is less distinct due to the overlap of tissue types, an attempt is made in Figure 2.2. Loosely, the mammogram can be divided into fatty tissue that is dark and glandular tissue that is high in pixel intensity. The pectoral muscle is a characteristic feature of mediolateral oblique view mammograms and presents itself as a bright triangle in one corner of the image. Another feature is the nipple which may not necessarily be seen ‘in profile’ on the image, a close-up of image mdb029 with nipple in profile, and mdb009 with nipple not in profile are shown in Figure 2.3 (a) and (b) respectively. Near the border of the breast lies a region of fatty tissue termed here the near-skin tissue. This region results due to the poor compression of the breast at the edge causing a gradual decrease in thickness towards the skin-line. This region is usually not visible on the original image without enhancement; it is shown in Figure 2.4. The near-skin tissue is used in Chapter 7 to help determine the orientation of the breast in the mammogram image.

![Image Background](a) ![Mainly Fatty Tissue](b)

Figure 2.2: Original image in (a) and its attempted decomposition into separate breast regions in (b).
Lesions and microcalcifications are some of the abnormal regions found on a mammogram. Some examples from the MIAS [19] database are shown in Figure 2.5 with the truthing data superimposed. As the figure shows, there are several different lesion types and lesions can either be malignant or benign. Segmentation of the abnormal tissue types is not undertaken in this thesis.
Breast Tissue Types

The appearance of the breast tissue in a mammogram varies between images. The process of involution leads to the change from a predominantly bright, glandular tissue filled image in younger women to a darker, mostly fatty image, although the process can be reversible through either pregnancy [9], or arrested through hormone replacement therapy [9]. Breasts can be divided into a number of classes based on the appearance of their glandular tissue [20]. The MIAS database separates images into three classes based on density, these will be relevant in Chapter 10, and an example of each of the Fatty, Fatty-Glandular, and Dense-Glandular classes are shown in Figure 2.6.
Conclusion

An introduction to the mammogram has been presented in this Chapter, having in mind a reader with no medical background. The tissue types that make up the breast were introduced, along with their mammographic appearance along with different breast positioning to obtain different views. The wide range of breast appearances discussed in this introductory Chapter serve to stress the problems of developing generic algorithms to process such a range, this will be the topic of the following Chapters.
Chapter 3

COMPUTER ASSISTED DIAGNOSIS OF MAMMOGRAMS

Introduction

A brief overview of computer processing in the analysis of mammograms is presented in this Chapter. This overview serves to highlight the scope and breadth of research topics currently being pursued, and state the uses of mammogram segmentation techniques presented in this thesis in each case, rather than act as a detailed literature review. More detailed reviews of relevant literature are presented in the chapters that address similar problems.

Computer processing in the diagnosis of mammograms has several end goals. It is envisaged that a computer can act as a second reader, either to help the radiologist examine the mammogram by pointing out certain regions, to help classify suspicious regions, or to act as an expert system, analysing risk factors for the patient.

There are several abnormalities that may be present in a mammogram and these are associated with different procedures for the radiologist to follow in their detection and classification. Similarly, the computer assisted diagnosis is often also divided along these lines; there is no universal ‘automatic cancer finder’, rather a collection of methods specifically tuned for different abnormality types, such as asymmetry between left and right breast, changes over time, masses, microcalcifications, and measurement of breast density.

Abnormality detection through multiple image comparison

Computer-based comparison of images has been attempted to automate tasks performed by the radiologist, there are several aims in combining information from more than one image, such as bilateral comparison, temporal comparison and 3D reconstruction. Bilateral comparison is used to detect asymmetry between left and right breast [21] - [32], its presence indicating possible abnormality. Several views of the same breast can be used to relate the mammogram to a 3D position in the breast [33], [34], [35], and developments over time can
be found through temporal examination, comparing mammograms taken at several screenings [36], [37], [38], [25], [39].

In order to compare two images, it is necessary to register them so that equivalent points in each image can be identified. This is usually done by identifying some corresponding regions on the images and finding a transformation from position on one image to position on the other. Several attempts have been made using a rigid registration, where the deformability of the breast is not taken into account. Yin and Giger [21] use the nipple and breast border to determine rotation and translation of one image to match the other, Stamatakis et al. [23] used only the maximum point of curvature of the breast boundary to determine rotation and translation parameters, and Faizon et al. [27] used the nipple to determine only translation parameters. Rigid registration is sometimes used as an initial step. For example, Kok-Wiles et al. [25] and Woo et al. [28] both used rigid transformations based on the breast boundary, followed by non-rigid transformations to match salient parenchymal regions in the case of [25] and image warping to align the breast boundary in the algorithm of [28].

Accounting for the non-rigid behaviour of the breast leads to more accurate registration. The breast undergoes a certain amount of distortion due to compression and the fact that it is a flexible organ. The transformation required for non-rigid registration is not simple, due to the different deformation behaviour of tissue types under compression. It has been noted that denser regions may move more than less dense regions under compression [25], and that the way the breast is positioned affects its deformation [34]. To add to this, registration is complicated through such issues as the use of different compression settings to acquire temporal mammograms, due to changes in the breast [31], and the fact that the left and right breasts are not exactly symmetric in the first place [9].

Several algorithms for non-rigid registration of deformable objects have been applied to mammograms, including thin plate splines [24], [26], [39], curved epipolar lines [34], [35], and the radial basis function [30], [31]. The registration algorithms, both rigid and non-rigid, often rely on reference points taken from a segmentation of the image, such as is undertaken in this thesis. Points along the breast boundary, segmented in Chapter 8, are used extensively [21], [24], [39], [23], [26], [28], [34], [35]. It has been noted that the pectoral muscle position is independent of the breast tissue movement [29] and can lead to lower registration accuracy if used to generate control points [39]. Despite this it has been used to determine registration
control points in another study [24]. Its location is also useful to remove the effect of the pectoral muscle intensities [24], to remove it from the registration completely [26], or to use its straight line approximation as an axis for an anatomy-based coordinate system [29]. The pectoral muscle is segmented in Chapter 9. Location of the nipple (not segmented in this thesis) on the breast boundary seems to be a well-accepted landmark for registration [21], [36], [27], [29], [30], [33], [34], as it is a distinctive feature of the breast boundary.

Also of note is that some algorithms forgo the registration of the mammograms, segmenting similar regions within the breast in each image for comparison. Vujovic et al. [36] use the intersection of elongated structures in the breast tissue as control points, although the nipple is still used as a reference point. Miller et al. [22] segmented non-fat tissue regions in each image and used global comparison with shape, brightness distribution and topology measures to compare regions without the need for image registration. Kok-Wiles et al. [25] used ‘salient parenchymal structures’, found using a saliency metric based on intensity gradients, to match both temporal and bilateral mammograms.

Detection and Classification of Lesions and Microcalcifications

Part of the goal of multi image comparison in the previous section is to detect lesions; lesion detection is however possible from a single mammogram image, as is the detection of microcalcifications – the subjects of a wide volume of publications. Whichever way they are detected, lesions and microcalcifications also need to be classified into their malignant and benign types – a technique many have also tried to automate. Both mass and microcalcification processing tend to follow a similar set of image processing steps. The image is first enhanced to highlight some property of the desired regions – commonly done through either spatial filtering or a time-frequency based method such as wavelet decomposition. Candidate regions are extracted from the enhanced image and some basic classification performed to remove false positives. Further classification may then be performed to classify region types or to separate malignant from non-malignant regions – using such frameworks as neural networks, statistical classifiers, or some sort of decision tree mechanism, taking into account features such as shape, size, and texture properties (such as statistical distribution of regions, and other measures of texture).
As far as the segmentation undertaken in this thesis is concerned, the search for abnormality in a single image generally uses the breast boundary to constrain processing to the breast, avoiding spending time processing non-breast regions, and eliminating false positive findings from those regions. Knowledge of the pectoral muscle may also be used in single image analysis, as it was in the bilateral comparison of Karssemeijer et al. [24], as “breast tumors are frequently projected in the lower areas of the pectoral” [24] and the presence of the intensity gradient at the pectoral muscle edge “may easily generate false alarms or miss tumors” [24], hence location of the pectoral muscle may be used to subtract its contributing intensity from the image.

Many attempts have been made to segment and classify lesions, however they are not discussed in detail here as they are outside the scope of this thesis. Approaches to detection have been based on concepts including template matching, wavelets, and measures of texture. The list of methods used is extensive, only several are summarised here.

Lai et al. [40] used a modified median filter to enhance the image and a template based approach to segment possible circumscribed masses. Several circular template sizes were used to target different mass sizes of a circular shape, with the template set to locate regions based on brightness and uniform density of the tumour, with don’t-care pixels to allow fuzzy edges. Brzakovic et al. [41] used fuzzy pyramid linking to extract suspicious regions, and a hierarchy of Bayes classifiers, using measures based on shape and intensity characteristics to classify different mass types.

Laws’ texture measures [42] have been used quite extensively to segment lesions; Kegelmeyer [43] uses Law’s texture measures together with Analysis of Local Oriented Edges (ALOE) as features to determine a probability of abnormality for each pixel using a binary decision tree. The ALOE is performed as the presence of a lesion disrupts the breast ductal pattern, which normally runs from the nipple towards the chest wall. Polikowski et al. [44] also made use of Law’s texture measures, using a system based on several modules, enhancing masses through a difference of Gaussians filter and extracting Law’s texture measures, along with features based on shape, size, and contrast to classify regions as malignant or benign using a multi-layer perceptron.
It has been reported that success of lesion detection using Law’s texture measures is highly
dependent on pre-processing [45], Pfisterer et al. [45] comparing four texture-based
algorithms. In the comparison, hexagonal wavelets, Law’s texture energy masks, wavelet
enhanced Law’s textures, and steerable filter enhanced Law’s textures were used to locate
masses on mammograms. Pfisterer et al. conclude that the steerable filter enhanced Law’s
method shows the best results for the trial, while their hexagonal wavelet approach showed
the lowest true positive rate.

Microcalcification detection and classification is also a popular topic with many algorithms
existing in the literature. Microcalcifications represent a sharp transition in intensity as they
are quite bright, hence pre-processing to highlight small objects representing high frequency
is used by most approaches. Use of wavelets is a popular method to extract the high
frequency regions of the mammogram in the search for calcifications [47], [46], [48], [49],
others include Markov random field models [51], [50], use of spatial filters [52], [53], such as
box-rim filters [54], or through background subtraction with a model of the background
produced by such methods as Gaussian blurring [49] or polynomial modelling [54], [49].
Kobatake et al. [56] also use subtraction with several smoothed images, each produced with a
top-hat transformation using several structuring elements to increase sensitivity to straight
lines in several orientations to remove false-positive causing ‘elongated shadows’ due to
glands and blood vessels. Another way to remove these false positives is direct segmentation
of curvilinear structures in the mammogram [55]. Other methods to segment potential
microcalcifications include mathematical morphology [57] - [61], directional recursive median
filtering [62], fuzzy logic [63], and fractal theory [64], [65].

Lee at al. [66] compared four methods of microcalcification detection, that of Karssemeijer’s
Markov random field [50], Strickland’s wavelet based algorithm [67], Guillemet et al.’s
greylevel “isophote” contours [68], and Wallet et al.’s adaptive threshold based method [69].
The algorithms were tested on images from two databases, with better results for one
database, indicating a sensitivity to the image source. The criteria for success was detecting at
least 80% of the microcalcification clusters with the least false positive clusters detected. The
algorithms by Wallet and Guillement did not meet the 80% sensitivity requirement,
Karssemeijer’s algorithm being favoured as it produces less false positives than Strickland’s
wavelet based algorithm.
In addition to the properties of individual microcalcifications their grouping is important. A large percentage of mammograms will contain benign microcalcifications [70]. However, the more calcifications per unit area, the more likely the cause is to be malignant [9]. Automated segmentation of microcalcifications can lead to complicated algorithms using more than one of the methods outlined above. For example, Doi et al. [71], [72] follow a series of steps – wavelet transform based processing to enhance microcalcifications, global thresholding, morphological erosion, local thresholding, texture analysis, and clustering. Using a variety of operations leads to a number of parameters whose tuning is problematic due to their interdependence. Adaptive tuning of parameters can in this instance be used to improve the success rate of the algorithm while keeping false positives low. Using a genetic algorithm to set the parameters in the case of the algorithm of Doi et al. led to a sensitivity increase from 80% to 87% with a false positive rate of 1 per image [73].

Classification of extracted microcalcifications can be performed using a variety of classifiers, similar features are generally used by all approaches, such as microcalcification shape, size and properties related to texture or intensity distribution [74], [75], [49], [47], [76], [77]. Features used are similar in characteristic to those used for lesions except that the much smaller area of microcalcifications results in the measures being based on a smaller sample of pixels. Taylor et al. [78] identified several features that have been successfully applied by two or more authors, they are reproduced here:

- Number
- Mean area
- Standard deviation of area
- Mean grey level
- Standard deviation of grey level
- Mean contrast
- Compactness of shape
- Standard deviation of compactness
- Eccentricity of shape
• Standard deviation of eccentricity of shape
• Compactness of cluster shape
• Mean separation of calcifications

Kaufmann et al. [60] defined a cluster as three or more microcalcifications within a 1cm diameter circle, cluster features such as calcification number, cluster area, and statistical measures of inter classification distance, combined with individual calcification features were used to classify clusters into benign and malignant categories, evaluating k-nearest neighbour and Bayesian classifiers. Similar sets of features have been used by others [77]; Fukuoka et al. [79] and Hara et al. [80] use a 3 layer feed forward neural network as the classifier. Bottema et al. [70] determined the convex hull of a microcalcification cluster and used the ratio of the maximum and minimum distances from the convex hull boundary to its geometric centre to separate clusters of large ductal carcinoma in situ related calcification from benign calcifications and artefacts.

**Risk Assessment**

Risk of developing breast cancer has been correlated to such factors as age and family history [9]. The mammographic appearance of a breast has also been correlated with the risk of developing breast cancer. The appearance of the mammographic parenchymal pattern was first associated with breast cancer risk by Wolfe [20], who separated parenchymal patterns into four classes.

Segmentation of the breast, the inner breast tissue, and the pectoral muscle are often pre-requisites for automatic risk assessment, hence the direct relevance of this thesis to the field. Furthermore, Chapter 10 is a direct attempt at classifying mammograms on the basis of their glandular tissue content, providing a method of such risk assessment. The subject of classifying mammograms based on their glandular tissue is therefore dealt with in further detail in that chapter.

Current methods of breast density quantification from mammograms are based on the extraction of several features from the segmented breast, either from the image directly [81], [82], [83], [84], [85], or from its histogram [86]. The features are passed into some kind of classifier, classifying the mammogram into one of several classes (usually 3 or 4). The work
in Chapter 10 takes a different approach, rather than extracting features, closeness measures between breast tissue histograms are applied – allowing classification by determining distance between an image histogram and an ideal histogram for each class.

**Computer Assisted Mammogram Diagnosis and The Role of This Thesis**

The uses of mammogram regions segmented in this thesis have been stated. Regions extracted provide a way to limit a search for abnormalities to the breast tissue, provide reference points for image registration, and to isolate breast tissue for risk-assessment. Besides these, this thesis also introduces some concepts not specific to mammograms, such as the expanding window for determining local thresholds, presented in Chapter 9 to segment the pectoral muscle, the Closest Point Distance for histogram comparison, presented in Chapter 10 for segmenting the image database, and several ways of decomposing an image using global thresholding presented in Chapter 5 and 6.
NOTATIONAL AND EXPERIMENTAL FRAMEWORK

This chapter serves to list the technical parameters of the images and equipment used in the experiments of this thesis, from image acquisition equipment to the manipulation undertaken to make them viewable in print. The computational platform and implementation language is also listed to act as a reference.

MIAS Mammogram Database

The Mammographic Image Analysis Society mammogram database [19] consists of 322 images, representing 161 breast pairs in the mediolateral oblique view. The images originated as the product of the film-screen mammogram process in the United Kingdom National Breast Screening Program. The films have been digitised using a Joyce-Loebl scanning microdensitometer with a 50\(\mu\)m x 50\(\mu\)m per pixel spatial resolution and an 8 bit per pixel greyscale intensity resolution. The microdensitometer converted optical density of the film to greyscale intensity linearly in the optical density range of 0 to 3.2.

The images have been reviewed by a consultant radiologist to identify abnormalities and truth data is available with the database. No abnormality is present in 204 images, abnormalities – both malignant and benign include: Calcifications (25 images), Circumscribed Masses (20 images), Spiculated Masses (21 images), Architectural Distortion (20 images), Asymmetry (17 images, and Miscellaneous – as ill defined masses (15 images). The images are further divided by their tissue density into three classes, Fatty (106 images), Fatty-Glandular (104 images), and Dense-Glandular (112 images).

Images in the database used in this thesis are identified by their file name to enable others to compare results presented here against other algorithms on the same images. The file names begin with the ‘mdb’ prefix, followed by the file number. Images are numbered consecutively from 1 to 322. The images are presented as consecutive left-right pairs of each
patient, so that odd numbers indicate a left breast image and even numbers a right breast image with the right breast following the left breast in order.

**Image Reduction**

The spatial resolution of the images was reduced from the original $50\mu m/pixel$ to $400\mu m/pixel$ using the mean of each non-overlapping 8x8 pixel neighbourhood to create the reduced image; the greyscale intensity range was kept at 8 bits/pixel (256 grey levels).

The reduction in image size as a result of reduced resolution leads to advantages in that the images can be stored and accessed more easily and operations performed on the images are faster. Since this thesis covers the segmentation of relatively large regions of the mammogram, algorithm performance in terms of accuracy is not adversely affected by the reduced resolution.

**Image Display**

Due to the limited page size of this thesis and closeness of greylevels in parts of images, it is difficult to see some of the features of the mammogram if the original image is reproduced in print.

**Histogram Equalisation**

In the majority of images, better visibility of the image for the reader has been achieved through histogram equalisation of the original image. This is a process where *pixel brightness values are re-assigned in an attempt to have the same number of pixels correspond to each brightness value*, resulting in a linear cumulative histogram. This operation is ‘standard’ and covered in many image processing books [87], [88]. A very brief overview is given due to its effect in images with a large dark background such as mammograms.

The basis of histogram equalisation is that peaks in a histogram identify high probability pixels, often corresponding to some region [87]. This is especially the case in the mammogram since tissue type corresponds directly to X-ray attenuation and as a result the greyscale intensity of that tissue type image. Valleys in the histogram are assumed to correspond to less important features or regions. The effect of re-distributing intensities through histogram equalisation is that the peaks are spread out – increasing the contrast of
the regions, and the valleys are compressed as adjacent low probability histogram bins are aggregated to produce a bin with the same probability as the others in the histogram.

Figure 4.1: A characteristic of images with a high proportion of black pixels is that the equalized histogram is shifted right, shown in (a). This is evident in the histogram equalised image, original shown in (b) and histogram equalised in (c), where the background of the mammogram becomes quite light. Anchoring the lowest intensity in the equalised image to 0 produces image (d) showing greater contrast between breast and background.
The problem with the histogram equalisation of mammograms is that the image background of intensities near 0 forms a majority of the image. Since bins cannot be split, the image background intensity is shifted by compressing the histogram towards the bright regions in order to produce a cumulative histogram, as a result the background ends up quite bright; an example is shown in Figure 4.1. A workaround was used in [89], known as zero-anchored histogram equalisation where the lowest intensity in the equalised image is mapped back to intensity 0. This leads to a distinct delineation in many images between the breast and background and highlights any noise present against the image background. There are some cases however where the background intensity in the equalised image is already close to zero and hence little visible distinction exists between the breast and background, one workaround is to map the lowest intensity to the highest possible intensity, or white. This has been used in one case in the chapter on extracting the skin-air interface.

Other Presentation Techniques Used
Another way to represent an image is by re-mapping intensity values randomly into a colour map, this highlights regions which within areas that may seem uniform due to a subtle difference in intensity.

Segmentation results are usually shown superimposed over the original or histogram equalised original to show accuracy. This is done in colour in many instances so the reader finds it easier to differentiate between the mammogram image and the superimposed segment.

Computational Platform
All experiments in this thesis were conducted on a Dell Inspiron 4000 notebook computer with an 800Mhz Intel Pentium 3 Mobile processor, 128MB of PC-100 RAM and a 4200rpm hard disk. Algorithms were implemented in Matlab 5.3 and run under the Microsoft Windows 2000 operating system. This thesis was written in Microsoft Word 2000, with most graphs and images being generated using Matlab 5.3 with additional annotations and graphics added using Microsoft Word 2000 or Microsoft Excel 2000.
MINIMUM CROSS-ENTROPY THRESHOLDING

Introduction
The notion of cross-entropy distance has been applied to intensity thresholding by others using two separate image models, each resulting in several algorithms. Evaluation of the algorithms in the literature has been mainly qualitative and there has been limited comparison of results. Due to optimisations made, one algorithm has been attacked as not representing cross-entropy at all; this claim is investigated and the optimisations are shown not to affect the algorithm’s validity.

Following a brief introduction to thresholding, an analysis of cross-entropy and its related thresholding algorithms is made. The symmetric and non-symmetric cross-entropy forms are evaluated and shown to display similar characteristics. This leads to the conclusion that while non-symmetric cross-entropy is not a true distance, it is faster to compute while giving similar results in the thresholding algorithms as the symmetric cross-entropy.

Several algorithms split the cross-entropy calculation within the image. It is shown that because of intensity scaling and the fact that cross-entropy does not satisfy the triangle inequality, this split will result in different behaviour of the final threshold. A new cross-entropy thresholding algorithm is introduced combining aspects of the existing algorithms.

An analysis of the new and existing thresholding algorithms is made with evaluation based on the thresholding of synthetic bimodal Gaussian histograms. Classification error is computed as a function of the Gaussian parameters for each algorithm to determine image classes where one algorithm outperforms others. The results are validated on real images that fit into each sub-class. It is noted that while the new algorithm combines aspects of existing ones, it is better suited to some classes of image than the existing algorithms.
Thresholding

A greyscale image is a matrix of pixels, each appearing at a brightness from black to white. In the computer, it is convenient to assign a number to each pixel brightness; commonly low numbers are used for low intensity or dark pixels, and high numbers for bright pixels.

An intensity histogram is a graph, showing the number of pixels of each intensity in the image. An example of a simple image and its histogram is shown in Figure 5.1. This image contains five intensities, indicated as 0-4, the numbers superimposed over each pixel.

![Image Histogram](image)

(a) ![Image Histogram](image)

Figure 5.1: An example 5x5 pixel image (a) with pixel intensities assigned numbers 0 – 4. The image histogram is shown in (b).

The goal of intensity thresholding is to determine a brightness that will separate the image into object and background. This is possible if the object pixels belong to a different part of the histogram than the background pixels. The definition of object and background are subjective, depending on which part of the image is of interest. For the example image, if the top two rows are of interest, they are referred to as the object, and are able to be separated from the background by a threshold between one and two if the object is defined as pixels greater than the threshold. At other times it may be useful to define the object as pixels below the threshold.

Once a suitable threshold is found, pixels on either side of it are grouped together as being equivalent; a typical way to represent this is to produce a binary image. Pixels below the
threshold in the original image are set to zero in the binary image and pixels above the
threshold are set to one. Pixels at the threshold may be assigned either above or below.
Defining pixels at the threshold to be assigned to the object, a threshold $T = 1$ will separate
the object from the background in the example image. The result of thresholding is shown
in Figure 5.2 with the binary image produced, shown in (a), and the threshold indicated on
the original image histogram in (b).

There are a number of ways to threshold an image without user intervention; some well
known approaches include algorithms by Otsu [90], Kittler and Illingworth [91] and Kapur
[92]. Survey papers evaluating some thresholding methods include [93], [94], and [95]. One
way to threshold an image is through the use of cross-entropy. This cross-entropy will be the
focus of this chapter and plays a large role in this thesis. The two accepted interpretations of
minimum cross entropy for threshold selection, along with their related algorithms, are
examined and a new algorithm is proposed, combining some of the ideas in each of the
existing methods.

![Image Histogram](image.png)

Figure 5.2: Image from Figure 5.1 thresholded at $T = 1$ and
converted to a binary image in (a). The mapping from original
image intensity to binary image is shown on the histogram in (b).

**Minimum Cross-Entropy For Intensity Thresholding**

Cross-Entropy was formulated as a measure of distance between two probability
distributions. First proposed by Kullback [96], it takes several names in the literature,
including directed divergence and Kullback-Leibler number. For two probability distributions, \( P = \{p_1, \ldots, p_n\} \) and \( Q = \{q_1, \ldots, q_n\} \), where \( \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1 \) (required for a probability distribution), the cross-entropy distance between \( P \) and \( Q \) is given as:

\[
HCE(Q, P) = \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right)
\]  
(1)

For cross-entropy to be a true distance measure in metric space, it needs to satisfy four conditions.

1. \( HCE(Q, P) \geq 0 \)
2. \( HCE(Q, P) = 0 \) if and only if \( Q = P \)
3. \( HCE(Q, P) = HCE(P, Q) \)
4. \( HCE(Q, P) \leq HCE(Q, R) + HCE(R, P) \)

Cross-entropy satisfies the first two properties, however in the form of Equation (1) it violates the third condition, symmetry, as \( HCE(Q, P) \neq HCE(P, Q) \). While it is useful as a distance measure in its non-symmetric form, it can also be symmetrised using the sum of \( HCE(Q, P) \) and \( HCE(P, Q) \) to produce \( HCE_s(Q, P) \).

\[
HCE_s(Q, P) = \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) + \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right)
\]  
(2)

Overlooked by most papers on cross-entropy based thresholding is the fact that cross-entropy does not satisfy the triangle inequality, condition 4, which can be demonstrated using an example.

If three probability distributions are defined as:

\[
P = [\frac{9}{10}, \frac{1}{10}], \quad Q = [\frac{1}{3}, \frac{2}{3}], \quad R = [\frac{5}{8}, \frac{3}{8}]
\]
Then \( HCE_S(Q, P) = 1.6379 \) and \( HCE_S(Q, R) + HCE_S(R, P) = 0.3512 + 0.4638 = 0.8150 \), hence for this case \( HCE_S(Q, P) > HCE_S(Q, R) + HCE_S(R, P) \).

The general concept of cross-entropy has found use in many applications where the data can be modelled as probability distributions. Recent examples of cross-entropy applications include a cross-entropy error function for neural networks [97], monitoring signals in nuclear power plants [98], and modelling the Danish pork industry [99].

Four groups, Li and Lee [100], Brink and Pendock [101], Pal [102] and Xue, Zhang and Lin [103] have used the cross-entropy concept to produce algorithms for the thresholding of images. For the case of images, the probability distributions \( P \) and \( Q \) are determined from the intensities in the original and thresholded image either directly from the pixels, from the histogram, or from a model of the histogram. The threshold is set so as to minimise the cross-entropy distance between the original and thresholded image.

Both Li and Lee, and Brink and Pendock have used the same model to determine a probability distribution from an image, resulting in similar algorithms. Their approach will be referred to as the LLBP approach for short.

There is only one difference between the work of Li and Lee, and that of Brink and Pendock, that is the reversal of the roles of \( P \) and \( Q \). In Brink and Pendock’s algorithm, \( P \) refers to the original image and \( Q \) to the thresholded image. Li and Lee assign the original image probabilities to \( Q \), using \( P \) for the thresholded image. The reversal is not trivial due to the non-symmetry of cross-entropy, but besides this difference identical algorithms are produced, resulting in similar thresholds, as observed by Zimmer et al. [104]. The symmetric form of cross-entropy has also been examined by Brink and Pendock, equivalent to the sum of their non-symmetric cross-entropy and Li and Lee’s reversed non-symmetric cross-entropy.

The LLBP approach will be compared to that of Pal, who uses a different model to derive \( P \), the probability distribution from the original image. Also, rather than determining \( Q \) from the thresholded image Pal models the histogram on either side of the threshold by a Poisson
distribution, using the model histogram to determine the probabilities for $Q$. The approach of Pal was extended by Xue, Zhang and Lin [103] who use the same model to derive $P$ but choose to model the histogram using the uniform distribution. Their algorithm will also be examined.

Of the groups, only Xue et al. mention the failure of cross-entropy to meet the triangle inequality. It is argued that the triangle inequality is not necessary, as only two classes are being considered when computing cross-entropy between the original and thresholded image. While this may be true it also means that cross-entropy is not a true metric distance, which has other repercussions when comparing the algorithms.

The LLBP and Pal approaches, while using different models for the probability distributions also take a different approach as to how the cross-entropy calculation is performed, leading to different results due to the non-metric properties.

The LLBP approach calculates a single cross-entropy distance for each threshold – using all pixels in the original image to determine one probability distribution and all pixels in the thresholded image to determine the other probability distribution. In the Pal approach, the cross-entropy calculation is split into two parts. The cross-entropy of the object (pixels below the threshold) is calculated separately and added to the cross-entropy of the background (pixels above the threshold), determining separate probability distributions for the object and background. In splitting the calculation, the object and background have to be normalised separately so that their respective probabilities sum to one. This introduces a scaling which along with the non-metric properties of cross-entropy means that the result is not the same as when interpreting each entire image as a single probability distribution.

In this light, a new algorithm with both symmetric and non-symmetric versions is introduced, created through the use of the probability model from the LLBP approach, combined with the effect of splitting the cross-entropy calculation, as was done in the Pal approach. The new algorithm is evaluated against the existing approaches to show some of its unique behaviour.
The LLBP approach

Both Brink and Pendock, and Li and Lee use Jayne’s monkey model [105] to obtain probability distributions from both the original and thresholded images. In the monkey model, the intensity of a pixel represents the number of photons striking that pixel, or its ‘photon count’. A probability distribution may be obtained by normalizing the ‘photon count’ of each pixel by the ‘photon count’ of the whole image (or illumination). The set of pixel intensities within the image thus becomes a probability distribution, with the normalized pixel intensities in each image summing to one. One optimisation that both Li and Lee, and Brink and Pendock make is to discard the normalization. Without normalization, the set of pixel intensities do not sum to one and thus do not form a strict probability distribution, this has been argued by Pal in [102] as changing the algorithm from Equation (1) so that the algorithm produced cannot be called minimum cross-entropy thresholding. The normalisation can however be discarded as the normalisation factors reduce to constants due to the way that the thresholded images have been defined by the two groups, now shown.

If the intensity distributions of the original and thresholded, n pixel sized images are \( R = \{r_1, \ldots, r_n\} \) and \( S = \{s_1, \ldots, s_n\} \), where each element \( r \) and \( s \) is the intensity of a single pixel, and the total image illumination (sum of intensities) in the original is \( L_R \), and in the thresholded is \( L_S \), then the normalised probabilities are:

\[
P = \frac{R}{L_R} \quad \text{and} \quad Q = \frac{S}{L_S}.
\]

And Equation (1) becomes

\[
HCE(Q, P) = \sum_{i=1}^{n} \frac{S_i}{L_S} \log \left( \frac{S_i}{L_S} / \frac{R_i}{L_R} \right)
\]  

Rather than assigning the thresholded regions the arbitrary values of 0 and 1, the background and foreground are given the mean intensity values for their respective regions. Defining the thresholded image in this way means that the illumination of the thresholded image will be constant across all thresholds and always equal to the illumination of the original image, therefore \( L_R = L_S = L \) (a constant for the image) and the equation simplifies to:
\[
HCE(Q, P) = \frac{1}{L} \sum_{l=1}^{n} s_l \log(\frac{s_l}{r_{l}})
\]

The effect of normalisation is thus limited to a linear scaling, the resulting minimum cross-entropy distance corresponding to the same intensity threshold as it would if image intensities had not been normalised.

The optimisations made by the groups using the monkey model will now be examined as they are also used in the new algorithm. Taking the example image from Figure 5.1, five thresholds are possible, the resulting images shown in Figure 5.3. Taking the example image for a threshold of \(T=2\), shown in Figure 5.4, and disregarding the normalisation, pixel
Intensities from the original image are the source of the probability distribution $P$, and the thresholded image pixels form the probability distribution $Q$. The cross-entropy between the original and thresholded image can then be determined using Equation (1), with $n$ being the number of image pixels and $p_i$ and $q_i$ corresponding pixels in the original and thresholded images.

$$P = \{0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4, 4, 4\}$$

$$Q = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5, 3.5\}$$

Notice that all pixels in the original image with the same intensity will map to only one of the intensities in the thresholded image. Hence calculations in the sum of Equation (1) are repeated for pixels of identical intensity in the original image. It is necessary only to perform the calculation once for each intensity, multiplying the result by the number of times that intensity occurs. Information on what intensities exist in the original image and the number of times they occur can be supplied to the algorithm from the image histogram.

Figure 5.4: Original image along with its histogram and thresholded image at $T = 2$. Intensities of each pixel in a region are shown superimposed. Above and below threshold means are used for the background and object regions in the thresholded image.
Only two intensities exist in the thresholded image; the intensity means from the original image above and below the threshold. Rather than taking the thresholded image intensities from the actual image or its histogram, they can be computed directly from the original image histogram, saving on the need to produce an actual thresholded image. Since each member of the probability distribution corresponding to the thresholded image will only contain either of the two means, the calculation can be made more efficient by splitting the algorithm into two parts, one for pixels corresponding to the mean below the threshold in the thresholded image, and one for pixels corresponding to the mean above and substituting the means into each of them.

Keeping with the notation introduced in the example from Figure 5.3, the optimised algorithm from Brink and Pendock’s algorithm becomes:

$$HCE(T)_{Brink} = \sum_{j=0}^{T} f_j \mu_y \log(\frac{\mu_y}{j}) + \sum_{j=T+1}^{J} f_j \mu_A \log(\frac{\mu_A}{j})$$  \hspace{1cm} (5)$$

and the optimised algorithm from Li and Lee is:

$$HCE(T)_{Li} = \sum_{j=0}^{T} f_j \log(\frac{j}{\mu_y}) + \sum_{j=T+1}^{J} f_j \log(\frac{j}{\mu_A})$$  \hspace{1cm} (6)$$
where $T$ is the threshold being considered and $j$ an intensity within the image, whose maximum intensity is $J$. Three parameters are taken from the histogram: $\mu_B$, the mean of the image intensities below the threshold, $\mu_A$, the mean of the image intensities above $T$, and the pixel count $f_j$ for pixels $j$. Being histogram-based makes the algorithm fast for large images with a greyscale range lower than the number of pixels.

The symmetric version of the algorithm is the sum of $\text{HCE}(T)_{\text{Brink}}$ and $\text{HCE}(T)_{\text{Li}}$. The optimised version of the algorithm is shown in equation (7).

$$HCE(T) = \sum_{j=0}^{T} f_j (\mu_B \log \left( \frac{f_B}{J} \right) + j \log \left( \frac{j}{\mu_B} \right)) + \sum_{j>T}^{J} f_j (\mu_A \log \left( \frac{f_A}{J} \right) + j \log \left( \frac{j}{\mu_A} \right))$$  

(7)

These simplifications result in an algorithm where the image intensity does not need to be normalised in order to determine $P$ and $Q$, and an algorithm based on the histogram alone.

It is important to note that although it appears that Equations (5), (6) and (7) are calculating the cross-entropy between backgrounds in the original and thresholded images and adding it to the cross-entropy between the objects in the original and thresholded images, this is not the case. Such an operation would require the background and object regions to be normalised by their respective background/object illumination (to form probability distributions), since these illuminations will be different for each threshold normalisation can not be discarded. The algorithm is split into the two sums only to remove direct dependence on the thresholded image histogram.

**Pal’s Approach**

Pal defined the distributions $P$ and $Q$ differently. Also, unlike the LLBP approach, the object and background for each threshold are treated separately, defining the object as pixels below the threshold and the background as pixels above. The total cross-entropy for a threshold $T$ is given by:

$$HCE(T) = HCE(P_O, Q_O) + HCE(P_B, Q_B)$$  

(8)
where \( HCE(P_o, Q_o) \) is the cross-entropy between the P and Q of the object (pixels below threshold T) and \( HCE(P_b, Q_b) \) is the cross-entropy between the P and Q of the background using the same threshold. The symmetric definition from Equation (2) is used to evaluate cross-entropy distance.

In the Pal approach, the original image probability distributions of object and background are defined using the probability of each intensity in the respective region’s histogram. For a threshold T, the object and background distributions, \( P^o = \{p^o_1, \ldots, p^o_T\} \) and \( P^b = \{p^b_{T+1}, \ldots, p^b_J\} \) are determined in terms of the histogram frequencies as:

\[
p_j^o = \frac{f_j}{\sum_{j=1}^T f_j}, \quad 0 \leq j \leq T
\]

\[
p_j^b = \frac{f_j}{\sum_{j=T+1}^J f_j}, \quad T < j \leq J
\]

This means that probabilities are assigned to each intensity rather than to each pixel in the image, as in the LLBP approach. For the example image from Figure 5.3, at threshold \( T = 2 \),

\[ P^o = \{\frac{5}{15}, \frac{5}{15}, \frac{5}{15}\}, \quad P^b = \{\frac{5}{10}, \frac{5}{10}\}. \]

Where each probability term corresponds to pixels of a single intensity (there are 5 pixels of each intensity), 15 is the total pixel count for the object and 10 is the pixel count for the background.

To determine the probability distribution of the thresholded image, \( Q^o = \{q^o_1, \ldots, q^o_n\} \) for the object and \( Q^b = \{q^b_1, \ldots, q^b_m\} \) for the background, a Poisson distribution is fitted to the respective original image histograms using Poisson parameter \( \lambda^o \) for the object and \( \lambda^b \) for
the background. Pal has developed an algorithm for fitting a Poisson distribution to the histogram [106], which defines the Poisson parameters as:

\[
\lambda^O = \frac{\sum_{j=0}^{T} j f_j}{\sum_{j=0}^{T} f_j} \quad (11)
\]

\[
\lambda^B = \frac{\sum_{j=T+1}^{J} j f_j}{\sum_{j=T+1}^{J} f_j} \quad (12)
\]

Using (11) and (12), the distributions \( Q^O \) and \( Q^B \) can obtained by evaluating their Poisson distribution expressions:

\[
q^O_j = \frac{e^{-\lambda^O} \lambda^O_j}{j!}, 0 \leq j \leq T \quad (13)
\]

\[
q^B_j = \frac{e^{-\lambda^B} \lambda^B_j}{j!}, T < j \leq J \quad (14)
\]

Thus, the Pal approach does not rely on producing an actual thresholded image.

The resulting cross-entropy is calculated using:

\[
HCE(T)_{\text{Pal}} = \sum_{j=0}^{T} p^O_j \log\left(\frac{p^O_j}{q^O_j}\right) + \sum_{j=0}^{T} q^O_j \log\left(\frac{q^O_j}{p^O_j}\right) + \sum_{j=T+1}^{J} p^B_j \log\left(\frac{p^B_j}{q^B_j}\right) + \sum_{j=T+1}^{J} q^B_j \log\left(\frac{q^B_j}{p^B_j}\right) \quad (15)
\]

As all probability distributions can be calculated from the histogram data, Equation (15) can be evaluated using statistics from the histogram alone.

A variation on Pal’s approach is made by Xue, Zhang and Lin [103], where the Poisson distributions used to model the object and background are replaced by the uniform distribution.
The definition of the original image probability distribution is the same as used by Pal, using Equations (9) and (10) to determine $P^O$ and $P^B$. In the uniform distribution, all values within a specified interval have equal probability of occurrence. To define the distributions $Q^O$ and $Q^B$, the uniform distribution for the ‘thresholded’ histogram at threshold $T$ is defined as:

$$f_j^T = \begin{cases} 
\sum_{k=0}^{T} f_k \frac{T}{T+1}, & 0 \leq j \leq T \\
\sum_{k=0}^{J-T-1} f_k \frac{J-T-1}{J-T-1}, & T < j \leq J
\end{cases}$$

From (16) the distributions $Q^O = \{q_1^O, ..., q_n^O\}$ and $Q^B = \{q_1^B, ..., q_m^B\}$ are defined as:

$$q_j^O = \frac{f_j}{\sum_{j=1}^{T} f_j^T}, 0 \leq j \leq T$$

$$q_j^B = \frac{f_j}{\sum_{j=T+1}^{J} f_j^T}, T < j \leq J$$

This model, where intensities within each interval from the original histogram are mapped to a single value in the thresholded histogram is closer to the idea of a thresholded image in the LLBP approach where pixels on each side of the threshold are mapped to a single value.

**Modified LLBP approach**

Pal (and Xue, Zhang and Lin) make an interesting point in computing object to object and background to background cross-entropy separately, rather than as one calculation, as in the LLBP approach. The same can be done for the LLBP approach. As dictated by the LLBP approach, $P$ and $Q$ are defined using the monkey model as sets of normalized pixel intensities.
As before, the algorithm to split the calculation is in the form of Equation (8), or using the non-symmetric cross-entropy

\[ HCE(Q, P) = \sum_{i_o=1}^{n} q_{i_o}^{O} \log \left( \frac{q_{i_o}^{O}}{p_{i_o}^{O}} \right) + \sum_{i_b=1}^{m} q_{i_b}^{B} \log \left( \frac{q_{i_b}^{B}}{p_{i_b}^{B}} \right) \]  

(19)

where \( i_o \) refers to pixels in the object and \( i_b \) to pixels in the background. As in Pal’s algorithm, the object is defined as pixels below the threshold. The number of pixels in the object is \( n \) (pixels below and equal to the threshold), and \( m \) is the number of pixels in the background (pixels above the threshold).

Since the illumination of the object and background will be different for each threshold, the normalization, which was discarded when treating the whole image as a single probability distribution, becomes important. For the object in the original image, each pixel is normalised by the object’s illumination, or area under the histogram of the object. The object in the thresholded image consists of pixels of identical intensity. Since these intensities must sum to one, each thresholded image object pixel is simply assigned the value 1/object area, where object area is the number of pixels in the object. The same is done for the background, normalising the original background by its illumination and setting each background pixel in the thresholded image to 1/background area. For the example image at threshold \( T = 2 \), the normalised original and thresholded images are shown in Figure 5.5 along with the histogram. This figure can be compared to that of Figure 5.4 for the LLBP approach cross-entropy algorithm.

Since the normalised thresholded image will contain pixels of value 1/n for all \( q_o \) and 1/m for all \( q_B \), these values can be substituted directly into Equation (19) in a similar way as in the LLBP approach to get:

\[ HCE(Q, P) = \sum_{i=1}^{n} \frac{1}{n} \log \left( \frac{n}{p_{i}^{O}} \right) + \sum_{i=1}^{m} \frac{1}{m} \log \left( \frac{m}{p_{i}^{B}} \right) \]  

(20)

For the example image,
\( \text{Po} = \{0/50,0/50,0/50,0/50,0/50,1/50,1/50,1/50,1/50,1/50,2/50,2/50,2/50,2/50,2/50,2/50\} \),

\( \text{Qo} = \{1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15,1/15\} \)

i.e. \( 1/n = 1/15 \),


\( \text{Qb} = \{1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10,1/10\} \),

i.e. \( 1/m = 1/10 \).

The same process used to simplify the LLBP approach is again used to convert the algorithm to use only the image histogram. Since the terms in the sum will be equal for pixels of identical intensity in the original image, the calculation is performed once and multiplied by the number of pixels of that intensity. As in the LLBP approach, three algorithms can be defined, one symmetric and two non-symmetric, obtained by switching P and Q. The non-symmetric algorithms are given as Equation (21) and Equation (22).

\[
HCE(Q,P) = \sum_{j=0}^{T} f_j \frac{1}{n} \log \left( \frac{1/n}{p_j^o} \right) + \sum_{j=T+1}^{I} f_j \frac{1}{m} \log \left( \frac{1/m}{p_j^b} \right) \tag{21}
\]

\[
HCE(P,Q) = \sum_{j=0}^{T} f_j p_j^o \log \left( \frac{p_j^o}{1/n} \right) + \sum_{j=T+1}^{I} f_j p_j^b \log \left( \frac{p_j^b}{1/m} \right) \tag{22}
\]

The symmetric algorithm is derived by summing the non-symmetric algorithms from Equation (21) and Equation (22) to give:

\[
HCE(P,Q) = \sum_{j=0}^{T} f_j \left( \frac{1}{n} \log \left( \frac{1/n}{p_j^o} \right) + p_j^o \log \left( \frac{p_j^o}{1/n} \right) \right) + \sum_{j=T+1}^{I} f_j \left( \frac{1}{m} \log \left( \frac{1/m}{p_j^b} \right) + p_j^b \log \left( \frac{p_j^b}{1/m} \right) \right) \tag{23}
\]
Algorithm Evaluation

Previous evaluation of the algorithms in the literature has been limited. The algorithm by Pal has been evaluated and compared to that of Li and Lee in [102] on the basis of four synthetic histograms and two sample images. Li and Lee themselves have used four synthetic histograms and one image in [100] to document the evaluation of their algorithm. Xue et al. compare their algorithm against others including Pal using four images but do not show the histograms. The most thorough testing has been presented by Brink and Pendock in [101], where segmentation error is determined as a function of the changing weighting of two synthetic Gaussians of constant standard deviation for several Gaussian means.

Colour coding is used in the evaluation to indicate thresholds from the different algorithms. The coding used is indicated in the key of Figure 5.6 with different heights used to differentiate between algorithms when two or more result in the same threshold.

Thresholding Algorithm Colour Coding

(a) Brink and Pendock
(b) Symmetric Brink and Pendock
(c) Li and Lee
(d) Alternative LLBP1
(e) Alternative LLBP2
(f) Symmetric Alternative LLBP
(g) Pal
(h) (Using dashed line) Xue, Zhang, and Lin

Figure 5.6: Colour coding of algorithms used in subsequent graphs: (a) Brink and Pendock (Equation (5)), (b) Symmetric Brink and Pendock (Equation (7)), (c) Li and Lee (Equation (6)), (d) Alternative LLBP1 (Equation (21)), (e) Alternative LLBP2 (Equation (22)), (f) Symmetric Alternative LLBP (Equation (23)), (g) Pal (Equation (15)), (h) Xue, Zhang and Lin (also Equation (15) with Q redefined using Equation (17) and Equation (18)).

Since all algorithms presented work on data obtainable from histograms alone, segmentation can be tested using synthetic histograms without any actual image. The synthetic histogram evaluation used by Brink and Pendock will be used to compare all the algorithms. The evaluation is extended to cover effects due to changing standard deviation as well as Gaussian weighting in the synthetic histogram. Examples of where a certain algorithm outperforms another are given using real images. Both symmetric and non-symmetric
versions of the LLBP and modified LLBP algorithms and the symmetric Pal and Xue et al.
algorithms are evaluated. The algorithms are also ranked according to their speed.

**Synthetic Histogram Evaluation**

A Gaussian peak in the histogram is used to model each image region, a bimodal Gaussian
histogram modelling the simplest image with uniform background and foreground. Since the
Gaussian parameters of each peak are known, segmentation error can be computed.

For a synthetic bimodal Gaussian histogram, where greylevel is denoted by $g$, the contents of
histogram bin $p_g$ can be defined as:

$$ p_g = p_g^0 + p_g^1 $$

$$ p_g^0 = \frac{P_0}{\sigma_0 \sqrt{2\pi}} e^{-\frac{(g-m_0)^2}{2\sigma_0^2}}, \quad p_g^1 = \frac{P_1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(g-m_1)^2}{2\sigma_1^2}} \quad (24) $$

In (24) $P_0$ and $P_1$ are the relative weights of the Gaussians, $m_0$ and $m_1$ are the means, and the
standard deviation of each Gaussian is represented by $\sigma_0$ and $\sigma_1$.

The error for a specific threshold is computed by calculating the area of the mis-classified
portion of each Gaussian – representing the number of misclassified pixels, although the
total error will be dependent on peak size due to the use of $P_0$ and $P_1$. To compare errors for
different $P_0$ and $P_1$ values, the weightings can be disregarded in the error calculation as
suggested in [107] to get an unbiased error. Assuming that $m_0 \leq m_1$, the total unbiased
classification error is given by:

$$ E(T) = \sum_{g=0}^{T-1} p_g^0 + \sum_{g=0}^{T} p_g^1 \quad (25) $$

Plotting the total unbiased error as a function of $P_0 / P_1$ is a useful performance measure since
it is rare that the object of interest and its background will be the same size.

**Varying Gaussian Ratio**

To emulate the evaluation performed by Brink and Pendock, both peaks in the synthetic
histogram were initially given identical standard deviations, $\sigma_0 = \sigma_1$. To see how the
thresholding algorithms perform when the weighting of the two Gaussians is varied, the ratio $P_0/P_1$ is changed by increasing $P_0$, the proportion of the low intensity Gaussian. Synthetic histogram examples are shown in Figure 5.7(a-c) for a $P_0/P_1$ of 1/1 (a), 10/1 (b) and 100/1 (c). Thresholds selected by the various algorithms when presented with the histograms in Figure 5.7(a-c) are indicated on the histograms in Figure 5.7(d-f).

From Figure 5.7, it can be seen that all the algorithms produce thresholds to separate the two Gaussians when the weighting is 1/1 and 10/1. For the graph in Figure 5.7(c), where the Gaussian weighting is 100:1, only Pal’s algorithm thresholds to separate the two peaks, while the others set their threshold towards the mean of the dominant peak.

To see how algorithm performance changes with the Gaussian ratio, a plot of threshold versus $P_0/P_1$ is shown in the graph of Figure 5.8(a), with resulting unbiased classification.
error shown in Figure 5.8(b). The ratios examined are between 1/1 and 100/1, identical to the range used in [101].

Examining the behaviour of the symmetric and non-symmetric versions of each approach it can be seen that they follow the same trend, the thresholds being only slightly different in each case. This behaviour is observed whether dealing with synthetic histograms or real images. Due to this, only data from symmetric algorithm versions will be shown in detail in subsequent evaluation.
Figure 5.9: Histograms and modified LLBP cross-entropy for each intensity for three cases (a-c). (i) shows the histogram for each case, (ii) shows the object cross-entropy and (iii) shows the background cross-entropy. The total cross-entropy is shown in (iv) with local extrema of interest indicated.
In Figure 5.8, the modified LLBP method shows some interesting and seemingly strange behaviour, with the thresholds first tending towards the less dominant Gaussian and then switching rapidly towards the dominant Gaussian mean when Gaussian Ratio is near 40. This behaviour can be examined by looking at the value of cross-entropy for different thresholds as the Gaussian ratio is increased. Figure 5.9(a-c) shows the modified LLBP cross-entropy for each of the same three Gaussian mixtures as Figure 5.7. For each case, (i) shows the histogram, (ii) shows the cross-entropy between original and thresholded image for pixels above the threshold as a function of the threshold. Figure 5.9(iii) shows the cross-entropy for below threshold pixels. The total cross-entropy, effectively the sum of the graphs in Figure 5.9(ii) and Figure 5.9(iii) is shown in Figure 5.9(iv) for each case (a-c).

The graphs of total cross-entropy show that two local minima have a role in determining the threshold across the range of Gaussian weightings. A minimum occurs at the point between the two Gaussians, $\wedge$, this minimum being most distinct in Figure 5.9(a-iv) when the Gaussians are equally weighted. A second minimum, $\oplus$, occurs in the total cross-entropy when one Gaussian is weighted more than the other. This minimum corresponds to the mean of the dominant Gaussian and grows in prominence as the weighting increases.

The minimum at the dominant Gaussian mean has a local maximum associated with it, $\vert$, on the side closer to the less dominant Gaussian. Figure 5.9(a,b,c-ii) shows that the local maximum arises from the background cross-entropy displacing the minimum cross-entropy from the dominant Gaussian. Thus the threshold initially moves towards the less dominant Gaussian as the weighting increases. When the weighting is large enough that the local minimum associated with the dominant Gaussian mean becomes the global minimum, the threshold is shifted there.

*Varying Gaussian Mean Spacing*

The shape of the graphs in Figure 5.8 depends on the distance of the Gaussian peaks from each other, or the difference between $m_0$ and $m_1$. An important attribute of the algorithms is the result they produce when only one peak is present in the histogram. The threshold produced in this case gives an indication of what happens when the other peaks in the
histogram are not significant enough to affect the algorithm. Figure 5.10(a) shows a one peak Gaussian, with the results of thresholding shown in Figure 5.10(b).

In all three cases the algorithms set the threshold at the Gaussian mean. This is expected as in the case of two Gaussians present, the ideal threshold, $T$, should be set between them:

$$T = m_0 + \frac{(m_1 - m_0)}{2}$$

As the difference between $m_0$ and $m_1$ decreases, the threshold should stay between them until the means are equal and $m_0 = m_1 = T$. In the threshold versus Gaussian ratio graph of Figure 5.8, the threshold is set at the mean of the dominant Gaussian by the LLBP algorithms in a region where the peak dominates, indicating that smaller peaks may be ignored by the algorithms. The Pal and Xue et al. minimum cross-entropy algorithms are affected in a similar way, but only in some cases where the Gaussian peaks overlap.

The behaviour of the error with varying Gaussian ratio can be shown across a range of mean spacings through a 3d surface graph, as shown for the symmetric algorithms in Figure 5.11. The distance between peaks in the graphs of Figure 5.11 is shown for a difference range of $m_1 - m_0 = (0, 1\sigma, 2\sigma, ..., 20\sigma)$. As before, the threshold selected using the weighted histogram is used to calculate the unbiased classification error.
Figure 5.11: Error resulting from the minimum cross-entropy thresholds as functions of both changing Gaussian ratio \((P_0/P_1)\) and peak distance \((m_1-m_0)\) in multiples of \(\sigma\). Graphs are shown for symmetric algorithms (a) Pal, (b) Symmetric Brink and Pendock, (c) Xue et al. (Note axis rotated for clarity) and (d) Symmetric Modified LLBP.

Slices can be taken from the 3d graphs at certain peak spacings to better show the algorithms’ behaviour as the weighting of the lower intensity peak increases. Superimposing slices from each algorithm produces graphs highlighting differences in behaviour in certain regions. Three such graphs are shown in Figure 5.12 along with example histograms to visualise the spread of the peaks.
It is clear from the graphs of Figure 5.11 that the algorithms of Pal and Xu et al. perform the best, maintaining accuracy across a the range of Gaussian ratios and mean spacings. The algorithms in the LLBP approach keep an accurate threshold for a number of weightings after which the threshold tends towards the mean of the more dominant Gaussian. As the distance between the Gaussian means increases, all algorithms improve in accuracy. The error function of Pal’s algorithm reduces to a flat line while the LLBP algorithms produce
accurate thresholds for a larger range of weightings, the gradient of the error function flattening and the point where the dominant Gaussian mean is selected in favour of the correct threshold moving towards the right.

Thresholding results for histograms with overlapping peaks can be demonstrated on the mammogram by looking at the image corner containing the pectoral muscle, shown along with its histogram in Figure 5.13(a).
The high intensity peak of the histogram belongs to the pectoral muscle and the peak corresponding to the darker pixels is from the breast background. Results from thresholding are shown in Figure 5.13(b-e) on the histograms from all algorithms and as binary masks produced from the symmetric algorithm versions. The symmetric Modified LLBP algorithm produces an identical threshold to that of the Pal approach and all algorithms except Xue successfully set the threshold between the two peaks.
Figure 5.14: Histogram of the pectoral corner with the low intensities set to zero (a) in order to show the behaviour of Xue cross-entropy when flat regions are encountered at the ends of the histogram. The object cross-entropy (b) is affected by flat regions at the low intensity end, while the background cross-entropy (c) is affected by flat regions at the high intensity end. While the total cross-entropy (d) has a local minimum between the two histogram peaks, the global minimum corresponds to the flat regions.
The Xue algorithm fails to segment the two peaks due to an inherent limitation arising from use of the uniform distribution. The uniform distribution model results in a histogram composed of two flat regions. For an image whose intensity range does not stretch across the entire intensity spectrum a flat region will exist on one or both ends of the histogram which will be close to the flat region in the thresholded model. Thus Xue threshold is biased towards flat regions at histogram edges that do not necessarily represent boundaries between any peaks. An example is shown in Figure 5.14, again for the mammogram corner. This time the lower end of the histogram has been modified to zero the pixels count for intensities 0 to 50, leading to flat regions at either end of the histogram. A plot of the Xue cross-entropy as a function of threshold is shown in Figure 5.14 as (b) object cross-entropy, (c) background cross-entropy and (d) total cross-entropy. It can be seen that although a local minimum does exist between the two peaks, the flat regions correspond to a lower cross-entropy.

A standard image whose histogram shows several peaks merged together is the Baboon face, shown in Figure 5.15(a) along with its histogram. The histogram is more complex than that of the pectoral muscle corner. It can be separated into three peaks, two prominent ones, separated by a distinct valley, and a low peak with a large standard deviation merging into the prominent peaks. Thresholds chosen by each of the three algorithm groups are shown in Figure 5.15(b-e), along with a binary mask produced by applying the symmetric algorithm from each group.
Visual inspection of the histogram indicates that the valley between the two prominent peaks is the most obvious place to set a threshold, though only the Pal and alternative algorithm do so. The Modified LLBP algorithm sets the threshold in the middle of the valley, while the Pal algorithm moves the threshold towards the brighter peak.

In contrast to the other algorithms, thresholds set by the LLBP approach algorithms are positioned outside the valley, separating the low standard deviation peak from the group of two prominent peaks.

The Baboon is another image where the brightest intensities of the spectrum are not represented in the image. As in the mammogram corner image, this causes the Xue et al. algorithm to give a threshold corresponding to the end of the images intensity range.

Figure 5.15: (a) Baboon Face image with its histogram. Thresholds shown in (b-e) on histogram with binary mask shown for symmetric algorithm results. Thresholding results are shown for (b) Pal algorithm, (c) LLBP algorithms, (d) Modified LLBP algorithms, and (e) Xue et al. algorithm.
With regard to the relative size of histogram peaks in the case of actual images, the performance comparison is subjective, depending on the definition of how high the ratio of a set of peaks must be before one is discarded as insignificant or noise.

The histogram of the ‘bridge’ image, shown in Figure 5.16(a), may be interpreted in a number of ways. The histogram can be interpreted as containing two peaks, the main large peak and the small peak at the upper end of the intensity scale. Comparing this histogram to the synthetic ones used in algorithm evaluation, the Bridge histogram has a high value of Gaussian ratio $P_0/P_1$. Confirming results obtained in the synthetic evaluation, only the Pal algorithm will set a threshold to separate the two peaks, as shown in Figure 5.16(b), while the Xue et al. algorithm is the next closest in Figure 5.16(e). The threshold chosen by the Pal algorithm separates the light planks of the bridge from the rest of the image.

Another way to look at the histogram is to divide the large peak into two sub-peaks that overlap. The dark peak represents the ground and foliage found on the right side of the image, while the brighter peak represents objects mainly found on the left side of the image, like the white water. Results from the LLBP approach algorithms shown in Figure 5.16(c) from the modified LLBP algorithms shown in Figure 5.16(d) show thresholds to separate these two peaks, ignoring the small, but distinct peak that the Pal algorithm isolates.
Figure 5.16: (a) ‘The Bridge’ image has a histogram composed of several peaks, most prominent a large low intensity peak and a smaller high intensity peak. Thresholding results are shown for (b) Pal algorithm, (c) LLBP algorithms, (d) Modified LLBP algorithms, and (e) Xue et al. For each case, a binary thresholded image is shown for the threshold from the symmetric algorithm, while thresholds from symmetric and non-symmetric algorithms are shown on the histogram.

Examining Changes in Standard Deviation

While Brink and Pendock used Gaussians of constant and identical standard deviation to evaluate their algorithm, the spread of peaks present in the histogram of a real image can vary widely. For some images, characteristic differences in peak standard deviation are present. For example, distinct, yet low spread, peaks are present in any image where a significant proportion is occupied by a plain object, such as the classic Cameraman image in Figure 5.17(a). The cameraman’s black coat is present as a sharp peak at low intensity in the histogram. The mammogram also shows similar characteristics, the pixels making up the image background, where no tissue is present, being predominantly black. In the histogram, this translates to a peak of very low standard deviation compared to the other peaks, as
shown in MIAS image14R in Figure 5.17(b). The size of the peak depends on the image, MIAS image 1L in Figure 5.17(c) having an extremely large background peak.

Figure 5.17: Both the cameraman (a) and the mammogram images (MIAS image mdb028 (b), MIAS image mdb001 (c)) show a distinct peak of low standard deviation, caused by a large region of uniform intensity, the peak being extremely prominent in the case of (c).
The effects of varying standard deviation can be shown by using another surface graph, shown in Figure 5.18 and Figure 5.19. The Gaussian ratio axis is the same as before, histograms are obtained by varying, $P_0$, the size of the low intensity peak. The standard deviation analysis is split into two parts. In the first part the standard deviation of the high intensity peak is varied along with the weighting of the low intensity peak (Figure 5.18) and in the second part both the standard deviation and weighting of the low intensity peak is varied (Figure 5.19).

It is easier to interpret the results by examining cross-sections of the graphs. For example, the 2D error graph shown in Figure 5.8 for equal standard deviation of both peaks can be reproduced by taking a cross-section from the middle of the standard deviation axis in either Figure 5.18 or Figure 5.19.
Figure 5.18: The effect on threshold of changing the standard deviation of the high intensity peak while changing the Gaussian ratio by varying the proportion of the low intensity peak.
All algorithms show low error where the ratio of Gaussians is near 1:1 and the standard deviation is the same. Results vary for each algorithm when there is change between the parameters defining the Gaussians.

In the graphs of Figure 5.18 the standard deviation of the high intensity peak is varied. As $P_0$ is larger than $P_1$ in these cases, the threshold in the graphs of Figure 5.18 is not affected as much by the standard deviation change as that of the Figure 5.19 graphs, where the standard deviation of the dominant peak is being modified. This is particularly true for the Symmetric Brink and Pendock algorithm, with cross-sections at various standard deviations being almost identical, looking like the graphs produced in Figure 5.8. The error of the Symmetric Brink and Pendock algorithm increases when the standard deviation of the dominant peak is increased. Figure 5.19(b) shows that in the area where the standard deviation of the

Figure 5.19: The effect on thresholding error of changing the standard deviation of the low intensity peak while also changing its proportion to vary the Gaussian ratio.
dominant low-intensity peak is higher than the other peak, the error increases rapidly as the dominant peak gets larger. Examining the threshold graph in Figure 5.19(b) shows that this is caused by the non-dominant peak being ignored and the threshold being set at the mean of the dominant Gaussian peak. On the other hand, when the standard deviation of the dominant peak is lower than that of the non-dominant, the error performance improves across the whole range of Gaussian ratios.

In the modified LLBP algorithm the characteristic performance seen in the threshold and error graphs of Figure 5.18 and Figure 5.19 remains unchanged. As the Gaussian weighting increases, the threshold moves towards the less dominant peak, up to a point where the threshold changes abruptly to the mean of the dominant peak. What the standard deviation difference affects is the point where the abrupt threshold change occurs, occurring for lower Gaussian ratios as standard deviation of either peak increases. This is an interesting property, since lowering the standard deviation of either the dominant or non-dominant Gaussian leads to better performance in the Modified LLBP algorithm. Figure 5.20 shows two examples of a bimodal Gaussian histogram with different weighting and standard deviation of each peak. The histogram of Figure 5.20(a) has ratio $P_0/P_1 = 50$, with $\sigma_0 = 10$ and $\sigma_1 = 1$. Thresholds from each algorithm are indicated on the histogram, showing that the Modified LLBP algorithm is the only one to separate the two peaks. Good performance of the Modified LLBP algorithm is also shown in Figure 5.20(b), where $P_0/P_1 = 5$ and $\sigma_0 = 1$, $\sigma_1 = 10$. Here the Modified LLBP algorithm outperforms the Pal and Xue et al. algorithms whose performance is affected detrimentally in this case by the standard deviation changes.
For the Pal algorithm, performance when the weighting is changed remains approximately constant for each standard deviation tested. The error however increases as the difference in standard deviation of the two peaks increases, being lowest when it is equal. The error rises rapidly as the standard deviation of the non-dominant peak decreases. This corresponds to the threshold being set at the mean of the dominant peak, meaning the less dominant peak is being ignored. Although the Pal algorithm shows a low error across a wide range of peak ratios and standard deviations, peaks below a certain value of standard deviation may be ignored.

Looking at the Xue et al. algorithm, performance characteristics are similar to Pal’s algorithm, with lower error with changing standard deviation of the non-dominant and similar performance as the dominant peak standard deviation is changed.

Figure 5.20: Examples where Modified LLBP algorithm outperforms others when one peak is of lower standard deviation. (a) Dominant peak is of higher standard deviation, $P_0/P_1 = 50, \sigma_0 = 10, \sigma_1 = 1$. (b) Dominant peak is of lower standard deviation, $P_0/P_1 = 5, \sigma_0 = 1, \sigma_1 = 10$. 
Overall, all algorithms perform adequately on the synthetic histograms within a certain range, although each displays slightly different characteristics. Their behaviour can be further examined by applying the algorithms to the already mentioned cameraman and mammogram images.

For the cameraman, all algorithms except Xue et al. place the threshold in the valley between the low standard deviation peak corresponding to the cameraman’s coat and the group of peaks on the other side of the histogram, shown in Figure 5.21(a). The LLBP approaches set the threshold in the middle of the valley, while both the alternative and the Pal algorithm set the threshold closer to the peak corresponding to the coat.
Figure 5.21: The Cameraman image with all thresholds indicated on histogram (a) (refer to colour coding from Figure 5.6). Thresholds are shown individually for each algorithm group in (b-e) with binary masks shown for symmetric algorithm results for (b) Pal algorithm, (c) Modified LLBP algorithms, (d) LLBP algorithms, and (e) Xue et al. algorithm.

As the valley represents a relatively low number of pixels, the binary masks produced from each threshold look similar. The main difference from the one produced using the
symmetric Brink algorithm in (d) and the lower thresholds in (b) and (c) is the visibility of the camera and details of the cameraman’s face.

Once again the Xue et al. threshold is set where the intensity range of the image begins, this time at the low intensities. While this can be corrected by using only intensities that occur within the image intensity range to determine the threshold it does not remove the problem inherent in the Xue et al. threshold in this case as another relatively flat histogram region exists at the bright end of the histogram. The Xue et al. cross-entropy graph for the cameraman is shown in Figure 5.22, showing that the next lowest minimum is near the intensity 200, the edge of the flat high intensity region.

The histogram of a mammogram is similar in appearance to the Cameraman image, as seen in the images of Figure 5.17(b) (mdb028) and Figure 5.17(c) (mdb001). The difference in these two mammogram images is the size of the background peak in the histogram.

![Figure 5.22: Histogram of 'The Cameraman' image, (a), with Cross-entropy for each threshold in the intensity spectrum (b). The two lowest minima arise from the flat regions at either end of the histogram.](image)
mammogram in Figure 5.17(c) has a more uniform background, leading to a sharp spike in the histogram. The decomposition of the two mammograms is shown in Figure 5.23 for mdb001 and Figure 5.24 for mdb 028. The placement of the thresholds for these two images is similar to the case of the Cameraman, except for the Pal threshold, which ignores the large background peak in the image of Figure 5.17(c), similar to the case of the synthetic histogram in Figure 5.20(b).
Figure 5.23: MIAS image mdb001 with all thresholds indicated on histogram (a). Thresholds are shown individually for each algorithm group in (b-e) with binary masks shown for symmetric algorithm results. Refer to colour coding of Figure 5.6 for algorithm details.
Figure 5.24: MIAS image mdb028 with all thresholds indicated on histogram (a). Thresholds are shown individually for each algorithm group in (b-e) with binary masks shown for symmetric algorithm results. Refer to colour coding of Figure 5.6 for algorithm details.

**Performance in Terms of Speed**

Execution times on a Pentium 3 800MHz processor for an implementation in Matlab 5.3 are given in Table 5.1. The algorithms were run on a synthetic histogram with 256 intensity levels, simulating an 8-bit image.

In all cases, the non-symmetric version of an algorithm is faster than the symmetric version as less computation is required. In terms of speed alone, the Xue et al. algorithm is the fastest, followed by algorithms in the LLBP approach. The Modified LLBP algorithms take longer to run due to the need for normalization of the object and background at each threshold. The Pal algorithm is the slowest due to the time taken by the statistical model fitting to the histogram.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Speed (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li</td>
<td>0.21</td>
</tr>
<tr>
<td>Brink</td>
<td>0.20</td>
</tr>
<tr>
<td>Brink Symmetric</td>
<td>0.32</td>
</tr>
<tr>
<td>Pal.</td>
<td>1.30</td>
</tr>
<tr>
<td>Modified LLBP – Non-Symmetric</td>
<td>0.24</td>
</tr>
<tr>
<td>Modified LLBP – Non-Symmetric</td>
<td>0.25</td>
</tr>
<tr>
<td>Modified LLBP – Symmetric</td>
<td>0.36</td>
</tr>
<tr>
<td>Xue</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 5.1: Executions speeds in seconds of each algorithm on a 256 level histogram.

**Discussion**

Symmetry does not make a large difference to the threshold. For each algorithm, both symmetric and non-symmetric distances followed the same trend as the parameters of the synthetic Gaussian were changed. Using non-symmetric forms of cross-entropy has been in the past argued to be inferior, the argument being that non-symmetry implies that the measure is not a true distance, but as cross-entropy does not obey the triangle inequality, it is not a true distance whether symmetric or not. The advantage of using non-symmetric cross-entropy is speed, as only half the computation of the symmetric distance is needed with a similar final threshold being assigned.

All algorithms perform well when the image histogram is composed of two peaks of approximately the same proportion with equal standard deviation, irrespective of whether they overlap. In this case, the non-symmetric LLBP approach algorithms by Brink and Pendock [101] and Li and Lee [100] are preferred due to their speed.
If the histogram is bimodal, but one peak is of much larger proportion than the other (over approximately 20 times), the algorithm by Pal [102] has a better chance of separating the two, although it is the slowest. The LLBP approach algorithms and the new Modified LLBP algorithms both tend to set the threshold towards the mean of the dominant peak in these cases, so if in fact it is desired to ignore the smaller proportioned peak, these algorithms should be used.

To separate peaks of low standard deviation compared to the rest of the histogram distribution, as arise in many images with a uniform background, the Modified LLBP algorithm threshold is useful. The Modified LLBP algorithm tends to set the threshold closer to the low standard deviation peak than the original LLBP approach algorithms. The Pal and Xue et al. algorithms may ignore the peak altogether if the standard deviation is too low, regardless of its prominence.

The Xue et al. algorithm shows bias towards flat histogram regions due to the uniform distribution used to model the histogram. The problem is that flat regions do not necessarily represent a division between two regions, as was seen when the intensity range in the image does not cover the entire spectrum.

**Conclusion**

A thresholding algorithm based on cross-entropy was introduced and compared to two other families of algorithms based on cross-entropy. Evaluation was performed on synthetic histograms composed of two Gaussians as their parameters were varied. Using synthetic histogram analysis, the characteristics of each algorithm were determined, showing which image types were suited to each algorithm. Results were verified on actual images.

It was found that symmetric and non-symmetric algorithms of the same type produce similar thresholds with identical characteristics as image parameters are varied and so it can be concluded that the more computationally efficient non-symmetric versions can be safely used. Of the algorithms tested, all are suitable on simple bimodal histograms, however for more complex images each algorithm has its advantages. Recommendations of algorithm suitability for different image types were made and presented in the preceding discussion.
The next chapter extends the simple single value thresholding evaluated here to multi-level thresholding for simple mammogram segmentation.
Chapter 6

MULTI LEVEL THRESHOLDING FOR SYSTEMATIC MAMMOGRAM DECOMPOSITION

The algorithms presented in the previous chapter process an image to find a single threshold between two regions. In this chapter, a simple method of guiding the thresholding process to segment multi-region images is presented. The threshold from the minimum cross-entropy thresholding algorithm is used to constrain the histogram for a subsequent application of the algorithm. This technique is used to generate a rough decomposition of mammogram images, dividing the images into skin-air interface, inner tissue, glandular regions and pectoral muscle.

Introduction

Most images contain several regions, resulting in a histogram that is not bimodal. To extract all regions of interest it may be necessary to obtain several thresholds, known as multi level threshold selection, or multithresholding.

On first glance, intensity thresholding using a set of global thresholds would seem to be an ideal solution to mammogram segmentation. Since radiological density corresponds to pixel intensity on the image, distinct regions should exist in the histogram for each tissue type. In practice, superposition of tissue types leads to the histogram regions merging into one another and differences in the thickness of tissue regions result in an increased spread of the peaks. Finally, noise affects the image, leading to problems of distinguishing the breast from the background. Nevertheless, useful results can be obtained from automatic global thresholding. Depending on the application, these are accurate enough to use directly or can be used as a starting point for refinements such as the ones introduced in following chapters.

Since the problem of multi-threshold selection arises in many applications, a variety of algorithms have been developed. Some algorithms are extensions of specific bi-level thresholding algorithms, others offer a general framework for extending some multi-
thresholding algorithm classes. Although a number of survey papers exist on the evaluation of thresholding algorithms, they usually do not concentrate on the multi-thresholding aspect.

Several properties of the intensity histogram can be used to determine multiple thresholds. Boukharouba et al. [108] selected thresholds on the basis of inflection points found in the cumulative histogram. Kohler [109] proposed a thresholding algorithm where the optimisation criterion to determine the threshold is to extract more high contrast edges and fewer low contrast edges than any other threshold, implemented by maximising the average contrast of edges detected. This is extended to multi-thresholding by removing edges detected by previous thresholds and repeating the algorithm. Other multi-threshold selection methods include the use of wavelet analysis of the histogram, by Olivo [110], and algorithms to model the histogram by a Gaussian mixture, such as by Jeng-Horn Chang et al. [111].

Pun [112], Kapur et al. [92] and Otsu [90] among others – have discussed the extension of their bi-level thresholding algorithms to multi-thresholding. Usually such extensions, including those in [112], [92] and [90], require the number of thresholds, or image classes to be known. Extensions have however been proposed by others to determine the number of classes to segment the image into.

Peng-Yeng Yin et al. [114] derive a general algorithm for extending bi-level thresholding algorithms that rely on function optimisation to multi-level thresholding, demonstrating the algorithm using both Otsu’s [90] algorithm and that of Kapur et al. [92]. The number of thresholds is determined iteratively, working from 2-level thresholding to N-level thresholding, choosing the final threshold number based on a uniformity measure of the result. For each level of thresholding, after the thresholds are initialised, they are used to divide the histogram into sections. An updated set of thresholds is found by applying the chosen bi-level thresholding algorithm (Otsu [90], Kapur [92], etc.) to each histogram section. This process of threshold refinement continues until the function of the thresholding method is optimised.

Ohlander et al. [115] used a simple and systematic technique to recursively decompose images into regions. The algorithm first segments the most dominant regions and works recursively to decompose each region into sub-regions. For a given region, a set of histograms is computed. The peaks in each histogram are found and a set of precedence
criteria is used to find the ‘best’ peak out of the peaks from the set of histograms. Setting a
threshold on each side of the peak partitions the region into two sub-regions, one
 corresponding to the histogram peak and the other to the rest of the histogram. These sub-
regions are extracted from the image and each can be further decomposed using the same
algorithm. Use is made of colour as well as greyscale components, separate histograms being
used for each of the components of the RGB, HIS and YIQ colour models. In addition to
the colour based histograms, texture information is used in the form of edges per unit area
and variance to give the algorithm ability to separate texturized and homogenous regions.

The same idea behind Ohlander’s work was used by Wang and Haralick [116] using the edge
pixels of the image to determine histograms. Separate histograms are generated for bright
and dark edge pixels and one of the highest peaks from the two histograms is selected as the
threshold. The threshold splits the image into two regions and the procedure is repeated on
each.

The ideas of Ohlander [115] are used in this chapter to extend bi-level thresholding using a
simple framework, based on the assumption that the region extracted by one threshold may
be further decomposed independent of the outside regions. The technique needs user
guidance to determine the number of thresholds and the regions which should be
decomposed further. Automation is possible for image classes with similar histograms, once
an initial image is segmented with user guidance.

**Method**

The intensity histogram of a mammogram consists of a number of peaks corresponding to
regions in the image. The method presented relies on the assumption that for images with
multiple peaks in the histogram, finding the boundary between two adjacent peaks requires
knowledge of only those peaks. In terms of the image regions, this means that if a region is
to be divided into sub-regions, only pixels within this region need to be considered.

The minimum cross-entropy algorithms can be interpreted as functions that separate the
image into the two most dominant intensity regions, or histogram peaks, the threshold being
found in the valley between them. In actual fact, each of these regions may not be a single
peak but itself be composed of a number of peaks, corresponding to less dominant image
regions. Thus the histogram on one side of the first threshold can be treated as a separate image histogram, again able to be separated into two peaks, similar to the ideas of Ohlander [115], except their peak finding algorithm and precedence heuristic is replaced by minimum cross-entropy threshold selection. Another difference is that rather than extracting a region corresponding to a histogram peak using thresholds on both peak sides, the histogram is instead divided by only one threshold, corresponding to a valley between peaks.

To demonstrate the idea, consider the synthetic histogram in Figure 6.1(a). The histogram consists of two overlapping high intensity peaks and a low intensity peak. A single application of the symmetric Brink Minimum Cross-Entropy separates the low intensity peak from the high intensity peak group, shown in Figure 6.1(b). To find a threshold between the two high intensity peaks, the histogram bin counts below the first threshold are set to zero, eliminating the low intensity peak from the histogram, shown in Figure 6.1(c). Applying the symmetric Brink Minimum Cross-Entropy algorithm on this histogram results in a threshold that separates the two high intensity peaks, shown in Figure 6.1(d).
Figure 6.1: A synthetic histogram with three peaks (a) is segmented in (b) using the symmetric Brink Minimum Cross-Entropy algorithm. If pixel counts below the first threshold are set to zero, only the high intensity peaks remain in the histogram (c), and can be separated with another application of minimum cross-entropy thresholding (d).

Figure 6.2: (a) Two greyscale discs on a black background. (b) Binary image after first application of Brink Minimum Cross-Entropy thresholding separating the two discs from the background. (c) Binary image after applying the Brink Minimum Cross-Entropy thresholding algorithm using only pixels belonging to the two discs to generate the histogram.
The same idea can be demonstrated in the image domain using binary masks. An example is shown in Figure 6.2(a), using a simple image of two overlapping discs on a black background.

This image contains three distinct pixel intensities, meaning that the histogram will contain three sharp peaks. Setting the threshold, this time using the Modified LLBP version of minimum cross-entropy, results in the binary image in Figure 6.2(b), the two disks grouped together as one object. The minimum cross-entropy algorithm can now be run again; this time only considering pixels in the original image which correspond to 'ON' pixels in the mask derived from the first threshold. The resulting threshold segments the smaller, brighter disc as shown in Figure 6.2(c).

By constraining the data presented to the algorithm, each successive application segments less dominant regions. Each application of the thresholding algorithm changes the range of thresholds that the succeeding application of the cross-entropy algorithm will consider. This way the once off application will find a global minimum threshold while successive applications will find local minima in the cross-entropy function, corresponding to boundaries between less dominant regions.

**Usage Guidelines for Minimum Cross-Entropy Multi Level Thresholding**

Multithresholding schemes are usually general, splitting arbitrary images into a pre-defined number of levels, some automatically determining this number. For specific images however, a single region is usually of interest. For example, in some applications the pectoral muscle in the mammogram may be of interest, while in others, the label may be the desired region, hence rather than segmenting into arbitrary regions, the algorithm must be somehow guided towards the threshold that segments the region of interest. This can be achieved on some subsets of images where the histograms are similar in appearance, such as in mediolateral oblique view mammograms. The user must first guide the thresholding algorithm, specifying how many thresholding iterations to perform and how to constrain the histogram before each application. If the thresholding algorithm performs similarly across different images of that type, the same set of steps can then be applied to automatically extract the same region in other images. To find the initial set of thresholding operations, the user must know something about the intensity characteristics of the image.
Histogram Analysis

A rough idea of some initial thresholds can be gained by looking at the image to see if the desired region is separable from its background using intensity. This can be used as an initial test to see whether intensity thresholding should be explored. The accuracy of this approach is limited by the human visual system. Analysis of the histogram provides more accurate knowledge as to the intensity distribution of regions of interest relative to the distribution of the background or noise.

One tool for image histogram analysis is the GNU Image Manipulation Program (GIMP) [117]. This program allows the user to select different portions of the image histogram and displays a corresponding binary mask of the image pixels represented by that portion of the histogram. Another method is to ‘try’ different thresholding algorithms and look at the image and histogram to find an applicable algorithm. Such an image thresholding analysis program has been implemented in Matlab as a graphical user interface front end to the minimum cross-entropy thresholding algorithms, of which a screen shot is shown in Figure 6.3. Here the menu for selecting the operation is shown in one window, and the original image, histogram and binary image shown in separate windows. In the function window, each thresholding algorithm is broken down into three operations. Pressing the large button labelled with the cross-entropy algorithm name runs that algorithm on the entire original image, showing the threshold chosen on the histogram and showing a binary image. The smaller buttons above and below the main thresholding button run the algorithm on a part of the histogram constrained by the threshold set by the previous operation. The button labelled ‘+’ runs the algorithm on pixels above the threshold and the button labelled ‘-’ runs it on pixels below the threshold. The result of running the non-symmetric Brink thresholding algorithm on the Lena image, using the button ‘Brink Cross-Entropy’ is shown in Figure 6.4(a). Pressing the ‘-’ button just below the ‘Brink Cross-Entropy’ button runs the same algorithm on pixels below the threshold (i.e. on the two low intensity peaks), resulting in the threshold shown in Figure 6.4(b). This recursive application is not limited to using the same algorithm for each step, taking the threshold shown in Figure 6.4(b), and pressing the button labelled ‘+’, just above the ‘Pal Cross-Entropy’ button applies the Pal thresholding algorithm to pixels above the current threshold, result shown in Figure 6.4(c).
Figure 6.3: The iterative thresholding evaluator consists of three windows. The function window (a), contains the operations to perform on the image. The operations are performed on the original image, shown in window (b) to produce a binary image, shown in window (c). Depending on the operation, it may be subject to intensity constraints, the intensity histogram shown in window (d).
(a) Result of ‘Brink Cross-Entropy’ button press.
(b) Threshold image below threshold from (a) using Brink Cross-Entropy.
Figure 6.4: An example of the threshold selection process. In (a) the non-symmetric Brink thresholding algorithm has been run on the original image, producing the binary mask and threshold shown. In (b) the algorithm is run again, this time on pixels below the first threshold to produce another threshold separating the two lowest intensity peaks. In (c), the Pal thresholding algorithm is run using pixels above the second threshold, segmenting the brightest peak from the histogram.

Intensity Based Segmentation of Mediolateral-Oblique View Mammograms

Histograms of mammograms in the mediolateral oblique view show peaks in similar positions. Techniques developed on one image using properties common to the class can be applied to other images of that class without further user intervention. The variance of the mediolateral oblique view image class however is a cause for concern. It has been stated in Chapter 2 that mammograms fall under a set of types, related to the level of glandular tissue
present. It will be shown that even though this variance exists, useful information can be extracted from mammograms using an automated thresholding algorithm.

The mammogram can be divided in terms of intensity into four tissue types. For the mammogram of Figure 6.5(a), MIAS image mdb007, region outlines set by global thresholds are shown in Figure 6.5(b). The positions of the regions are indicated on the image histogram in Figure 6.5(c). The regions are separated in the image but overlap in the histogram; they correspond to (i) the pectoral muscle, (ii) fatty breast tissue, (iii) denser glandular tissue and (iv) the image background.

Example mammograms from each of the three breast types present in the MIAS database are shown in Figure 6.6 along with their histograms. Although the peaks corresponding to the various tissue types vary in appearance across the images, they are located in similar positions relative to each other on the histogram.
Depending on the images being segmented, an appropriate thresholding algorithm must be chosen to maintain consistent performance across the range of images. It was demonstrated in the previous chapter that the Pal minimum cross-entropy sets the initial threshold
dependent on the standard deviation of the histogram peak corresponding to the background. An example was given where the Pal algorithm was shown to delineate the skin-air interface for MIAS image mdb028, shown in Figure 5.24(d) and the brighter parts of the pectoral muscle and glandular tissue for MIAS image mdb001, shown in Figure 5.23(d). This would result in non-uniform performance across a series of mammogram images. The Xue et al. algorithm is also unsuitable due to the bias to flat regions of the histogram. This leaves a choice between the algorithms of the LLBP approach and the algorithms of the Modified LLBP approach.

The Modified LLBP approach algorithms set an initial threshold which segments more of the breast than the LLBP algorithms, producing an outline closer to the actual skin-air interface, as was shown in Chapter 5. Although this is the case, the skin-air interface can still be extracted using the recursive multi-level thresholding method and the advantage is that combining the rough breast outline with the skin-air interface outline extracts the tissue near the skin, a region that is also useful. A further advantage of the LLBP algorithms over the Modified LLBP is speed.

From the findings of Chapter 5, the non-symmetric cross-entropy is faster to compute while offering a similar threshold as the symmetric form, the non-symmetric minimum cross-entropy thresholding algorithm by Brink and Pendock [101] is therefore used here to decompose the mammogram histogram.

**Results**

Results from the decomposition of the three examples in Figure 6.6 are shown for each successive threshold in Figure 6.7–Figure 6.10, indicating the threshold on the histogram and showing the binary mask produced from the threshold. The outline of each binary mask is shown superimposed over the original images in Figure 6.11 to show the correspondence of the thresholds to the image regions.

**Initial Threshold – Near-skin tissue**

The first threshold is determined using the whole image. This region contains tissue near the skin-air interface, where the breast is poorly compressed. Thresholds and binary masks for the three example images are shown in Figure 6.7 and the outline of the binary mask is
shown as the red line in Figure 6.11. The threshold is placed near the bottom of the large valley in the low intensity part of the histogram. The result is a rough outline of the breast, with tissue near the skin-air interface being cut off.

Figure 6.7: The initial threshold corresponds to the majority of the breast tissue, it is shown superimposed over the histogram along with the binary mask resulting from setting pixels higher than the threshold to 1.
Below Initial Threshold – Skin-air interface

The histogram below the first threshold contains pixels belonging either to the image background or the tissue near the skin cut off in the first binary mask. Running the thresholding algorithm on this part of the histogram results in a threshold which separates the two, shown in Figure 6.8 for the examples. It can be seen from the histogram and also the binary masks that a clean delineation of the skin-air interface using a global threshold, if possible, is not easy since the background and breast regions of the histogram overlap. The division between breast and background is hard to see in the histogram of the image in Figure 6.8(a) but dips in the histogram are found in the case of examples from Figure 6.8(b) and Figure 6.8(c). Intensity differences of the skin-air interface across the image result in a smooth outline in parts of the image with a rougher outline in others, best seen from the green line in Figure 6.11 which corresponds to the outline of the binary mask. The skin-air interface extracted in this way can however be used as a first approximation and refined using local thresholding combined with knowledge of the skin-air interface shape. Such refinements are discussed in Chapter 8.

The region of the breast between the intensities of the first and second threshold is also useful. This area of poorly compressed tissue has found a number of uses in the past. These will be discussed further in Chapter 7 which uses characteristics of this region to determine orientation of the breast in the mediolateral oblique view mammogram.

Another image region extracted along with both the inner tissue border and the skin-air interface are bright labels and any scanning artefacts which appear in both binary masks. Their appearance in both masks can be used to differentiate them from the rest of the image, Chapter 7 presents algorithms to extract them and differentiate the scanning artefacts from labels using their shape.
Figure 6.8: Skin-Air interface threshold, obtained by thresholding below the initial threshold. The division between background and breast is not clear on the histogram of the glandular example (a), but is easier to see for the dense image (b) and the fatty image (c) as a dip in the histogram.
Above Initial Threshold – glandular tissue, pectoral muscle

Looking at intensities above the first inner tissue threshold, the histogram is composed of peaks belonging to the glandular tissue and pectoral muscle. The results of thresholding using this part of the histogram are shown in Figure 6.9, the outline of the binary mask shown as the blue line in Figure 6.11. The outline from this threshold is not as reliable as the skin-air interface. In some images a clear outline of the pectoral muscle and glandular tissue is shown, in others the threshold is too low to differentiate the pectoral muscle from the glandular tissue as in Figure 6.9(c). Taking a further threshold on the histogram to the right of this threshold gives the results shown in Figure 6.10, with the yellow line in Figure 6.11 corresponding to the outline. For images where the pectoral muscle was not previously delineated, this threshold generally shows its outline. For images where a pectoral boundary was visible from the last threshold, the brighter parts of the pectoral muscle and glandular tissue remain. It is interesting to note that while the pectoral muscle edge from this threshold may be internal to the actual pectoral edge, the shape remains similar while the outline is less affected by overlapping glandular tissue. Thus if the angle of the pectoral muscle is of interest, the inner border is a more reliable source of information, a finding that is used in Chapter 9 for more accurate approximation of the pectoral muscle boundary.
Figure 6.9: Threshold above the initial. Depending on the size of the peaks in the image this either delineates the pectoral muscle and glandular tissue or extracts the brighter part of the breast. Note: Figure 6.12 shows that this is not dependent on image type.
Figure 6.10: The second threshold above the initial. This either extracts the brighter parts of the pectoral muscle and glandular tissue as shown in (a) and (b) and this typically delineates the pectoral muscle if it was not extracted in the first threshold above the initial, shown in (c).
Performance Across the Image Database

The recursive decomposition was tested on the MIAS database to measure deviations in performance. Three examples of each type are shown in Figure 6.12 with outlines of the binary thresholding masks superimposed. The same colour coding is used as in Figure 6.11, with the red line to indicate the first threshold, at the near-skin tissue, and the green line indicating the mask from the skin-air interface threshold, derived from the histogram region below the first threshold. The blue and yellow lines are the two glandular tissue and pectoral muscle thresholds, with blue being the first threshold set using the histogram on the right side of the near-skin tissue threshold and the yellow line corresponding to the threshold from the histogram section on the right of the threshold corresponding to the blue line.
Figure 6.12: Binary mask outlines from each threshold superimposed over the original image. Three images from each of the three types are shown, with glandular mammograms in (a) mdb031, (b) mdb042, and (c) mdb083, dense mammograms in (d) mdb063, (e) mdb068, and (f) mdb126, and fatty mammograms in (g) mdb011, (h) mdb038, and (i) mdb134.
The skin-air outline from the global threshold is generally good for an approximation of the division between breast tissue and background. For applications where the region of interest is the glandular tissue, this method is suitable for limiting the examinable area as the glandular tissue is not cut off by this threshold.

The skin outline tends to be jagged in parts of the image, caused by the outline being placed slightly internal to the breast, in the fatty tissue near the skin. This is caused by variations in intensity of tissue near the skin across the image and is not easy to avoid when using a global threshold. Another form of noise is scratches which run down some of the images, an example of which can be seen in Figure 6.12(b). These scratches are sometimes extracted along with the skin-air interface, as seen in the figure. Chapter 8 looks at a combination of local and global thresholding to segment a smoother skin-air interface, while Chapter 7 includes a section on pre-processing to remove scratches from the image.

The outline from the initial threshold, in the near-skin tissue, roughly follows the skin-air interface and also contains the glandular tissue within it. As this region is internal to the real breast boundary it is separated from the noise in the image background. Performance of the global multi thresholding algorithms is consistent enough that this property is used in Chapter 7 to eliminate labels and scanning artefacts from the images. The area of near-skin tissue, between the initial and skin-air thresholds, while having several uses, is used in Chapter 7 to determine image orientation, succeeding in all images from the MIAS database.

The extraction of the pectoral muscle using the global threshold is less reliable than that of the skin-air interface and the near-skin tissue. The threshold for the pectoral muscle is affected by differing proportions of glandular tissue in the breast (glandular tissue peak size) and the relative density of the glandular tissue (peak position in the histogram). As a result, the pectoral muscle boundary is extracted by either the first, second, and in some cases third threshold from the inside of the near-skin tissue boundary. While it is possible to detect which of the binary images corresponds to the pectoral boundary by looking for the characteristic straight line results are not accurate in all cases as both thresholds may result in a straight line near the pectoral muscle (Figure 6.12(e,f,h)). Chapter 9 examines this issue and presents a new way to extract the pectoral muscle using a local threshold.
Conclusion

It has been shown that cross-entropy thresholding can be guided to segment an image that is composed of several regions. Extraction of each region is performed as a sequence of thresholding steps. Generally the minimum cross-entropy threshold is set in a valley between two peaks on the histogram. After each thresholding step, the histogram on each side of the threshold can be treated separately, used in the next thresholding step. This way, the thresholding algorithm first separates the two most dominant regions in the image, after which each region is decomposed into its sub-regions.

The technique was demonstrated on several mammograms, extracting the regions from the mammogram, such as the breast region, near-skin tissue, glandular tissue, and pectoral muscle using a non-symmetric minimum cross-entropy algorithm by Brink and Pendock. The evaluation demonstrated on the MIAS database showed that for images of a certain class, where histogram regions are similar between images, the thresholding can be automated, as demonstrated on the skin-air interface and near-skin tissue. It was seen with the pectoral muscle that as differences between image features increase across an image class, the automated multi thresholding becomes less reliable.
SEGMENTING IMAGE BACKGROUND AND DETERMINING ORIENTATION

A rough segmentation of the mammogram was achieved in the last chapter using iterative minimum cross-entropy, resulting in a set of binary masks. In this chapter, further refinement of the regions in the masks is undertaken, segmenting the image background into labels, scanning artefacts, tapes, and scratches, and using properties of the image to determine breast orientation.

High intensity objects outside the breast area, such as scanning artefacts and labels are differentiated from breast tissue using object-based logic operators. An algorithm is presented which differentiates rectangular labels from other high intensity image background regions based on their shape and the Radon transform is used to locate the edges of tape artefacts.

One undesirable image feature is scratches present in some images due to the action of the rollers. Scratches have the potential to affect the skin-line estimation and neighbourhood based image functions. A difference operator is used to find the image columns affected by scratches. Their presence can either be flagged, or they can be removed by the interpolation of surrounding pixels to replace the scratch pixels.

Once noise in the image background is located, the orientation that the image has been scanned in is determined. Knowing breast orientation in the image simplifies further segmentation steps, such as location of the skin-air interface and the pectoral muscle. Orientation is determined using a two-stage process. First, the thickening of tissue near the skin is used to locate the chest wall. Second, in the case of mediolateral oblique view mammograms, the characteristic asymmetry of the breast is used to locate the image top. The results of this application are significant, as most previous algorithms concentrate on determining the chest wall while assuming that the image is not upside down or rotated onto its side. This algorithm makes no such assumption and successfully determines orientation on every image in the MIAS database.
**Removing Noise from the Image Background**

Ideally, the image background of the mammogram, i.e. the pixels that do not correspond to the breast, would be of intensity zero. In practice, a number of regions may exist in the image background where this is not the case. These regions can be regarded as noise, although some have useful applications. The noise regions can be loosely classified into high intensity noise and low intensity noise. Low intensity noise is that with pixel brightness near the level of the skin-air interface, usually from the imaging process. Since this noise is near the skin-air interface its removal is more complicated, hence it will not be discussed here but treated fully in Chapter 8 on extracting the skin-air interface. High intensity noise is made up of regions whose brightness is near that of the glandular tissue and pectoral muscle, such as labels (regions corresponding to lead markers) and scanning artefacts (corresponding to an image of the scanner light source). In-between the high and low intensity noise are regions corresponding to tapes that have been placed on the image, and also some lower intensity labels, which may have high intensity components. Examples of these regions are shown in Figure 7.1 on MIAS image mdb274.

![Figure 7.1: Several types of noise present on the image background, example using MIAS image mdb274.](image-url)
A simple method to remove noise from a binary mask of the breast is first presented. While this method removes most of the noise, it also removes any parts of the breast that are fragmented from the main breast region. To combat this, a variation of the algorithm is presented which targets only high intensity noise, namely the scanning artefacts and labels. Thirdly, an algorithm is presented which differentiates high intensity rectangular labels from the other high intensity noise, the algorithms also being described in [118].

**Simple Background Noise Removal From Binary Masks**

A simple way to eliminate most of the noise from a binary mask is to find the largest object in the image and assume it is the breast. The algorithm is implemented to work on the binary mask, example shown in Figure 7.2(a) for image mdb274. Individual white pixel regions are labelled, shown using an arbitrary colour coding in Figure 7.2(b). The number of pixels within each region is counted and the region with the largest pixel count is removed, leaving the noise, shown in Figure 7.2(c).

![Figure 7.2: (a) Binary near-skin mask of image mdb274. (b) Regions of white pixels are labelled, shown using a random colour coding. The number of pixels in each region is determined and the largest region assumed to be the breast, leaving the others as noise (c).](image)

This technique is effective for constraining the area of interest to one region within the mask and is useful for algorithms that do not depend on any great accuracy of breast tissue, like the orientation algorithm described later in this chapter.
Eliminating High Intensity Regions from the Image Background

The simple noise removal algorithm does not make any attempt to differentiate the regions of noise from parts of the breast that have been separated from the main breast region by the threshold. By limiting the focus to the removal of high intensity noise it is easier to preserve regions belonging to the breast. The trade off is that some of the medium intensity noise, such as tapes and lower intensity labels, will remain.

High intensity noise regions include the bright rectangular labels and high intensity scanning artefacts, examples indicated for MIAS image mdb148 in Figure 7.3(a). It is useful to detect these regions as their spatial location near the skin-air interface may affect neighbourhood based operators, as in the case of the local thresholding algorithm used to delineate the skin-air interface in Chapter 8. The high intensity regions have been the source of error in image orientation algorithms based on the difference of intensities between the breast tissue and image background [89], [119]. In these references, rather than removed, they were considered a weakness of the method. Removal of bright noise is also important in the image orientation algorithm presented later in this chapter.

Figure 7.3: MIAS image mdb148 with label and scanning artefact shown in (a). The binary near-skin mask is shown in (b) and the glandular mask in (c).
Method

The high intensity noise removal algorithm depends on the threshold for the inner edge of the near-skin tissue and the thresholds for glandular tissue and pectoral muscle. These were obtained in the previous chapter. Using the example image from Figure 7.3(a), the near-skin tissue threshold is obtained by applying the minimum cross-entropy thresholding algorithm of Brink and Pendock to the entire image, producing the binary mask in Figure 7.3(b) showing the label and scanning artefact as well as the inner breast. The glandular tissue threshold is obtained by applying the algorithm again to pixels above the first threshold, producing the glandular tissue binary mask shown in Figure 7.3(c). Although it has been shown that in some images this threshold does not delineate the glandular tissue or pectoral muscle, this does not affect the performance of the noise removal algorithm.

In the glandular mask, part of the label and bright scanning artefacts are still present but the breast has been fragmented into a number of regions. As the second threshold is higher than the first, all regions of white pixels present in the glandular mask are subsets of some region in the near-skin mask. Also, as the first threshold is near the level of the skin-air interface and the second near the glandular tissue intensity level, fragments of low intensity tissue near the skin found in the near-skin mask will be excluded from the glandular mask.

As a precautionary step, morphological opening of the binary masks is performed using a 3X3 neighbourhood. The opening is needed as some images in the database contain a line of pixels at the image edges that is relatively bright. The inclusion of this line in the binary mask may connect any noise present to the breast tissue. In practice this causes problems in only one image, mdb009 where this connection prevents a bright scanning artefact from being located.

Running the simple noise algorithm on the near-skin mask results in a mask that does not include the main breast region. Due to decreasing intensity near edges, the noise regions in the glandular mask will be eroded. To extract non-eroded noise regions while ignoring breast fragments an object-based AND operation is performed using the glandular mask and the noise mask from the near-skin threshold, shown in Figure 7.4, again for MIAS image mdb148.
High intensity noise is preserved as it occurs in both masks. Breast fragments from the near-skin mask are eliminated as they do not occur in the glandular mask. Further breast fragments may exist in the glandular mask; these are also eliminated as the breast region from the near skin mask was removed in the simple noise algorithm.

In the event that lesions or microcalcifications that are relatively high in intensity exist near the skin-air interface they may be separated from the breast in the primary mask and if they are bright enough to be retained in the secondary mask, they will be removed as noise. The resulting image therefore should not be used in examining the breast tissue for such regions, rather as a pre-processor to determine image characteristics such as orientation and the skin-air interface.

**Results**

The MIAS database was found to contain 162 images with high intensity rectangular labels. Bright scanning artefacts are present in 30 of the database images.
Of the square high intensity labels, all except one was removed, representing a removal rate of 99.38%. The image in which the label was not removed, MIAS image mdb006, is shown in Figure 7.5(a). The label was not removed due to a tape artefact present in the image joining the label to the breast tissue in the near-skin mask, shown in Figure 7.5(b), preventing it from being separated.

Of the bright scanning artefacts one was not removed, representing a removal rate of 96.67%. The artefact was not removed from image mdb149 where, like the label in image mdb006, the scanning artefact was joined to the breast by a tape artefact, thus assumed to be a part of the breast. If the pre-processing opening operation had not been done on the masks, the scanning artefact in image mdb009 would also have escaped detection due to it being joined to the main breast region by a single pixel line at the bottom edge of the image.

**Locating Rectangular Labels**

While some of the extracted regions, such as tapes and scanning artefacts, are just noise and may be discarded, the labels present on the image background may contain some information such as identifying the breast as either left or right and indicating that the breast is in the
mediolateral oblique view, hence it is useful to differentiate these two bright background region types.

**Method**

To segment the image, the near-skin mask is used. The label candidates are all regions present in the binary mask, apart from the breast, located as the largest object as before. To locate the specific type of label and differentiate it from the other artefacts, such as scanning artefacts and fragments of the breast, several constraints, related to size and shape, are imposed on candidate regions.

**Constraints**

Two types of constraint are imposed on the candidate regions, size and shape. Each region is examined in turn and the constraints applied to classify them as either rectangular label or noise.

![Image](a) ![Image](b) ![Image](c)

Figure 7.6: For each region in the near-skin mask, holes are filled to improve classification. To fill the holes in the label in (a), it is isolated in the mask (b). The size of the black pixel areas is determined and all except the largest area of black pixels are filled (c).

For very small groups of pixels that are not labels it is possible that they will meet the shape criteria and be removed, hence the size of regions considered is constrained. Region size is measured in pixels, for a more accurate measure. Holes in the regions are first filled. To fill the holes of a region, such as the label in MIAS image mdb145, shown in the near-skin mask
in Figure 7.6(a), the region is first isolated in the binary mask, Figure 7.6(b). Separate regions of black pixels in this image are labelled and the largest group is taken to be the background. All other regions of black pixels apart from the background are classified as holes and filled, producing the image in Figure 7.6(c). A constant minimum for label size was set at 1000 pixels, labels found in the images used (reduced to 400 µm per pixel) being ~3500 pixels.

![Figure 7.6](image)

Figure 7.6: Vertical or horizontal lines passing through a rectangle, shown in (a-i) with a horizontal line, will intersect two edges of the rectangle (unless going through a vertical edge column or horizontal edge row), intersecting edge pixels marked. This can be used to differentiate rectangles from regions with a rough outline (a-ii), where more rows and columns will exist with more than two edge pixels. In practice, (b), the edges of rectangular labels will be jagged, due to noise and tilt and contain more than two edge pixels, as shown for the example row.

The second constraint is shape. A property of a rectangle, invariant to rotation, is that it contains two pairs of parallel sides. From this, two criteria are derived, a limit on the number of edges that each row/column must contain and the variation in distance between those edges.

Taking the perimeter of the region for an ideal rectangle shown in Figure 7.7(a-i), most rows and columns (except those containing a whole side) within the region will contain two pixels, one from each parallel edge. This criterion can be used to differentiate rectangles from jagged regions, as the roughness of their perimeter will lead to more than two pixels in each row/column, also shown on the example in Figure 7.7(a-ii). In practice, due to the tilt of the
rectangle and noise, Rows/Columns may contain more than two pixels, an example being shown for the label in MIAS image mdb107 in Figure 7.7(b), hence the constraint is relaxed to require at least half of the rows and half of the columns of each region to contain exactly two pixels from the perimeter image.

The second shape criterion is that the opposing edges are parallel. Only rows and columns that meet the first shape criterion are considered. For each row that has two lines running through it, the distance between those lines is found. For a rectangle, these distances should be constant across the shape. This will not be the case when the label is on an angle, the corners will present two lines in each row/column that are not parallel. Due to noise and erosion of edges (some labels appear as having rounded corners), these distances will not always be equal. To allow for some flexibility the standard deviation of the distances was used, allowing it to be up to 11 pixels (determined through experimentation). The same procedure is followed for the columns to determine if the lines running through are parallel.

If the constraints on size and shape for a region are met, it is classified as a label and may be extracted from the original image. The main problem with the algorithm occurs when part of the label is off the image. Such a label is shown in Figure 7.8. Here an edge is introduced into the label that is not necessarily parallel to any of its real edges, hence the label does not present as a rectangle. When the test for parallel lines is applied to the label's right edge and the edge caused by the boundary of the image, it fails. Another effect of the label being partly off screen is that its area is reduced. This could cause the minimum size test to fail.

Results
The labels in the MIAS database images are at several different angles and in some cases only part of the label is visible on the image, the rest of the label being off the image.

Figure 7.8: The label shown does not present itself as a rectangle due to the image edge.
Of the 162 labels present 81.48% are removed. This represents 30 not removed; all of which project beyond the dimensions of the image.

In 16 images something other than the label is removed. In all of these cases this other object was either a rectangular scanning artefact (12 images) or a tape artefact (4 images), which is a less serious error than removing a part of the breast. The segmentation of such artefacts is not desired as they do not contain the information which may be searched for in the label and their identification would need to be considered in any post-processing algorithms.

**Summary of Background Noise Removal**

Combined, the algorithms presented for background noise removal provide a hierarchical decomposition of the image background. The simple noise algorithm locates most of the noise in a binary mask with the risk of extracting breast tissue. The high intensity noise results are a subset of the simple noise, obtained by using information from the glandular mask. Figure 7.9 shows three images for which the resulting image background noise masks have been superimposed on the original. Rectangular labels are shown in yellow, bright noise not classified as labels is shown in green and noise extracted using the simple algorithm not classified as a bright artefact is shown in red.
Tape Artefact Location

Markings left by tapes, or other shadows presenting themselves as horizontally running strips, no matter what the cause, will be referred to here as tape artefacts for simplicity. Tape artefacts usually overlap the breast tissue and therefore are more difficult to remove than the vertically running scanning artefacts. The tape artefacts are generally not homogenous in intensity, so that intensity thresholding is not in most cases an adequate segmentation technique. Some examples of scratching are shown in Figure 7.10.
The common property of tape artefacts is the horizontal line corresponding to their edge. This property is used to locate the artefact, flagging its presence so that the corresponding tissue may be removed from computation, rather than removing the effect of the artefact from the affected breast tissue. Since rectangular labels will also have horizontal lines, they are first removed using the bright artefact removal algorithm. Also, since the tape edge is most distinct when it overlaps the breast, only the left half of the image is used for breasts facing right, this required the orientation algorithm presented further in this chapter to re-orient left facing images.

Location of the tape edge is performed using the Radon transform, originally formulated in [120]. The Radon transform can be used to detect straight lines, similar to the Hough transform, by mapping an image into a space with position, angle, and strength of a straight line as its dimensions. The horizontal running edges of the image are first enhanced using a 3x3 horizontal Sobel [121] mask, shown in Figure 7.11.
The Radon transform is then applied, for horizontal lines only. Figure 7.12 shows original image mdb002 in (a) flipped onto its side to put the tape edge location onto the x-axis. The label removed, and edge enhanced image is shown in (b), and results of the radon transform from the left half of the image (bottom half of rotated image) shown in (c), superimposed over the original with the x-axis representing the image row, and the y-axis the degree that that row contains a straight line of pixels. From this graph, the correspondence of straight edge in the image to peak in the graph can be seen. In order to apply a static threshold to the Radon transform graph, it is first normalised to remove effect from background by subtracting the average in a local 14-point sized sliding window, producing a normalised graph shown in Figure 7.12 (d), also superimposed on the original. A threshold, set through experimentation, is then applied to locate candidate tape edges. The threshold must be set for a trade off between the percentage of artefacts detected (sensitivity) and the percentage of cases detected as artefacts that are actually artefacts and not false positives (specificity).

Figure 7.12: (a) Original mdb002, (b) Sobel edge enhanced for tape edges, (c) superimposed radon transform for tape edges, (d) normalised radon transform for tape edges.
The location, above or below the edge, of the tape is determined through a simple rule. If a tape edge is detected in the top quarter of the image, the tape is assumed to be above the edge, if the edge is detected in the bottom quarter of the image, then the tape is assumed to be below the edge.

**Tape Location Results**

The threshold for tape edges determines the sensitivity and specificity of the algorithm, with higher threshold decreasing the number of false positives but also decreasing the percentage of located tapes. The tape location algorithm is used in the next chapter to aid the skin-air interface detection algorithm, there a threshold level of 35 is used – however that level is chosen to account for mode filter edge enhancement, used in another part of the skin-air interface algorithm, as well as the horizontal Sobel filter described here. With that threshold, no false positives are found, while some of the less distinct horizontally running shadows are missed.

**Scratch Removal**

Scratching of the film is a frequent artefact present on the images in the MIAS database. A typical scratch can be seen in MIAS image mdb147, shown in Figure 7.13(a). The scratch runs vertically down a column and varies in intensity across its length. The intensity of scratch pixels varies between images and its visibility depends on the regions that the scratch passes through.
While scratches present as seemingly insignificant features in the mammogram, sometimes being almost invisible, they can be 'amplified' during computer processing. For example, processing the image in Figure 7.13(a) by the Extended Russ operator of Chandrasekhar [89] results in the image shown in Figure 7.13(b) with the scratch appearing more significant.

A simple algorithm is implemented to locate vertical scratches, using a difference operator to determine the likelihood of a scratch being present in a column. False positives are eliminated by examining the likelihood of adjacent columns containing a scratch and by locating bright scanning artefacts and labels.

**Method**

A scratch represents a distinct change in the intensity profile of a row. The difference between a pixel and its left and right neighbour is a simple measure of roughness in the horizontal direction for that pixel. Taking the average roughness of pixels in a column results in column roughness measure. The presence of a scratch is associated with an increased roughness, hence the average roughness is a measure of scratch likelihood. MIAS image mdb147 is shown again in Figure 7.14 along with the scratch likelihood function. The high points in the function correspond to (a) scratch, (b) label, and (c) scanning artefact.
Due to the scratch likelihood measure being an average intensity difference for a column, similar peak amplitudes arise from faint scratches covering the whole column and the edges of short, bright objects such as labels. To alleviate this, the likelihood function is weighted using the length of a likely scratch. To determine the length, the number of pixels that are significantly different to their neighbours is counted, a significant difference being defined as greater than 0.02, or 2% of the intensity range which is from 0 to 1. The scratch likelihood of a column is scaled if less than half of the column shows significant difference, dividing it by the total number of pixels in the column minus the number of significant scratch pixels.

The edges of scanning artefacts that are present in the MIAS database typically encompass a whole column. To account for these artefacts in the scratch likelihood function, they are first located in the image, as shown earlier in this chapter. Once the location of these regions is
determined, the pixels corresponding to their vertical edges can be flagged and removed from the computation of the scratch location, rather than the whole artefact. Removal of the high intensity background regions for scratch location is not desired, as the scratch may run through the artefact. Removal of the artefact will lead to a loss of scratch information, making the scratch more difficult to detect.

A fixed low threshold is applied to the scratch likelihood data. This has the effect of breaking the function into areas corresponding to separate peaks. For each peak, the filter is centred on the highest point. Two neighbourhoods of size 10 columns are defined on either side of the peak centre, a distance 4 columns away from it to exclude the peak from them. It is then checked that the maximum value of the function within the neighbourhood boundaries is less than 40% of the centre column scratch value.

This scratch filter is illustrated in Figure 7.15 with an example from image mdb003. In this image the scratch runs down the centre of a label so that peaks stemming from the label edges surround the scratch peak. Figure 7.15(a) shows the scratch peak meets the requirements and Figure 7.15(b) gives an example of a non-scratch peak being rejected.

Once a scratch is located, its position can either be passed to subsequent image processing routines or the scratch can be removed from the image. To remove a scratch, a smoothing
operation is performed on its area where scratch pixels are replaced by the average of their left and right neighbours.

**Testing Data Set**

Each MIAS database image was examined manually to determine images affected by visible scratching. To do so, each image was viewed as the histogram equalised original to better visualise light scratches on a dark background and also an inverted image where dark scratches on a light background are easier to see.

The extended Russ operator, by Chandrasekhar [89] was then used to transform the original images. This neighbourhood-based operator has the effect of amplifying the intensity difference of the scratches and also increases their apparent size, demonstrated by the example in Figure 7.13. This property was used to confirm the location of scratches located on the original images and to locate more subtle ones. Once located, the scratches can be smoothed and the performance verified by transforming the scratch free image with the Extended Russ operator.
Results and Discussion

All together, 48 scratches are present in the MIAS database, covering 44 images. The algorithm finds 40 of the scratches, representing approximately 83%. Eight scratches are relatively short (less than ¼ of the image), and are not found by the automatic algorithm. Three false positive scratches are detected. In all three images the false detection was due to bright scanning/tape artefacts not located during pre-processing.

In image mdb149 the high intensity noise removal algorithm fails to detect a bright artefact due to it being connected to the breast tissue. The scratch location algorithm therefore has no knowledge of the tape and classifies its edge as a scratch.
The other false positives are encountered in images mdb274 and mdb280. The problem with these images is illustrated by mdb274 in Figure 7.16. The image has very significant artefacts due to a combination of tapes and scanning artefacts. Several intensity layers are present in these artefacts but the artefact location algorithm identifies only the outside boundary. The intensity change inside the tape boundary (a) is classified as a scratch, this problem also occurs in mdb280. A scratch exists in the image at b) and is classified correctly.

Cleaning was performed on all located scratches including the false positives. The extended Russ operator was then applied to transform the images. The result for MIAS image mdb147, transformed original shown in Figure 7.17(a) is shown in Figure 7.17(b). Such cleaning has shown to be affective in that the cleaned scratches no longer pose a problem for the extended Russ operator. The false positives had no significant visible effect on the transformed images.

![Figure 7.17: Original image mdb147 transformed with extended Russ operator. b) Scratch removed image transformed with extended Russ operator.](image)

**Conclusion**

While scratches seem insignificant on the original image, further processing can amplify them. Vertical scratches can be found by looking for the characteristic intensity spike occurring in each row where the scratch is visible. To avoid false positives from scanning artefacts and labels, post-processing in the form of a filter for the scratch likelihood function was used.
Mammogram Orientation

Orienting the mammogram in a particular direction is required in several computer aided diagnosis schemes. Some, (eg. [81],[119]), offer a solution that differentiates between left facing and right facing breasts while assuming the image is not upside down, others omit a comprehensive solution. Since there is no orientation standard for mammogram digitisation, the cases of images rotated onto their side (landscape mode) and upside down images need to be considered. In this section an algorithm to determine the orientation for mammograms in the mediolateral oblique view is presented.

In reference [81], Karssemeijer used a number of features to classify mammograms into different types (more on this in Chapter 10). His segmentation algorithm assumes that “the chest side is on the left, and that the image is not upside down” [81]. It is stated that the chest side is determined automatically, without an algorithm description and the problem of upside down images is not addressed.

A simple orientation algorithm was utilised by Chandrasekhar in [89]. Examining the image dimensions under the assumption that image height was greater than image width determined if an image was rotated onto its side. After this re-orientation, the chest wall is assumed as either on the left or right side of the image. Four pixel regions are extracted from near the corners and the brightest region assumed to be nearest the chest wall. The algorithm was tested on MIAS database images, however no quantitative results are reported. Upside down image are not considered and the algorithm was reported to fail “when the background is not all dark” [89] and when scanning artefacts were present.

Lou et al. [119] investigated three methods to determine image orientation, but limit themselves to differentiating between left facing and right facing images. Two of the techniques are rejected due to their failure in certain cases, while the third is used in their implementation.

Similar to Chandrasekhar’s method, their method is based on pixel value summation. Two rectangular areas from the left and right sides of the image are taken, and the area with the largest pixel sum assumed to be nearest to the chest wall. Problems are identified by Lou et al. relating to non-breast image regions, such as scanning artefacts, labels and also compression paddle projections from digital mammograms. It is concluded that these will
affect the pixel summation so that “it does not hold that the larger sum always indicates the
side of the breast region”. A solution to this problem may be to remove such bright non-
breast features, using the algorithm described for scanning artefacts and labels (paddle
projections are not found in the MIAS database), however problems might still exist for
images such as the ones in Figure 7.18, showing MIAS images (a) mdb031, (b) mdb258 and
(c) mdb151 where bright tissue is present on both sides of the image. Depending on the size
of the rectangles on each side, the image side away from the chest wall may contain the
brighter pixels.

![Figure 7.18: Examples of images hard to orient based on simple
pixel distribution models. (a) and (b) and (c) containing relatively
brighter tissue on the image side away from the chest wall.](image)

The second method of Lou et al. uses a simple histogram texture measure to determine the
orientation. It is assumed that the image background and any artefacts will have a more
uniform pixel distribution than the breast region. The same rectangular regions are employed
as in the pixel summation method, and a smoothness measure is taken from each area’s
histogram. The image side with more histogram variation is taken as the breast side. It is
documented that the algorithm fails for mammograms with large breast regions (Figure
7.18(c) is a possible example).

Lou et al.’s third method uses the fact that the intensity profile of the breast near the skin
shows intensity decreasing in the direction away from the chest wall. The algorithm first uses
a clustering technique to remove local variation of pixels, horizontal intensity profiles are
then taken from a number of image rows and the direction of the global intensity decrease across the profile is taken as being away from the chest wall.

The problem of image orientation in rotated images and upside down images can be split into two. The first task is to determine the location of the chest wall, done by Lou et al. for right way up images – but needing some modification to generalise it. The second problem is to determine the image top. This task becomes easier once the chest wall location is determined, the image top being one of the image edges adjoining the chest wall.

**Determining Chest Wall Location**

The key to the success of Lou et al.’s orientation algorithm is the decreasing intensity of breast tissue near the skin-air interface. It was noted by Highnam et al. in [123] that when the breast is compressed by the flat compression plates, tissue is spread to a uniform thickness except at the edge, where “it bulges like a balloon” [123]. The bulging at the edge leads to tissue thickness decreasing, the effect on the image being the appearance of intensity contours near the skin-air interface, intensity being lowest at the edge where the bulge has the lowest thickness. The intensity contours are best seen by viewing the mammogram with greyscale values re-mapped to a random colour map, as suggested in [89], shown in Figure 7.19 for MIAS image mdb011.

![Figure 7.19: MIAS image mdb011, shown in (a) with the original greyscale intensity mapping and in (b) with the greyscale values re-mapped into a random colour map.](image)

126
For determining the chest wall, analysis is simplified by modelling the breast as a compressible balloon, similar to the model developed by Highnam et al. [123] for the estimation of compressed breast thickness. The balloon model can be used to determine the chest wall. Consider a round balloon compressed between two square parallel plates. Two regions result, shown in the diagram of Figure 7.20(a) for a top down view, showing the area of compression in (i) and for a side-on view, showing thickness variations, in (ii). There is a flat region of uniform thickness in the centre, surrounded by an unevenly compressed region, shown shaded, in which the thickness increases in the direction of the balloon centre, indicated by the dot-dash lines. The sum of the increasing thickness direction vectors around the balloon is zero.

Figure 7.20: Compression of a round balloon between two parallel plates (plate position indicated with dashed lines). In (a), the balloon is at the plate centre, while in (b) only a part of the balloon is subject to compression. The model is viewed top-down in (i) to show compressed area and from the side in (ii) to show thickness. The compression divides the balloon into two regions, a region of constant thickness in the middle surrounded by a region in which the thickness decreases towards the edge, indicated in shading. The thickness is seen in the side-on view as a bulge and arrows indicate the direction of thickening in both views. The model in (b) is similar to the compression encountered in a mammogram, where the chest wall constrains the area that is compressed.
In the case of the breast, the chest wall acts to constrain the compression on one side, this can be modelled on the balloon by moving the compression plates so that they do not compress the entire balloon, as is shown in Figure 7.20(b). Considering only the area acted upon by the compression plates, the vectors of increasing thickness direction will sum to a vector pointing to the compression plate edge adjacent to the uncompressed region of the balloon, this edge is analogous to the chest wall in the mammogram.

A translation of this simple and idealised model to a mediolateral oblique view mammogram is shown in Figure 7.21, with arrows to indicate the increasing intensity in the near-skin tissue. Although the compression plate in the oblique view is at an angle to the image edge, the chest wall is still most dominant on the left image edge as in the balloon model.

The poorly compressed tissue as a mammogram region in itself has been used in the past for nipple location [124] and in an algorithm to estimate the thickness of the compressed breast [123].

By recognising the near-skin tissue as a separate region, some of the limitations present in the Lou et al. algorithm can be overcome. The limitations of the Lou et al. algorithm are as follows:

1. Extracting an intensity profile across the entire image makes the technique prone to the local variations in parts of the image that have nothing to do with the decreasing
gradient of pixels near the skin, such as noise in the image background and the texture of the glandular tissue, requiring the clustering algorithm for pre-processing.

2. Using only a select number of profiles makes the technique more sensitive to non-ideal images, such as those in Figure 7.18(c), where the near-skin tissue is cut off in parts of the image.

3. The use of horizontal intensity profiles limits the method to finding the chest wall in images with it either on the left or right side of the image.

These problems can be addressed by either constraining or generalising parts of the algorithm.

The solution to the first problem of considering non-relevant image areas is to constrain the area of interest to the tissue exhibiting the property of decreasing intensity, namely the near-skin tissue. This has been addressed in Chapter 6 using multi-level minimum cross-entropy thresholding.

The second problem, of using a small part of the information available when certain intensity gradients are selected would be aggravated by limiting the area of interest to just the near-skin tissue. If the near-skin tissue in the row selected for profiling were cut off, no other information from that row would be available. Using all the pixels available from the near-skin tissue and averaging the result can solve this problem. This also removes some sensitivity to noise, hence Lou et al.’s clustering technique for noise removal is not employed. The exact solution leads into a solution for the third problem.

The third problem stems from the technique being constrained to images where the chest wall is either on the left or to the right. The act of extracting a horizontal intensity profile and measuring the decrease of its curve amounts to measuring the pixel gradient in the horizontal direction. For images rotated onto their side the same could be done but with vertical intensity profiles, measuring the pixel gradient in the vertical direction. The use of intensity profiles is however unnecessary, as pixel gradients can be computed using neighbourhood filtering, applying horizontal and vertical derivative filters within the area of interest.
Implementation

After removing labels and bright scanning artefacts, the image is segmented into background, near-skin tissue and breast tissue. This requires two thresholds, determined using the multi-level thresholding procedure described in Chapter 6. To remove further noise, the binary masks for inner breast tissue and near-skin tissue are searched for the largest object, all other objects are removed as in the simple noise removal algorithm described earlier in the chapter. Figure 7.22 shows MIAS images (a) mdb011, (b) mdb207 and (c) mdb133, along with the segmented images in Figure 7.22(d,e,f).

Figure 7.22: Original MIAS images (a) mdb011, (b) mdb207, and (c) mdb133, segmented into background, near-skin tissue and breast tissue.
Horizontal and vertical derivatives are used to determine the direction of increasing intensity in the near-skin tissue, whose net direction should be into the chest wall. The derivatives are obtained by filtering the image with the kernel masks depicted in Figure 7.23, such masks are typically used to determine the amount of intensity change in a single direction \cite{87}, with the kernel in (a) sensitive to intensity change along the horizontal, and in (b) along the vertical. The exact pixel direction vector can be determined from the two derivatives, however orientation is simplified to deal only with images rotated in 90° increments by using only the sign of each derivative, to determine the direction as either up or down or left and right. An example is given in Figure 7.24 using MIAS image mdb011, with pixels classified as skin-air interface being shown in white for cases of increasing intensity towards the left side of the image (a), towards the right side of the image (b), towards the image top (c), and towards the image bottom (d).
If the number of pixels with positive horizontal gradient (intensity increasing towards the left image edge) is denoted by $N_L$, number of pixels with a negative horizontal gradient as $N_R$, and similarly the number of pixels with positive and negative vertical gradients denoted by $N_U$ and $N_D$ respectively, than the net directions in the horizontal and vertical direction and their magnitudes can be defined as $VH$ and $VD$.

$$VH = N_L - N_R$$  \hspace{1cm} (26)

$$VD = N_U - N_D$$  \hspace{1cm} (27)

The signs of $VH$ and $VD$ in (26) and (27) indicate the net direction, the magnitude indicating the strength of the gradient.

The most dominant pixel direction is taken as being directed into the chest wall. In the example of Figure 7.24, $VD$ will be small due to pixels pointing up having a similar number of counterparts pointing down, $VH$ will be larger with overall direction pointing left, so the left image side is taken as the chest wall.
With the chest wall found, the image is rotated into an orientation with the chest wall on one side. The chest wall is placed on the left side in the implementation.

**Determining Image Top**

With the chest wall on the left hand side, the image can be in two possible orientations, right way up or upside down.

The asymmetry of the breast region about its vertical centroid is used to determine the image top. A characteristic asymmetry exists for the breast in the mediolateral oblique view, the breast generally being set at an angle, coming from the pectoral muscle in one corner to the breast tip in the opposite corner. To determine asymmetry, the threshold on the inside of the near-skin tissue is used to produce a binary breast mask. Since this threshold cuts off the tissue near the skin, an inaccurate breast outline results, the advantage is that most external low intensity noise is removed and due to the intensity gradients in the near-skin tissue the rough breast shape is preserved.

The vertical centroid of the binary mask is determined as the row with equal number of 'on' pixels above as below it. Asymmetric pixels in the mask are defined as binary ‘on’ pixels on one side of the centroid that do not have corresponding ‘on’ pixels on the opposite side. The vertical centroid and asymmetric regions are shown in Figure 7.25. Due to the characteristic asymmetry, regions of asymmetric pixels will exist. Because of the angle of the breast, an asymmetric region will exist at the breast tip below the vertical centroid in an image the right way up. Also, due to the effect of the pectoral muscle (exaggerated by the threshold being placed in tissue internal to the breast) another asymmetric region may exist where the pectoral muscle is located. As labels and scanning artefacts have been removed, any asymmetric regions that may have been attributed to them are eliminated.
Figure 7.25: Binary breast masks for the example images, showing the vertical centroid in red and asymmetric pixels in green. The breast tip features an asymmetric region closest to the right side of the centroid, situated below it due to the breast angle.

The image top is found by assuming the asymmetric region closest to the right side of the vertical centroid is the tip of the breast. The image is flipped if needed to place this asymmetric region below the vertical centroid – resulting in an image the right way up.

**Results and Discussion**

The algorithm has been tested on the entire 322-image MIAS database. Images were put through all possible 90-degree rotations with upside down and right-side-up images, examples shown in Figure 7.26. Correct orientation was determined for all cases, even where part of the near-skin tissue, used to locate the chest wall, was cut off, demonstrating robustness against shape variation and incomplete breast image, examples of which are shown in Figure 7.26.

Although the algorithm has been successful in all test cases, it has the potential to fail in images where the breast is cut off. Since the chest wall location algorithm depends on tissue near the skin, using an average gradient of these pixels, the exclusion of part of this tissue has the potential to affect the algorithm. Tissue near the breast tip, which is situated closest to the image edge, is most prone to being cut off – unfortunately this section of tissue is also most important in determining chest wall location, as direction vectors of pixels within it will point most directly into the chest wall. Consider for example MIAS image mdb152 shown in Figure 7.26(a). Here the breast tip is cut off and the detected near-skin tissue is along two almost parallel sides. In images like this, the pixel direction vectors will have a tendency to cancel out. Although in practice, this has not been found in either the MIAS database images
or documented by others also using the thickening property of near-skin tissue to determine orientation [119].

![Image](image_url)

**Figure 7.26:** Examples of images and orientations used to test the algorithm. MIAS database images (a) mdb152r, (b) mdb274r, (c) mdb318r, and (d) mdb283l.

**Conclusion**

The algorithm presented provides a solution for determining the orientation of a breast in a mediolateral oblique view mammogram. Two steps are performed to find the orientation.

The step to determine the location of the chest wall uses a compressible balloon model for the breast. The breast boundary in the image is composed of the skin-air interface and the chest wall. Tissue near the skin increases in thickness towards the breast centre, similar to the edges of a compressed balloon, evident through increasing intensity in the image. Taking direction vectors in the near-skin tissue and summing them gives an indication of the chest wall location.

The subsequent operation to determine the image top uses the chest wall location and the non-symmetric nature of the breast in the image around its vertical centroid. Taking mirror images of the breast around its vertical centroid is used to determine asymmetric regions. Analysis of the location of these asymmetric regions determines the top of the image. Both chest wall location and image top location were successful on all images from the MIAS database, resulting in correct orientation.

Overall, the methods presented in this chapter can be considered a pre-processing to simplify further mammogram segmentation, such as for the extraction of the skin-air interface and the pectoral muscle. The methods can be classed as segmentation of the image background, separating labels, scanning artefacts and scratches from the breast scratches.
orientation, related to the position of the breast in the image background can be thought of as an extension of this segmentation.
Chapter 8

EXTRACTING THE SKIN-AIR INTERFACE

Chapter 6 on multi-thresholding included the setting of thresholds near the skin-air interface, however to obtain an accurate outline across the entire skin line is not as simple as setting a global threshold. In this chapter, the skin-air interface extraction is simplified by modelling the breast border as a set of overlapping functions, splitting it into a skin-line found along the image columns (column-wise interface) and another along the rows of the image (row-wise interface). Using the initial global threshold estimates from the algorithm in Chapter 6, two algorithms are presented, one which applies the minimum cross-entropy to small local windows along the estimated outline and combines their estimates using a smoothness criterion. The alternative algorithm fits polynomials to the row and column interfaces generated by several thresholds and sets the threshold and order of the polynomial for each part of the skin-line to minimize error between extracted pixels and the fitted line.

The polynomial-based algorithm shows the best results due to the smoothness enforced by the polynomial functions. Another property of the polynomial-based algorithm is that it ignores the nipple when it is in profile, while the local threshold algorithm tends to extract it. An analysis of the polynomial algorithm performance on the MIAS database is shown, and the algorithm is found more accurate overall than an existing method.

Introduction

The skin-air interface is the boundary between the breast and the image background. Constraining further processing to the breast portion of the image is an important role of breast border segmentation, used for such a purpose in Chapter 10. Many techniques make do with an approximation near the actual border, using a global threshold either by applying a thresholding algorithm to the histogram [26], [39], [81], [125], [126] (setting the threshold in the broad histogram valley in the near-skin tissue, as the algorithm in Chapter 6), or trying different thresholds to determine one that maximises the smoothness of the extracted breast
boundary [123]. The outline obtained from a global threshold is often further refined using morphological processing to produce a cleaner estimate of the breast edge [26], [127].

An approximation to the skin-air interface internal to the actual boundary is relatively easy to obtain by a global threshold. Due to the intensity contours just internal to the breast (see image orientation algorithm in Chapter 7), the approximation will follow the actual breast boundary closely and only cuts off the poorly compressed fatty tissue, not glandular regions of the image. Thus the approximation to the skin-air interface from a global threshold is adequate for many tasks, such as the restriction of further processing to only the breast regions. The segmentation of the actual boundary however is complicated by noise at a similar greyscale level and the non-uniform intensity of the breast along its boundary – meaning that for accurate extraction of the breast boundary, a global threshold is not suitable, a fact widely recognised [119], [128], [129], [55], [130]. Besides merely restricting processing to the breast region, an accurate breast boundary has several uses, such as location of the nipple, as discussed in [124] and location of registration points to register left and right breasts.

Accurate segmentation of the breast boundary has been approached in a number of ways. Several authors have used the change in intensity gradient from the breast to the background to find the skin-air interface [119], [128], [129], [55], [131], [86]. For example, Mendez et al. [131] smoothed the mammogram using an 11x11 mean filter and positioned the breast with the nipple facing up. The breast was then divided into left, central and right sections. Using the gradient along a vertical pixel profile for the central section (section where breast edge runs closest to horizontal) and gradient along horizontal pixel profiles for the two side sections (here the breast edge runs most vertical), the criterion for finding the breast edge pixel in a profile was to find a sequence of nine adjacent pixels whose intensity was increasing into the breast, from pixel 1 to 9, the 10th pixel being classified as the breast edge. Although no analysis of pixel accuracy due to this method is reported, the boundary may be sensitive to background noise and seems to have an inherent problem due to the window sizes chosen. This can be explained using an ideal case of a horizontal profile with a distinct edge, shown in Figure 8.1(a) using numerical pixel values rather than greyscale shading. Use of the mean filter will naturally ‘blur’ the edge, creating an intensity gradient affecting pixels as long as the averaging window overlaps both regions (5 pixels for the 11x11 window used), this is shown in Figure 8.1(b). Requiring a pixel gradient in 9 pixels before the breast edge pixel will therefore introduce an error of 3 pixels, shown in Figure 8.1(c).
A similar approach for the gradient tracking was undertaken by Abdel-Mottaleb et al. [128]. First a rough estimate of the breast was obtained by thresholding midway between the background peak and main breast peak. The mammogram was positioned with the nipple facing left. The widest row of the breast was located, and the narrowest rows above and below this widest row were extracted. Since the breast has a fairly predictable shape, these rows were used to create a model of the expected shape so as to determine which direction pixel profiles were extracted around the breast. These pixel profiles were then searched for “intensity gradients appropriate to the subcutaneous region of the breast”. [128] resembling those typically found in the tissue near the breast edge.

As there is a physical difference between breast and background, the intensity gradient at the skin-air interface is expected to be high. In practice however, large amounts of noise near the breast border, or more distinct intensity gradients internal to the breast may lead to a boundary internal, or in places external, to the breast being extracted. This is a possible shortcoming of the methods by Mendez et al. and Abdel-Mottaleb et al., in fact the example image given in [128] is characteristic of a breast boundary that is actually internal to the breast. To counteract this problem several approaches may be used, including one presented in this chapter which uses a local application of the minimum cross-entropy algorithm to
refine the estimate near an approximation, and is also reported in [132]. Other approaches also exist. For example, Aylward et al. [129] found the first point along the boundary near its middle, where less noise is found, by examining the intensity gradient along a horizontal profile running through the centre of the breast. A ridge traversal algorithm was then used to find the rest of the breast boundary. Intensity gradients along horizontal profiles were also used by Zhou et al. [86] to determine an initial estimate after which more accurate tracking was implemented, again from the middle of the breast, using the previously found estimates and gradients normal to the tracking direction. Lou et al. [119] first thresholded the image to estimate the breast boundary and also used gradients normal to the estimated boundary to get closer to the actual boundary – this was followed by a process of refinement using further grey level analysis, and finally a linking and smoothing of the boundary to get a more accurate breast edge.

A different way of using image gradients has been implemented by Highnam and Brady [55]. Rather than examining the intensity transition gradient from background to breast it is observed that on the whole, the gradient of the image background is lower than that of the breast. A threshold value for the gradient separates breast from background and is followed by morphological operations to produce a cleaner separation.

Various other methods for automatic breast boundary detection are found in the literature. Bick et al. [133] constructed a modified histogram, including only pixels with a low range of intensities in their local neighbourhood. This histogram was used to identify pixels corresponding to directly exposed and unexposed image regions, followed by region growing to extract the breast region.

Suckling et al. [134] made use of a self-organising map neural network, using 18 features, including statistical measures and entropy from local neighbourhood windows and information about pixel pairs.

Heine et al. [135] use a multiresolution wavelet domain expansion, observing that as lower frequency expansion images are taken, the average power of the breast image portion increases and the average power of the image noise decreases. Taking a reduced size wavelet expansion image at a low frequency, a simple threshold is applied to eliminate noise outside the breast (not including bright artefacts). The largest object in the resulting binary image is
identified as the breast and processed with a box-car averaging filter to fill holes. Finally, dilation is used to expand the breast from the reduced image back to its original size. The segmentation example presented in [135], shows a fairly ‘rough’ breast outline, reporting that approximately 13% of the breast area pixels were missed on average.

A combination of region growing and k-means clustering was used by Lee et al. [64], where the image is divided into blocks and an energy function based on the mean and variance of pixels inside a block used to identify the block most likely to contain the background. Region growing is then used, starting with the background block, to extract the other blocks containing the background. The estimate of the breast-background regions is then refined using k-means clustering. Region growing is also utilised by Neiber et al. [136] for segmentation of the breast tissue. The grey levels in the image are divided into dark, medium and bright, with the largest set of connected medium pixels identified as the breast. Placing a seed-point inside the located breast, region growing based on intensity is used to get a better estimation of the breast tissue.

Saha et al. [138] used scale-based fuzzy connectivity to segment the breast region in their work on breast tissue density quantification. It is reported that the algorithm works ‘correctly’ however there is no detailed analysis of the closeness of the detected boundary to the skin-line. The fine accuracy is not clear, since evaluation of tissue density does not critically depend on an accurate breast boundary, global thresholding being reported as adequate in other work [81], and also in Chapter 10 of this thesis.

Ojala et al. [130] combined histogram thresholding and morphological filtering to obtain an initial segmentation. Initial boundary points are selected automatically from the segmented image and three contour modelling methods are then compared: Fourier descriptors, snakes and B-splines. Ojala et al. conclude that all three methods produce similar results, very close to the actual skin-line of the mammogram, although all rely on the setting of some parameters. Although no direct comparison of results was undertaken, it is hypothesised that the boundary obtained using contour modelling methods would show similar characteristics to the algorithm based on polynomial fitting, presented in this chapter. B-Splines are also used by Ferrari et al. [137], applied also to a binary image after morphological processing.
Algorithm Used for Comparison

Chandrasekhar developed a skin-air interface segmentation algorithm working on the principle of eliminating background noise before extracting the breast boundary [89], [139], [140], a similar principle as used by Heine et al. [135] although Chandrasekhar developed his algorithm several year prior to Heine’s published work. Chandrasekhar’s algorithm from [89] consists of several steps. Firstly, an approximation to the background region is obtained using a global threshold. Secondly, the extracted background is modelled by a polynomial in the spatial variables x and y on the basis of the Weierstrass approximation theorem [89] and the modelled background is subtracted from the original image. Thirdly, the background subtracted image is thresholded again, and since the background noise has been eliminated a cleaner outline results. This approach is similar in intended effect to Heine et al.’s wavelet decomposition method [135] where a low frequency image is used to obtain better separation of breast and background, allowing a global threshold to be set. The fourth step in Chandrasekhar's algorithm is postprocessing of the thresholded image. This involves extracted regions being labelled and the largest region taken as the breast, other regions merged with the background.

Chandrasekhar’s algorithm involves three parameters, the initial threshold, the degree of the fitted polynomial, and the final threshold. In the original work, the initial threshold to extract the background was determined by examining images in the MIAS database [19], being set at the static value of 12 for all MIAS images reported. The polynomial degree was chosen interactively between 0 and 3, and it was reported that if the polynomial degree was chosen carefully, a final static threshold of value 0 was adequate. The setting of the parameters was automated by Chandrasekhar and Attikiouzel in [140], a parameter known as ‘monochromaticity’ extracted from the cumulative histogram and used to guide the setting of the two thresholds. Candidate output images were produced using polynomials of order 0, 1, and 2, and the final output image was a union of several candidates, determined also using the monochromaticity value.

Chandrasekhar’s algorithm was tested on the MIAS database [19], with 30 of 32 images presented to a radiologist found ‘acceptable’ [140]. Results of the algorithm have since been used by others groups [30], [84], identifying it as an accepted algorithm. Because of the algorithms acceptance and availability of results from [140] for the whole MIAS database, it
makes a good level for comparison. Results of the skin-air segmentation algorithm from this chapter are therefore evaluated in terms of performance relative to Chandrasekhar’s algorithm.

This chapter is divided into two parts. First, the key idea of dividing the skin-air interface to simplify the problem is presented. Following this, two algorithms are outlined to find the skin-air interface using the skin-air simplification. While the concepts behind the algorithms are simple, the presence of noise, scanning artefacts, tapes, nipple in profile, and low contrast between skin and background requires some pre-processing, which is also presented.

**Row-Wise and Column-Wise Skin-Air Interface**

The key idea behind the algorithms in this chapter is the division of the skin-air interface into one obtained from the image rows (row-wise interface) and one from the image columns (column-wise interface). Due to the nature of the image, the breast in the mediolateral oblique view mammogram can be thought of as having four sides, the top and left sides correspond to the edges of the image, the actual skin-air interface can be split into a portion at the right of the image (for right facing mammograms) and a portion at the bottom of the image. The row-wise interface is defined as the skin-line such that each row of the image has at most a single skin-line pixel and represents the right edge of the breast, the column-wise interface such that each image column has at most a single pixel and represents the bottom breast edge. The complete skin-air interface is a combination of the row-wise and column-wise skin-air interface. This simplifies the problem by transforming each part of the skin-line into a function with one value for each row or column.

Consider the mammogram shown in Figure 8.2 (a), this is MIAS image mdb016, reduced in resolution to 30% so as to make individual pixels more visible. The image is shown again in (b) with label removed. A binary mask obtained by applying the Brink and Pendock minimum cross-entropy algorithm to the image through two iterations as in Chapter 6 is shown in Figure 8.2(c). The row-wise interface is relatively easy to obtain as the breast sits against the left hand side of the image. For each row the first white pixel in the row is classified as the start of the breast and the first black pixel to the right of this white pixel as the end, the pixel before this being used as the row-wise breast edge. The row-wise interface found from the example image is shown in Figure 8.2(d).
Figure 8.2: Splitting the skin-air interface into two parts: (a) reduced size original, (b) label removed, (c) binary mask as a result of global thresholding. The row-wise interface from the mask (1 pixel for each row) is shown in (d), and the column-wise interface (1 pixel for each column) in (e). The complete skin-air interface is a combination of these two.

Next, the outline of the bottom of the breast is desired. For each column, starting from the top of the image, the start of the breast is found as the first white pixel and the end as the first black pixel after that, the column-wise skin-air interface pixel being the white pixel just above it. The column-wise interface is however more difficult to obtain as the top of the breast does not sit against the top edge of the image and thus the first white pixel could be
noise, or the first black pixel after the first white one could be the end of the section at the top of the breast – running parallel to the pectoral muscle as seen in the Figure. The problem is overcome by a requirement that the first white pixel in the column be followed by a set number of white pixels in order to classify it as the start of the breast, this number of pixels can be set as a fraction of the image height, 1/6 used in the results presented here. The column-wise interface found through this method is shown in Figure 8.2(e). The behaviour of the row-wise and column-wise interface as functions can be seen using the axis orientations shown in Figure 8.2(d) and (e). Because of the single pixel requirement, the functions do not ‘wrap’ back over themselves.

In order to realize the skin-air interface split into row-wise and column-wise interface automatically, the orientation of the image must be determined; this is implemented using the algorithm presented in the previous Chapter. Further pre-processing already covered is the removal of labels and bright scanning artefacts from the greyscale image and the use of the simple noise removal algorithm on the binary images produced by thresholding to isolate the main breast tissue.

Skin-Line estimation from Simplified Interface

Two algorithms are presented which utilise the idea of a row-wise and column-wise skin-air interface. One uses a polynomial fitting technique, the other a local threshold-based technique.

Polynomial Based

The idea of the polynomial technique is to fit polynomials of several orders to the row-wise and column-wise interface generated from several thresholds and choose the polynomial order N and threshold T to minimise the mean squared error between the polynomial and the extracted row-wise or column-wise pixels. This relies on the idea that the skin-air interface is the smoothest section of identical intensity pixels near the breast edge, thus producing the least mean squared error when compared to the inherently smooth polynomial. As for the use of polynomials, this can be justified as the row-wise and column-wise boundary are in the form of functions that do not wrap back over themselves and the smooth behaviour of the breast outline is well suited to polynomial fitting.
Basic Algorithm

For the row-wise interface, as an example, the basic algorithm can be described by the structure shown in Figure 8.3.

The initial threshold is obtained using the global thresholding method as the threshold midway between that of the near-skin tissue (first iteration of Brink and Pendock minimum cross-entropy) and the global approximation to the skin-air interface (second iteration of Brink and Pendock minimum cross-entropy), these were discussed in Chapter 6. The lowest threshold evaluated is the lowest intensity found in the image, usually intensity 0. Polynomials of order 5 to 10 are fitted for row-wise and column-wise pixels at each threshold. These numbers were determined through experimentation and manual inspection of results.
Extension to Account for Variable Boundary Intensity

Due to the variability of intensity along the skin-air interface, both the row-wise and column-wise interface are split into two overlapping sections, each fitted with a separate polynomial using a separate threshold. Due to the position of the breast in the image, with space on the bottom and the right, one part of the row-wise interface is obtained from the top half of the image, the other part from the bottom two thirds of the image to account for the gap. Similarly, the left half of the image is used for one fit of the column-wise interface, while the right two-thirds form the other part, and example of this division is shown in Figure 8.4.

![Figure 8.4: The image is divided four ways to determine each of the row-wise and column-wise interfaces in two parts, thus lowering the impact of the breast boundary's changing intensity around the image.](image)

The polynomials for the final estimates in each section are only evaluated at the points that correspond to ‘on’ pixels in the thresholded image. Since some thresholds will result in discontinuous lines in the row-wise or column-wise interface, linear interpolation is used to join the polynomial estimates in order to overcome any inaccuracy due to over fitting. An example of the process is shown in Figure 8.5 on image mdb006, for the right column-wise interface. Figure 8.5(a) shows a close-up of the image after zero anchored histogram equalisation to increase boundary contrast – used here only to enable a clearer view and not in the image processing functions. Due to a relatively bright segment of the breast being
present near the edge in this image, the pixels of the right column-wise interface extracted by
the threshold corresponding to the ‘ideal’ polynomial do not include the adjacent darker
section, as shown by their superposition over the histogram equalised image in Figure 8.5(b)
in white. The fitted polynomial evaluated only at points existing in the column-wise interface
with linear interpolation is shown superimposed in Figure 8.5(c).

Figure 8.5: Generating a column-wise boundary (at most one
boundary pixel per column). Mode filtered original shown in (a)
using zero anchored histogram equalisation, boundary pixels from
thresholding superimposed in (b), and optimal polynomial with
linear interpolation in missing columns shown in (c).
Boundary Seeding - Biasing the Estimate Towards Previously Found Boundary Pixels

The polynomial-based algorithm has a tendency to deviate from the actual skin-air interface in certain regions. For example, when the nipple is in profile, quite sharp corners are introduced into the breast boundary by the nipple. An example is shown in Figure 8.6(a) for mdb061. This is a close-up of the nipple, the image has been rotated onto the side to save space, and also to better visualise the fitted function with the x-axis on the horizontal axis. This rotation is not just for visualisation, it is also implemented in the algorithm to enable the re-use of the column-wise function code for the row-wise interface. As the polynomial is more suited to relatively smooth curves, the pixels belonging to the nipple will be relatively far from the estimate and the mean square error between polynomial and breast boundary pixels increases. For the example image in Figure 8.6, a close-up of the nipple and fitted function corresponding to a threshold at the breast boundary is shown in Figure 8.6(b). Although this is the threshold closest to the skin-air interface, the sharp corners introduced by the nipple lead to a bad polynomial fit, resulting in a high mean square error between polynomial and the edge pixels. The final fitted line, corresponding to the lowest mean square error is shown in Figure 8.6(c). Although this lies internal to the breast, the edge pixels in this threshold do not include the nipple and thus produce a better fitting polynomial. Besides the presence of the nipple, noise in the image background near the skin-air interface could mean that the relatively smooth contours internal to the breast will result in lower error. This problem is overcome in part by including already determined sections of the skin-air interface with the boundary pixels currently being evaluated. This procedure will be referred to as boundary ‘seeding’.
Figure 8.6: Problems with sharp corners when seeding not used. (a) shows a close up of the original turned onto its side to examine the row-wise interface. (b) and (c) show results of two thresholds with edge pixels from thresholding in white and fitted polynomial in red. In (b) the edge pixels are closer to the actual breast boundary, however the complexity introduced by the nipple leads to a bad polynomial fit. The threshold in (c), which does not extract the nipple, results in a lower mean square error between the polynomial and breast edge pixels, even though the polynomial lies internal to the breast.

As the skin-air interface divisions overlap, combining a previously determined section with another estimate may lead to more than one y coordinate for some x coordinates, not satisfying the row-wise/column-wise condition. To overcome this, pixels from either the fitted line or current threshold line are removed such that a pixel at each x coordinate is derived from one estimate, the next pixel is derived from the other – this once again satisfies the one y for every x property. The example from Figure 8.6 is shown again in Figure 8.7 to see the improvement in the boundary estimate. Figure 8.7(a) shows boundary pixels from
the top row-wise and right column-wise skin-air interface in white. These pixels are included as additional boundary pixels to guide the estimate. Figure 8.7(b) shows the addition of the boundary pixels to the estimate that was previously judged as the best in Figure 8.6(c). The lines do not overlap and thus produce a high mean square error. Figure 8.7(c) shows boundary pixels closer to the skin, here again as in Figure 8.6(b), the nipple causes a higher fitting mean square error where it occurs but the correspondence of the boundary pixels with those determined previously causes an overall lower mean square error, allowing this boundary to be chosen in preference to others.

Figure 8.7: Boundary Seeding: Using previously found sections of the skin-air interface (a) to bias the final estimate towards an accurate boundary. Contours internal to the breast (b) will result in higher mean square error between polynomial and boundary, as the fitted function will sit between the currently determined boundary and the boundaries determined before. At thresholds near the skin-air interface (c), boundary pixels overlap those determined before resulting in a better overall fit for the polynomial despite a miss-fit existing in the place of the nipple.
**Order of Extraction**

The right column-wise interface was found to be least affected by noise, it is therefore extracted first and used to find the rest of the column-wise interface using the boundary seeding method.

The bottom row-wise interface is the most problematic as it may contain the nipple in profile. To help guide the algorithm, seeding must be performed at the top and bottom of this section. The right column-wise interface is first converted from a column-wise interface, where there is at most one pixel for each column, to row-wise, where at most one pixel for each row (in the right-way-up image) exists. This is done by discarding all but the right-most pixel in each row. This provides a desired skin-line near the bottom of the breast in row-wise interface form. A preliminary estimate of the top row-wise interface is obtained using the converted right column-wise interface for guidance. Since these two regions are relatively far apart, the result may be inaccurate and linear interpolation of found points is not used, as it would lead to a straight line cutting off the tip of the breast. The bottom row-wise interface is then found using boundary seeding with the top row-wise estimate and the converted right column-wise interface.

Once the bottom row-wise interface is found, it is used to find a more accurate top row-wise interface. The final extracted segments from MIAS image mdb147 are shown in Figure 8.8, with (a) showing the original image, and (b) – (e) showing sections of the fitted polynomials within the boundaries of the image divisions with: (b) top row-wise, (c) bottom row-wise, (d) right column-wise, and (e) left column-wise skin air interface segments.
Figure 8.8: (a) Original image (mdb147), with sections of skin-air interface as follows: (b) top row-wise, (c) bottom row-wise, (d) right column-wise, and (e) left column-wise.
Additional Processing to Improve Robustness

Unfortunately, the simple algorithms described above need help in order to work with some images due to several factors. Some of the pre-processing steps, such as orientation, have been outlined already in the section on the row-wise and column-wise interface.

Methods Presented Previously

The removal of tapes is performed after the contrast enhancement of the next section due to the edge-shifting properties of the mode filter that is used there. The edges of tapes are found using the Radon transform based algorithm presented in the last chapter. The area of the tape is removed from processing. However, this introduces a straight-line boundary for part of the skin-air interface and so the start row of the tape is saved in order for the algorithms to follow to be aware of it.

Mode Filter for Breast-Background Contrast Enhancement

The difference in intensity between the breast at its edge and the image background is often minimal. Median filtering is a popular technique for increasing the contrast of edges, however a sharper edge can be produced using the mode filter, as the mode is by definition the most likely value in a region. In both median and mode filtering, a sliding window is applied to the image, and the pixel at the window centre is changed to an intensity corresponding to either the median intensity or mode intensity (most common intensity). One disadvantage of the mode filter is its sensitivity to noise, especially with limited window sizes, while this can be alleviated by using the truncated median filter of Davies [141] as an approximation to the mode, however adequate and faster results are obtained using the normal mode, taken from a 3x3 neighbourhood as the local histogram bin with the highest level of pixels.

An important effect is that the mode filter displaces edges. This is not a well-known property, but can be illustrated modelling two adjacent regions by Gaussian distributions, as was done in Chapter 5 when evaluating thresholding algorithms. For a local window containing both regions, the mode will ideally be the mean of one of the regions (due to the shape of the Gaussian). The most common intensity (highest Gaussian) depends on the proportion of the two regions included in the window. Ideally, the mode will correspond to the region with the most pixels, so that the filtered value is set to the mean of the region on the centre pixel, for a simple straight-line boundary. This is the case only if the two regions are of identical
standard deviation. If the pixels in one region have a greater spread, then statistically more pixels from that region are needed to achieve the same frequency of any particular intensity than from a region of lower standard deviation.

Due to the texture of the breast, its pixels have a higher standard deviation than the relatively uniform (low standard deviation) background; hence the breast shrinks in the filtered image. This is corrected for in the final stage where the binary breast mask is dilated using a 3x3 neighbourhood.

**Error Scaling**
To help steer the fitted boundary towards the skin-line and away from the contours internal to the breast, the error between fitted polynomial and edge pixels of the binary thresholded image is scaled to introduce a bias towards lower intensities. The mean square error corresponding to each intensity is scaled by subtracting from the error the distance between that intensity and the maximum intensity in the image, multiplied by a scaling constant. The scaling constant used has been determined as 0.1 through interactive experimentation, performed by examining the error function in images where an internal contour was extracted in error.

**Missing Boundary Location**
The breast boundary at the top or the bottom may not be detected in some images due to high levels of noise present in those areas. Accounting for tapes, if enough of a boundary is missing, a further iteration of the line-fitting algorithm is run in the area, using seeding with the adjacent detected boundary to guide it. This process of fitting more lines in smaller areas is iterated until either the result converges (same estimate being produced each iteration) or the missing skin-line is of small enough size. Two extra areas are produced this way to add to the already found skin-line sections. These are referred to as the extra column-wise interface and the extra row-wise interface, their existence and dimensions depending on the existence and length of the missing boundary.

**Generating the Final Breast Boundary**
So far, separate breast boundary estimates have been generated around the actual boundary of the breast. To combine these estimates into the complete breast boundary two issues must be resolved, combining estimates in row-wise and column-wise form into a single
representation, and choosing the best estimate where more than one skin line estimate exists for a single section of the breast boundary.

**Eliminating Duplicate Skin-Line Estimates**

A simple heuristic is used to choose the best estimate out of two breast boundary sections. This is discussed first as it applies to both row-wise and column-wise boundaries. In some cases, the estimate of the breast boundary at the end of a section deviates into the breast tissue, or into a noisy area of the image background (the number of such cases is reduced significantly due to the ‘seeding’ of boundaries). As the sections overlap, an estimate of another section may provide a more accurate boundary. Two examples are shown in Figure 8.9, with (a) and (b) containing two zero-anchored histogram equalised images with all row-wise and column-wise boundaries superimposed. Close ups of the overlapping sections are shown inset. The overlap highlighted in Figure 8.9(a) occurs between the top row-wise interface which deviates into a noisy background region and the extra row-wise region. The overlap in Figure 8.9(b) is more complicated as it occurs between a column-wise and row-wise interface (this is not apparent from the image as it shows all boundaries superimposed in white).
Figure 8.9: Two images, mdb036 (a), and mdb176 (b) with all row-wise and column-wise boundaries superimposed in white. The area in the green circles is shown magnified in (c) and (d) as indicated by the arrow. The final fitted line, superimposed in red, is shown below the magnified section in (e) and (f).

Estimates from adjacent sections (e.g. column-wise 1 and column-wise 2) are combined to form a single line using a function. This function was written and is described in terms of combining two column-wise boundaries, however the same procedure is used to combine the row-wise boundaries after first rotating them into a column-wise representation. For columns where only one estimate generates a boundary, there is no conflict and such sections are used as the final boundary returned by the function. Sections of boundary where both estimates exist and are different are identified and the final boundary in those sections is determined heuristically by examining the boundary columns just before the start and just after the end of each such section. In the boundary column, one of three conditions must be true, 1) no estimate exists, 2) one estimate exists, or 3) both exist and are equal. Depending on the combination of these states for the column before and column after the section, a
single estimate or the mid-point in each column between the estimates is used, the rules summarised by Table 8.1.

<table>
<thead>
<tr>
<th>Column Before Overlap</th>
<th>Column After Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Estimate</strong></td>
<td>(a) Mean Estimate 1  Estimate 2 (d) Mean</td>
</tr>
<tr>
<td><strong>Estimate 1</strong></td>
<td>Estimate 1 Estimate 1 (b) Mean Estimate 1</td>
</tr>
<tr>
<td><strong>Estimate 2</strong></td>
<td>Estimate 2 (b) Mean Estimate 2 Estimate 2</td>
</tr>
<tr>
<td><strong>Both Equal</strong></td>
<td>(d) Mean Estimate 1 Estimate 2 (c) Mean</td>
</tr>
</tbody>
</table>

Table 8.1: Rules for determining which breast boundary estimate from two column-wise estimates is used where two possible boundaries exist for a section. The final boundary is determined by which of the estimates exist before and after the overlap occurred, resulting in either one of the estimate boundaries or the mean between them.

As the table shows, the mid-point, or mean between two estimates is used when either (a) no estimate exists before and after the overlap, (b) where one estimate exists before and the other after the overlap (the mean smoothly merges the two estimates), (c) both estimates are equal at both ends (a loop exists), and (d) a combination of equal estimates on one end and no estimate on the other. As an example, the overlapping section shown in Figure 8.9(c) is shown resolved in Figure 8.9(e), the boundary superimposed in red.

*Generating a Single Boundary for the Breast*

The procedure for combining two estimates with overlapping sections, as presented in the previous section, allows for the generation of a single column-wise interface, combining the left, right, and extra interfaces, and the generation of a single row-wise interface by combining the bottom, top, and extra row-wise interfaces. The combination of these resulting row-wise and column-wise interfaces is more complicated due to the different representations.
Before a single breast boundary is generated, the column-wise interface must be combined with the row-wise interface. This is done to enhance accuracy, so that when an overlap occurs it can be resolved, resulting in the two interfaces meeting. Combination is performed in two steps. The row-wise interface is converted to column-wise form and combined with the column-wise interface. The column-wise interface is likewise converted to row-wise form and combined with the row-wise interface.

Converting the row-wise interface to column-wise form is done simply by choosing the bottom most pixel in each column, however the area where this is done must be constrained so that the skin-line from the top of the breast is not extracted. Figure 8.10 illustrates this problem, showing the row-wise interface from image mdb133 in (a) along with its conversion to a column-wise interface and the boundary lines of constraint shown in (b), the valid column-wise line segment indicated as (i). Two problems arise if the boundaries are not applied. The first, where no boundary line exists at the bottom of the breast the bottom pixel in these columns is from the breast boundary near the top, shown in (ii). Second, in the region where there is a boundary near the bottom of the breast it is often not continuous, since each row in the row-wise image is allowed no more than one pixel, this results in columns which are without a pixel in places where the boundary is closer to horizontal. In these columns the boundary from the top is again extracted, shown as region (iii). Similar problems exist when converting the column-wise interface into row-wise form.
The row/column constraints shown in Figure 8.10 are derived as follows. The bottom-most row containing the row-wise boundary is found and the column where that boundary pixel exists (only one pixel will exist in this row by definition of row-wise image) used as the left-
most column for the conversion. Next, the right-most column where a pixel exists in the row-wise interface is found and the bottom pixel in this column marks the row above which the other boundary pixels will be ignored. The equivalent constraints are used for the column-wise interface.

Once these constraints are imposed, the bottom-most pixel in each column within the valid range is then used in the conversion to column-wise. The right-most pixel in each row from the column-wise interface within the constraints is used to produce a boundary in row-wise form.

The problem when the boundaries are converted is that gaps will exist. For example, the row-wise boundary has one pixel for every row along its boundary (segments were joined using the linear interpolation discussed earlier). However when the row-wise interface is converted to column-wise, each column is not guaranteed to have a pixel in it (see the converted estimate in Figure 8.10). If the breast boundary from the column-wise interface does not overlap to cover these gaps in the converted row-wise interface, gaps will result in the final skin-air interface due to the way it is generated. Rather than using further interpolation to close the gaps in the converted boundaries, the left and right column extremities of the column-wise interface are found and pixels in the converted boundary from the row-wise interface outside these extremities are eliminated. The same is done for the top and bottom extremities of the row-wise interface, with pixels from the converted column-wise interface set to zero outside them.

The row-wise and converted column-wise interfaces are then combined, as are the column-wise and converted row-wise interfaces. The example in Figure 8.9(d) where there is an overlap of estimates is shown resolved in Figure 8.9(e) using the same rules as before, the final boundary superimposed in red. If the skin-line does not stretch to the top of the image or to the bottom left, points at the top and bottom left are extrapolated using a straight line. Another option would be not to extrapolate at the top, as the missing section is usually attributed to the presence of a tape however the final use of the breast boundary would have to determine whether the sections affected by tapes are included or not.

This results in a ‘final’ row-wise interface and a ‘final’ column-wise interface. To combine these two into a single boundary, solid masks are produced by filling each column of the
column-wise interface up to the boundary pixel and each row of the row-wise interface up to its boundary pixels, these are the known portions of breast tissue. Next, the pixels of the column-wise mask above and to the right of the right-most part of the boundary are set to 1 as are the pixels of the row-wise mask that lie below and to the left of the bottom most point on the boundary, these are ‘possible’ regions of the breast. The two resulting masks are combined using the AND operator, the ‘possible’ breast regions included in each mask to prevent cut-off of the breast in these areas when the AND operator is applied. The two masks are shown in Figure 8.11 for image mdb133, with row-mask in (a), column-mask in (b) and the final breast mask in (c). To compensate for the standard deviation-based eroding effect of the mode filter used in contrast enhancement, the final breast mask is dilated using a 3x3 neighbourhood to return it closer to its original size. The histogram equalised original image and one with the final breast border determined as the outer pixels of the final breast mask are shown side by side in (d) and (e).
Local Threshold Based Skin-Air Interface

The second algorithm to be discussed was conceived to overcome the limitation of global thresholding, caused by the changing intensity of the breast edge along its length, and was
devised before the polynomial fitting approach. The theory of the local threshold-based algorithm was introduced previously in [132], and an extension which makes use of the approximation from the global threshold in presented in [142] and will be described here. Like the polynomial fitting technique, the local thresholding technique relies on splitting the skin-air interface into sections along the row-wise and column-wise skin-air interface, although each local area is smaller.

An approximation to the row-wise and column-wise breast boundary is made as the initial step. This is performed by edge enhancing the original image using the Prewitt edge operator and then thresholding at a low threshold close to zero. This relies on the property that internal to the breast lie a lot of structures such as the different vessels that have relatively strong edges, while the noise outside the breast is more randomly distributed and does not present such strong edges, similar to the observations of Highnam and Brady in [55] about high gradient in the breast and low in the background. The result can be seen in Figure 8.12.

![Figure 8.12: MIAS image mdb070 is shown in (a) with Prewitt edge enhanced image used to find the skin-air interface approximation shown in (b).](image)
Next a rectangular sliding window is used, 20 rows by 80 columns, slid across each row for the row-wise interface and 80 rows by 20 columns and slid down each column to determine the column wise interface. In a row, the row-wise skin-air interface approximation is the point where the difference between the number of white pixels to the left of the windows centre and the right is greatest. For the column-wise interface the difference between the top and bottom half is a maximum at the breast edge. This detection of the breast edge is similar to the gradient-based approach, discussed in the introduction to this chapter.

The approximated skin-air interface is refined at each point in the row-wise and column-wise interface by generating several candidates for the edge at each point. One candidate is generated from the result of global thresholding, using the second iteration of non-symmetric Brink and Pendock minimum cross-entropy, as discussed in Chapter 6. The global threshold does not depend on the approximation from the edge enhanced image discussed in the previous paragraph and may not lie near it. Other candidates for a point are obtained by applying the non-symmetric minimum cross-entropy thresholding algorithm by Brink and Pendock [101] (as discussed in Chapter 5) to a local area centred near the approximated breast boundary and determining a breast boundary candidate from the result – either as row-wise or column wise. The local area is the same size as the sliding window used to generate the approximation, in the case of the row-wise interface; the window slides centred down the row-wise approximation interface pixels. Since the local window encompasses 20 rows, the threshold produced from a single window position provides one candidate for each of the rows present. As successive windows overlap, several candidates for the position of the boundary in each row are generated. The same is done for the column in the column-wise interface.

The candidate positions for the boundary pixel in a row are grouped along with the one from the global estimate. The standard deviation of these is computed and if it is higher than a set value (determined as 5 through trials) it is concluded that there is some disagreement in the candidates. If parts of the skin-air interface have already been determined in the proceeding rows, these are called on to resolve the disagreement. The elimination of points using previous data is only performed for candidate lists with a high standard deviation. This is because an approximation using the gradient of the previous points is valid only for relatively small areas close to the points for which the gradient is determined. Letting it guide the algorithm for too many successive points will result in errors. As there are usually some
discrepancies between candidate points, a threshold must be set such that previous data is considered only when the standard deviation of a set of points is greater than the threshold.

Since the skin-line is relatively smooth and continuous, the relative direction of the skin-line should not change dramatically in a local area. For the row-wise approximation, the gradient of the previous five row pixels is determined as the difference between pixel positions in adjacent rows. These gradients are weighted so that the gradient of the pixel in the row nearest to the one currently being determined is most significant. From this, the likely location of the pixel in the current row is determined. All candidates outside a 20-pixel range of the likely location are eliminated; if no candidates are left, then the likely location is used as the boundary pixel. The same procedure is performed on the column-wise interface except that the boundary position in each column is considered.

Regardless of whether candidates have been eliminated using the approximation from the boundary gradient, further elimination is performed by looking for outliers. Points located more than two standard deviations from the mean are eliminated and a new mean determined and set as the breast boundary for that row or column.

As in the case of the polynomial fitting algorithm, binary masks are generated from the resulting row-wise and column-wise interfaces and these are combined using the AND operator to produce a mask for the entire breast.

The local and global thresholding techniques complement each other when they are combined. Figure 8.13(a) shows MIAS image mdb070 with the skin-air interface obtained from the global threshold. At the bottom of the image the breast border is accurate, however due to the changing intensity of the breast boundary across the image the same threshold extracts the noise in the circled section. Figure 8.13(b) shows the same mammogram, using the procedure outlined above but with global thresholding estimate excluded from the candidates. Here the problem lies in the column-wise interface, going from the bottom left of the image. The estimate goes astray due to a boundary internal to the breast and then gets lost, a weakness of a purely local approach which is sensitive to slight changes within a local area. Figure 8.13(c) shows the complete approach combining local and global estimates. The global error at the top of the image is resolved as are the majority of the local method errors, although a slight dip is noticeable where the local algorithm went
astray due to averaging, the combination with the global estimate leads the boundary back to the edge of the breast.

Figure 8.13: (a) Histogram equalised original with skin-air interface from the global threshold outlined and noise at the top of the image classified as part of the breast circled. The skin-air interface using only the local thresholding estimates is shown in (b). The column-wise interface has gone astray (circled) due to local variations. Adding the boundary from global thresholding to the local estimates results in (c), significant errors from the individual estimates have been removed.

Results

In viewing the final segmentation results, the skin-air interface from the polynomial based algorithm is clearly closer to the actual breast boundary than the skin-air interface obtained from the local threshold based algorithm. This occurs because the local thresholding algorithms, while trying to impose continuity by using prior points along the skin line, it cannot match the smoothness that is inherent to a polynomial fit. Comparing the outline for MIAS image mdb063 in Figure 8.14, the breast boundary determined by the local algorithm in (a) is clearly rougher than the one determined by the polynomial based approach, shown in (b). The skin-line is also compared to that from Chandrasekhar’s algorithm in (c).
Another weakness of the local threshold approach is that the skin air interface is determined from top to bottom, row by row for the row-wise interface and left to right, column by column in the column-wise interface. This means that a bad fit at the top or left of the image due to noise can affect the rest of the estimation process as previous points are in cases used to determine next points and there is no refinement of the points already determined. This is alleviated to some extent by only calling on the previous points if there is major disagreement between the current points estimates but still causes problems.

If viewed on a time line, the local threshold based algorithm initiated the concept of the row-wise and column-wise skin-air interface. Discussion will however centre on its successor the polynomial based algorithm due to its superior results.

**Properties of the Polynomial Based Boundary**

The most notable property of the polynomial based boundary is the smoothness it enforces despite local variations. This has two major effects:
1. Tolerance to incomplete breast boundary.

2. The nipple is ignored.

An example of the first effect is shown in Figure 8.15 on two zero-anchored histogram equalised images. In mdb071 (a) and (b), the breast border is missing some pixels. The polynomial-based result ignores these gaps and the fitted line runs smoothly around the breast where the boundary should be, shown in (a). In contrast, Chandrasekhar’s algorithm follows the rough boundary. In mdb146 (c) and (d), the skin boundary is missing from the image, leading to a ragged edge. The polynomial fits smooths the boundary (a), while Chandrasekhar’s algorithm again extracts a rough line.

![Figure 8.15: Enforced smoothness of polynomial boundary (a) and (c) compared to Chandrasekhar’s method (b) and (d). For missing sections of skin-line, as in mdb071, the boundary is interpolated by the polynomial (a), while for images with large sections of missing skin-line, as in mdb146, the boundary is smoothed (c). In contrast, the rough boundary is followed by Chandrasekhar’s algorithm in (b) and (d).](image)
The second property may or may not be beneficial. When the nipple is visible on the boundary of the breast the polynomial based algorithm ignores it, as the nipple presents some relatively sharp corners. If this is desired depends on the final application of the skin-air interface. To examine internal breast tissue or points along the skin-line, the nipple may not be important. Chandrasekhar [89] specifically states in his method for segmenting the nipple that when it is in profile, the nipple should be segmented as part of the skin-air interface. The nipple when it appears in profile presents as quite an obvious partially circular region adjacent to the determined skin-air interface and it should be straightforward to enhance the algorithm to detect it, this is however not approached here.

Accuracy of the Fitted Boundary
Through manual inspection by the author, approximately 75% of images in the MIAS database were fitted with a breast border that showed no obvious imperfection. For images where deviation from the actual skin-line is observable, it ranged from slight localised misalignment to a small percentage of images where it can be considered that the algorithm has failed. This section will describe the types of error introduced and the next section will examine the performance in context with an established skin-air interface extraction algorithm to show that the level of errors present is in comparison not significant.

Sub-optimal performance of the polynomial algorithm can be put into several classes, based on seriousness and cause. The classes identified are:

- Unacceptable deviation outside the breast (2 images)
- Slight deviation outside the breast (7 images)
- Contour internal to the breast (41 images)
- Straight line cut off (11 images)
- Bad fit from the ‘extra’ row/column–wise interface (20 images)

Unacceptable deviation outside the breast
Bright regions in the background that fail to be eliminated draw the fitted boundary outside the breast, leading to major error in two images, mdb260 and mdb280 shown in Figure 8.16.
This is in many ways the most serious error as a large proportion of background is classified as breast, making any measures applied to breast tissue (such as percentage of dense tissue) or skin-line evaluation for registration/asymmetry detection impossible. Although this is a serious deviation from the breast boundary it only affects 2 images, less than 1% of the database.

Figure 8.16: The two images with large deviation of the fitted boundary outside the breast tissue. (a) shows image mdb260 and (b) shows image mdb280, both with the skin-air interface superimposed.

_Slight deviation outside the breast_

Less serious deviation to areas outside the breast occurs in seven images. To be put into this class an obvious deviation from the actual skin-air interface must be present. These do not include images where the boundary is off by a few pixels due to poor definition of the skin-air interface. In four images, this occurs due to noise either at the top or bottom of the image, shown in Figure 8.17 for a portion of the image near the bottom of mdb132 in (a), and a section from the top of mdb068 in (b). For two images this occurs due to the presence of a label near the breast. Figure 8.17 also shows these two images, (c) mdb006, outline shown in black, where the label overlaps the breast and was therefore not removed in the pre-processing, and (d) mdb152, where noise due to the label overlaps the breast. In image md170, shown in (e), the sharp angle of the breast, combined with missing pixels causes a slight deviation of the fitted line outside the actual breast boundary.
Contour internal to the breast

The fitted line is placed internal to the breast to an extent in 41 of the database images. A proportion of this error can be attributed to the badly fitted ‘extra’ polynomials and the straight line cut-off which will be discussed in separate sections. Other reasons for this misalignment, examples shown in Figure 8.18, are presence of an overlapping label in mdb006, shown in (a), as well as causing a slight deviation outside the breast, as shown in Figure 8.17(c), causes the fitted line to run internal to the breast as the label breaks the smoothness of the breast edge. Image mdb026, shown in (b), has very poor definition at the
breast edge. Even zero anchored histogram equalisation does not show an obvious boundary (b-i) and an image with all pixels of zero intensity set to white is provided in (b-ii) to better visualise it. The poor definition at the skin line forcing the fitted curve towards a smoother internal contour. A contour inside the breast is also found in image mdb136, shown in (c), where breast is cut off at the bottom of the image. In images mdb237 (d), and mdb134 (e), despite the boundary seeding, the effect of the nipple at the skin boundary still forced the fitted boundary into the breast.

*Straight line cut off*

In 11 images (~3.4% of database) the straight-line interpolation between sections of the polynomial causes parts of the breast to be cut off. In all cases this occurs near the bottom of the breast. Figure 8.19 shows two typical examples from MIAS images (a) mdb069 and (b) mdb293.

*Bad fit from the ‘extra’ row/column-wise interface*

As was explained earlier, further polynomials are fitted when there is a significant gap between the row-wise polynomial and the top of the image, or the column-wise polynomial and the left side of the image. A problem arises with these ‘extra’ polynomials, as there are some images with significant gaps, however the skin-air interface in that gap is poorly defined. The result is often a less smooth breast border at the top or bottom left of the image with a ‘wavy’ appearance, shown through some examples in Figure 8.20. This effect has been noted in 20 of the database images (around 6.2% of the database).
Figure 8.18: Fitted line internal to actual skin-air interface. Instances of this problem occur due to: (a) label overlapping border (mdb006), (b) poor definition of the skin-line – unclear even on zero-anchored histogram equalised image (i) but evident when pixels of intensity 0 are set to 1 (mdb026), (c) breast is cut-off and internal contour extracted (mdb096). Presence of the nipple in (d) mdb238 and (e) mdb134 also lead to internal contours being preferred.
Figure 8.19: Two examples where linear interpolation of the fitted polynomials leads to breast cut off, examples are taken from images (a)mdb069 and (b) mdb293.
Comparison to Chandrasekhar's Results

Several differences are evident when comparing the fitted polynomial results with the results of Chandrasekhar's background subtraction algorithm. The two most distinct are that Chandrasekhar's algorithm extracts the nipple when it appears in profile and that the smoothness imposed the polynomials around the breast boundary results in a less jagged fitted boundary, especially when it is slightly internal to the breast.

The breast border resulting from each algorithm was compared qualitatively by superimposing the border as a 1 pixel wide boundary over the zero-anchored histogram equalised mammogram image. The comparison was performed by the author as a 'rough' indication of comparative algorithm performance. Comparison of the boundaries was performed by superimposing each boundary onto the image and viewing the two images side by side. Comparison was made of the level of deviation from the actual breast boundary on the image. For each image a decision was made as to which algorithm produced better results, if both results were acceptable, or if they were both poor representations of the actual
breast boundary. When comparing results from the two procedures the extraction or failure to segment a nipple in profile was not taken into account when determining accuracy.

The results are summarised in Table 8.2 and in graphical form in Figure 8.21.

<table>
<thead>
<tr>
<th></th>
<th>Number of Images</th>
<th>Percentage of Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both acceptable</td>
<td>195</td>
<td>60%</td>
</tr>
<tr>
<td>Both show error</td>
<td>14</td>
<td>4%</td>
</tr>
<tr>
<td>Polynomial fit better</td>
<td>92</td>
<td>29%</td>
</tr>
<tr>
<td>Chandrasekhar better</td>
<td>21</td>
<td>7%</td>
</tr>
<tr>
<td>Total</td>
<td>322</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 8.2: Performance comparison of the polynomial fitting algorithm with Chandrasekhar’s skin-air interface segmentation algorithm.

Figure 8.21: Graphical performance comparison of the polynomial fitting algorithm with Chandrasekhar’s skin-air interface segmentation algorithm. The graph shows that in 89% of cases the polynomial fitting algorithm performed as well or better than Chandrasekhar’s.
To summarise the results, the polynomial-fitting algorithm provides a better fit in 29% of images and a similar fit in 60%. This means that in 89% of cases the polynomial fitting algorithm performed no worse than Chandrasekhar’s. Chandrasekhar’s algorithm provided a more accurate breast border in 7% of images and both algorithms produced a badly fitted skin-air interface in 4%.

The results indicate that the polynomial fitting algorithm produces a better fit overall. To compare the relative results to a base line, Chandrasekhar reports in [140] that 32 randomly selected results were graded by a radiologist on a 5-point scale ranging from ‘useless’ (1) to ‘excellent’ (5) with a grade of (3) being ‘acceptable’. The radiologist graded 94% of images from the sample as ‘acceptable’ or higher.

Conclusion
This chapter has taken the previously found global threshold set near the skin-line intensities on the mammogram and successfully refined the estimate to obtain an accurate breast boundary. The calculation was simplified by dividing the skin-air interface into two sets of functions, one using the rows as the x-axis, the other using the columns. This simplification led to two methods, one based on local thresholding along the estimate, the other based on an optimal method – where the mean squared error between a polynomial and the estimated outline was minimised by changing the threshold and polynomial order in a given section of the skin-air interface. While the local-threshold technique at times results in a rough breast outline, the polynomial based algorithm produced an accurate skin-line in most images, with results comparing well against an established algorithm.
EXPANDING WINDOW THRESHOLDING FOR DELINEATING THE PECTORAL MUSCLE

In this Chapter, a novel algorithm is presented for selecting a threshold between two adjacent intensity regions. The algorithm works by taking a series of local windows around a point so as to include a changing ratio of the two regions. It is shown that when either a modification of the minimum cross-entropy function or the median function of the pixel intensities in each window is taken, an inflection point in the function as the region ratio changes defines a region of interest and a threshold to separate the two image regions. This algorithm is used to extract the pectoral muscle from mediolateral oblique view mammograms.

For the case of the pectoral muscle, the window is anchored in the pectoral muscle corner of the image, located using the orientation algorithm from Chapter 7. Once a threshold is obtained for the pectoral muscle, a straight line is fitted to the edge to overcome the problem of overlapping glandular tissue.

Automatic evaluation is used to select a best-estimate straight line and to integrate pectoral edge estimates obtained from global thresholding performed in Chapter 6. This automatic algorithm is also used to evaluate the estimate against the work of Kwok et al. [145], and also to integrate the results from the presented algorithms with their results, resulting in a better straight-line estimate.

After pectoral results are presented, to stimulate further work, a generalised form of the expanding window algorithm is presented in the form of semi-automatic segmentation where the user selects a pixel in an image and the algorithm determines an area of interest and threshold for a homogenous intensity region near that pixel.

Introduction

In the mediolateral oblique view, the pectoral muscle appears as a bright triangular region in one corner of the image. Processes that rely on pectoral muscle location include:
• Determining image adequacy [143];

• Matching data from mammogram pairs [29] [144];

• Exclusion of pectoral muscle tissue from further processing [81];

• Calibration of breast tissue intensities [81].

In the work of others, the pectoral boundary has usually been approximated by a straight line as the first step [29],[81],[89],[145]. The straight-line approximation has been determined by a number of techniques, including region growing [29], the Hough transform [89],[81],[144] and local adaptive thresholding followed by line fitting [145]. Once obtained, the straight-line approximation can be refined to follow the slightly curved outline of the pectoral muscle [144],[145] as a further step.

The image segmentation algorithm introduced in this Chapter provides another way to extract the straight-line approximation of the pectoral muscle. The algorithm was originally based on a modified version of minimum cross-entropy thresholding applied to a set of local windows, as first described in [146]. After this publication it was discovered that the median value had similar properties as modified minimum cross-entropy in certain cases while being faster to compute. Both algorithms are introduced and compared in this chapter. A comparison is also presented in [147].

Modified Minimum Cross-Entropy

In [101], Brink & Pendock do not define the mean intensity below a threshold, $\mu_B$, and the mean intensity above a threshold $\mu_A$ for obtaining cross-entropy clearly. Defining $g$ to be an intensity in the image, and $p_g$ to be the probability of that intensity occurring, the means can be defined as follows:

$$\mu_B = \sum_{g=0}^T gp_g$$

(28)

$$\mu_A = \sum_{g=T+1}^J gp_g$$

(29)
Since $\mu_B$ is the intensity mean considering only pixels below the threshold, $p_g$ is taken as the probability of pixel intensity $g$ within those pixels below the threshold.

$$p_g = \frac{f_g}{N_g}, \quad (30)$$

where $f_g$ is the number of pixels of intensity $g$ and $N_g$ is the total number of pixels below the threshold. Likewise, $p_g$ must only consider pixels above the threshold to determine the correct probability for determining $\mu_A$:

$$p_g = \frac{f_g}{N_g}, \quad (31)$$

These definitions of the means are equivalent to those given by Li and Lee in [100].

The formula for the probabilities given by Brink & Pendock in [101] can however be interpreted, due to their notation, as including all the image pixels in the probability calculation.

$$p_g = \frac{f_g}{N}, \quad (32)$$

with $N$ being the total number of pixels in the image. In [101] $p_g$ is substituted for both $p_g$ and $p_g$ in (28) and (29), making the new quantities $\mu_B$ and $\mu_A$:

$$\mu_B = \sum_{g=0}^{T} p_g \quad (33)$$

$$\mu_A = \sum_{g=T+1}^{J} p_g \quad (34)$$

As $p_g$ uses the total pixel number, $N$, instead of the number of pixels in the local region, (33) and (34) represent the region illumination divided by $N$, not region means, $N$ being a constant for all thresholds.
Graphing the resulting mean values for several thresholds in an image allows the comparison between the correct mean formulas for $\mu_B$ and $\mu_A$ along with those determined by using $p_{g,\text{real}}$. Figure 9.1 shows such graphs, produced using a mammogram image. For the area below the threshold, the mean and Brink's quantity follow a similar trend. Above the threshold however, the quantities diverge. Conceptually, as the threshold increases, pixels of lower intensity will be excluded from the area above the threshold. The exclusion of lower intensity pixels will lead to the average intensity of that area, $\mu_A$, increasing. Through this comparison it is obvious that $\mu_{\text{Brink}}$ does not represent the mean $\mu_A$, hence, use of $p_{g,\text{real}}$ does not lead to a probability distribution and Pal's argument from [102], that the algorithm is not minimum cross-entropy would hold (see discussion of Pal’s criticism in Chapter 5). The Brink and Pendock cross-entropy algorithm, when $p_{g,\text{real}}$ is used to work out the quantities $\mu_B$ and $\mu_A$ will hence forward be referred to as the modified minimum cross-entropy thresholding algorithm, or MX, with the resulting threshold referred to by the term MX value.

Checking results presented in [101] indicates that Brink and Pendock did use the correct mean formulas, utilising $p_{g,\text{h}}$ and $p_{g,\text{a}}$, and not $p_{g,\text{real}}$. The error was thus only introduced in the editing with results presented in [101] being representative of minimum cross-entropy
thresholding. Nevertheless, MX has useful properties of its own, in some ways similar to that of the median value in an image for the sub-class of bi-modal images.

![Graphs showing MX Intensity Threshold and Median Intensity Threshold for different distributions.](image)

**Figure 9.2:** Mixtures of two Gaussians with means of 50 and 150 and identical weighting (area under curve) in (a) and (b) and unequal weighting in (c) and (d). The threshold produced by the minimum cross-entropy algorithm is shown superimposed. The threshold is pulled towards the peak with the highest weighting, regardless of variance differences in the two Gaussians.

**Properties of MX and Median Values in Bimodal Images**

Several observations can be made on the MX and Median intensity values of a synthetic histogram composed of a mixture of two Gaussians. Such a histogram is a model for a two-class image, with one peak for the background and one for the object, as used in Chapter 5 in thresholding algorithm evaluation. When the Gaussians have the same weighting, this represents the same number of background as object pixels for the model image, in this case both the MX value and Median intensity correspond to a threshold at the midpoint between the two peaks. Examples of thresholding are shown in Figure 9.2 where accuracy is shown to hold in the case where the standard deviation is the same for both Gaussians (a) and
where it differs (b). When a weighting is applied to one Gaussian, the algorithms result in a value closer towards the mean of the dominant Gaussian. This is shown for two examples in Figure 9.2(c) and (d). If the chosen threshold is plotted as a function of the ratio of the two Gaussians, a smooth curve can be observed. A graph of this behaviour for a mixture of two Gaussians $P_0$ and $P_1$ with the same standard deviation, mean of $P_0$ equal to 50 and mean of $P_1$ equal to 150 is shown in Figure 9.3. This graph shows the threshold tending towards the mean of the dominant Gaussian. The inflection point of the graph corresponds to a threshold that separates the Gaussians, shown circled. It can be seen from the graph that the Median produces a sharper inflection point, its location at the mid-point intensity between the two Gaussian means while the MX inflection point corresponds to an intensity closer to the high intensity Gaussian. The inflection points are located approximately at the place on the graph where the Gaussian weightings are equal, again the Median being more accurate.

![Figure 9.3: Graph showing the progression of the MX and median intensities as the ratio of the two Gaussians in the histogram changes. In each case, the threshold intensity tends towards the mean of the dominant Gaussian with a point of inflection occurring near the point where the Gaussians are of the same weighting.](image)

The MX characteristics that lead to the behaviour of its minimum are examined in Figure 9.4. The figure shows four pairs of graphs (a)-(d) with a synthetic histogram in each case shown
above a graph of modified cross-entropy. The intensity where the minimum modified cross-entropy occurs is indicated by a stem. The graphs show a progression, where the weighting of the low intensity Gaussian is increased from (a) to (d), showing the effect on the modified cross-entropy and its minimum.

For histograms where the high intensity peak is dominant, a narrow minimum in the MX graph results in a region corresponding to the high intensity peak. The location of the minimum is at the mean of the high intensity peak when the weighting of the low intensity peak is zero (a).

As the weighting of the low intensity Gaussian increases, a dip occurs in the MX function corresponding to its intensities (b), this grows in prominence as the weighting of the low intensity Gaussian increases. When the weighting of the low intensity Gaussian is greater than the high intensity Gaussian, the dip it causes in the cross-entropy function becomes a narrow minimum. The location of the minimum converges towards the mean of the low intensity Gaussian as its weighting increases (d).

The middle ground between the two narrow minima is a broad minimum occurring when the Gaussians are equally weighted. The minimum corresponds to a threshold to separate the two classes (c).
It will be shown that for an image with two adjacent regions with different intensity distributions, as in the case of the pectoral muscle, a graph similar to that in Figure 9.3 can be produced by thresholding in a local window, anchored at a point, with its size varied to include different ratios of background to object pixels.

Bi-Modal Model for Pectoral Muscle Extraction

As was mentioned in Chapter 6, a mammogram image can be divided into four regions based on tissue type. Region outlines for MIAS image mdb007, in Figure 9.5(a), are shown in Figure 9.5(b) (figure repeated from Figure 6.5). The positions of the regions are indicated on the image histogram in Figure 9.5(c). The regions are separated in the image but overlap in the histogram; they correspond to (i) the pectoral muscle, (ii) fatty breast tissue, (iii) denser glandular tissue and (iv) the image background.

![Image](a) ![Image](b) ![Image](c)

Figure 9.5: MIAS image mdb007 (a) has been decomposed into four tissue type regions (b). The image histogram is shown in (c) with approximate region position shown. The regions correspond to (i) pectoral muscle, (ii) fatty tissue, (iii) denser tissue and (iv).

The pectoral edge represents a line between the pectoral region and the fatty breast tissue. A sub-image can be taken from the top-left corner that will contain only those two regions, excluding the denser glandular tissue and the image background. Such an image is shown in Figure 9.6(a). Its histogram in Figure 9.6(b) contains two peaks corresponding to the pectoral muscle and fatty tissue regions. This shows the segmentation of the pectoral muscle to be a two-class thresholding problem for a subset of spatially separable image pixels. The problem is investigated using the two-class model image shown in Figure 9.6(c). A bright region representing the pectoral muscle, Figure 9.6(i), is set on a dark background, Figure.
9.6(ii), for both mammogram (a,b) and model (c,d) images. In the model, the fatty breast region has been extrapolated to remove other non-pectoral image regions. A Gaussian intensity distribution has been used for both regions. The model image histogram is shown in Figure 9.6(d). The four parameters used to define this model (and this specific model image) are: a) mean intensity of pectoral muscle, or mid-point of the pectoral histogram peak (150), b) mean intensity of the model background histogram peak (50), c) the standard deviation of both Gaussian distributions (10) and d) the aspect ratio of the rectangle enclosing the pectoral muscle (4:7), the aspect ratio being defined here as ratio of width to height. Knowing the aspect ratio, it is adequate to state one dimension to specify a window size.

The MX value and Median intensity of several rectangular windows anchored to the top-left image corner are computed. The windows are of fixed aspect ratio and varying window size, increasing by an increment of $\Delta$. A wide range of aspect ratios has been tried on the model with similar results. Due to the exclusion of mammogram regions (iii) and (iv) in Figure 9.6(d) from the model, the window aspect ratio for a mammogram must be set to exclude these regions from the window for as many window sizes as possible. A useful aspect ratio is that of the window containing the entire pectoral muscle, this however varies between mammograms. An aspect ratio of (1:2) has been chosen here to demonstrate results on images with a variety of pectoral aspect ratios, including the model image and actual mammograms.
The characteristics of the cross-entropy minimum for varying window sizes and the behaviour of the median intensity in each window are summarised by the graph of MX and median values versus window size shown in Figure 9.7, which looks similar to the graph in Figure 9.3 as the same parameters were used for the Gaussians. The two regions converging to mean pectoral and mean background intensities are divided by a point of inflection, shown in the graph; in the case of the MX value, this point corresponds to the broad minimum occurring in Figure 9.4(c). The location of the inflection point has been determined by examining the first and second derivatives of the function. Because of intensity fluctuations in the model image, inflection points are a little off from the ideal values determined in the synthetic histogram. This time the MX value is closer to the threshold midway between the two Gaussians (100).
Figure 9.7: Using the model image in (a), MX values and median intensities of local windows are taken, the values shown in (b) as a function of window size. A similar trend is observed as for the model histograms.

Figure 9.8: Threshold intensities corresponding to MX value and median are taken from the mammogram in (a) as they were from the model image in the previous figure (same colour coding used). The graph in (b) shows similarity to the model image graph in the region marked by the dashed rectangle, points of inflection near window size 30 correspond to window sizes where approximately the same amount of pectoral tissue as background tissue is included.
Comparing this model to an actual mammogram, Figure 9.8(a) shows the threshold versus window width graph for MIAS image mdb207. The shape of this function is a close approximation to the shape of the model function for small window sizes, indicated in Figure 9.8(b) by the box, with behaviour of the two functions varying when other regions are included in the window. The inflection points, found from the first and second derivative, are shown in the figure, deviations from the model causing the occurrence of more than one point of inflection.

An inflection point in the graph provides two pieces of information: it gives an intensity that divides the two regions and also indicates a window size where the two regions are of similar proportion, providing an “area of interest”. The border of regions extracted using the inflection point threshold is shown in Figure 9.9(a),(b) for the model and Figure 9.9(c),(d) for the mammogram, using the inflection point near window size 30.
In order to determine thresholds without user intervention, the inflection points must be found by an automatic procedure. This is accomplished by fitting a series of quadratic functions to a sliding neighbourhood on the threshold versus window width function. First of all, this smooths the function. Second, points of inflection are found where the $x^2$ term of the quadratic changes sign. If the quadratics on either side of a point are taken as parts of a piecewise quadratic function the change in sign between them corresponds to a change in the direction of concavity, corresponding to zero second derivative.

The size of the sliding neighbourhood for point of inflection location is another parameter. If the neighbourhood is too small, erroneous inflection points will be found but if it is too
large, actual inflection points may either be missed or an error introduced in their location due to interpolation. To further complicate the size selection, the number of function points included in each neighbourhood is dependent on the resolution of the window size axis, corresponding to the $\Delta$ factor that the window expands by. Since a variety of $\Delta$ are tested, an adaptive solution has been implemented. The sliding neighbourhood size is initialised as a low value so as to initially find several inflection points. The neighbourhood size is then decreased until fewer than three inflection points are found. If needed, the size is then increased again so that at least one inflection point is found. In an ideal image the inflection point corresponding to the pectoral muscle is the most distinct and so should be preserved as the neighbourhood size increases. More than one inflection point is however allowed as other inflection points do exist and operation of the expanding window algorithms outside the two-region model could result in those being more distinct than the one corresponding to the pectoral muscle. Non-pectoral edge inflection points that occur inside the pectoral muscle are also useful for determining the angle that the pectoral muscle lies at, another reason for locating multiple inflection points.

**Deviation from the Model**

Several issues arise from fitting the model to a mammogram. First of all, the model is only valid for a limited range of window sizes as it does not take into account regions (iii) and (iv) of Figure 9.5(b). This issue of validity is especially true for the median-based approach as it is much more sensitive to deviation from the bi-modal model compared to the MX value, as seen in Figure 9.8. Because other intensity regions exist in the mammogram, there may be multiple points of inflection in the threshold function, only one corresponding to the pectoral boundary intensities. The second issue stems from the fact that the regions of a mammogram are not homogenous in intensity like the Gaussian based model image. The pectoral boundary can vary in intensity along its length and defining a single intensity to extract it leads to jaggedness in some parts of the border. Also, due to the glandular tissue having similar grey levels to the pectoral muscle, parts of it will also be outlined when using the correct threshold, as is shown in Figure 9.9(c),(d). These are overcome by the post processing described below.
Straight-Line Pectoral Edge Approximation

To locate the outline most likely to correspond to the pectoral muscle and to eliminate the jagged boundaries, a set of straight lines is fitted to the image. Each line is fitted to the border of pixel regions extracted using one of the inflection point thresholds. To fit the line, a simple gradient operator is applied to the binary border pixels, illustrated in Figure 9.10. A local window 11 columns wide and 13 rows high is taken around each candidate pixel and its top and bottom row examined. Due to its angle, the pectoral boundary is expected to run from some point near the top right corner of the window to a point near the opposite corner. The distance, in columns, of the right-most pixel in the top row from the left-most pixel in the bottom row is taken to be inversely proportional to the steepness, or gradient, of the candidate pixel. A tally of gradients and the number of pixels of each gradient is kept. Since the pectoral muscle is the most distinct and straightest boundary in the image, the most common pixel gradient can be assumed to belong to it.

![Figure 9.10: Determining the most probable pectoral muscle angle.](image)

After the edges of extracted binary regions are determined (a), each pixel is examined (b). In the example of (b), a window is taken around the candidate pixel (blue circle). The distance \(d\), taken as the column coordinate of the right-most pixel in the top row minus the column coordinate of the left-most pixel in the bottom row is assumed to be inversely proportional to the candidate pixel angle.

Once the angle that the pectoral muscle runs at has been determined, it remains to position a line of this angle at the appropriate location. The fitting of the straight line is done only with reference to the binary pixels that have been determined to lie at the pectoral gradient, this will include pixels on the boundary but also other pixels, including noise due to edges internal
to the pectoral muscle. Typically, the least mean squared error is used to fit functions to data. However, to remove bias due to the noise, rather than fit the line to satisfy all points as much as possible it was fitted to intersect as many points as possible.

Since several inflection points may exist, several binary pectoral masks and resulting pectoral gradient estimates will result. The intensity of the pectoral muscle near its edge is divided into a number of gradients, running parallel to the edge. If a threshold is a little bit internal to the actual pectoral edge, the gradient of the resulting pectoral line may still be a good estimate of the pectoral muscle gradient. In fact, if noise or overlapping tissue affect the pectoral muscle edge, the gradient determined from tissue internal to the pectoral muscle may be more accurate than one that can be determined from the actual edge. This is used to advantage by fitting several lines to each binary image, so that one line is fitted for each gradient. This means that for every \( n \) thresholds, there are \( n^2 \) straight line estimates of the pectoral muscle edge.

Once a series of straight lines is fitted to correspond to each threshold, it remains to differentiate between the actual pectoral boundary and the boundaries resulting from non-pectoral points of inflection. The pectoral boundary represents a division between two regions of the mammogram. One side of the boundary has relatively high intensities due to the presence of the pectoral muscle; the other side is relatively dark. A line fitted accurately will display the properties of this boundary. To determine which fitted line is most representative of the pectoral outline, a pixel neighbourhood is taken from each side of a line and the mean pixel intensity difference from one neighbourhood to the other is determined. In order to get a more accurate estimate of the line’s validity outlier values are eliminated. The straight line is divided into a number of regions and the intensity difference between the two sides of each region determined.

To get a better estimate, possible outlier regions are removed. The standard procedure for eliminating outlier points from a set is to remove any values more than two standard deviations away from the mean. This procedure was used initially and the results reported in [146] and [147] are based on it. There is however a problem with this procedure in a small percentage of images. The problem stems from the actual intensity gradient of the boundary changing in strength along its length. The largest intensity drop is observed at the top, with the glandular tissue towards the bottom of the pectoral boundary causing a lower intensity
drop between the pectoral muscle and adjacent tissue. When a line fits well at the top, the average intensity difference will be high as intensity differences from the rest of the boundary are not so significant. If the line fits well at the top but not elsewhere, the well fitted points should be considered outliers in order to give preference to lines which fit across the whole pectoral boundary. However, because of the range of intensity differences due to the badly fitted points the standard deviation increases, making the removal of outliers harder. This results in a line fitted well at the top of the boundary but badly elsewhere being selected over a line which fits better everywhere except at the top.

To obtain a better estimate of average intensity gradient along the majority of the pectoral boundary, the set of intensity gradients is taken and the procedure outlined below is followed:

Step 1: Determine mean of set.

Step 2: Replace value furthest from mean with the mean value.

Step 3: If the new standard deviation is higher than 0.05 go back to step 1.

The value 0.05 for the minimum standard deviation was determined by examining well fitted pectoral boundaries.

Once the average intensity gradients of the candidate fitted lines are computed, the line that maximises this gradient is chosen as the best estimate of the pectoral boundary.

**Secondary Approximation from Global Thresholding**

The straight-line fitting procedure is quite separate from the expanding window algorithm of Brink and Pendock, making it a generic procedure resulting in a straight-line estimate from a set of approximations to the pectoral muscle. Chapter 6, which dealt with mammogram decomposition through global multi-level thresholding, discussed two thresholds that may result in binary images that feature the pectoral muscle edge.

To re-cap briefly, the first threshold from the minimum cross-entropy algorithm outlines the near-skin tissue. Applying the algorithm again to those pixels extracted with the first
threshold often shows the pectoral muscle edge in the resulting binary mask, if not, the edge is shown in the subsequent thresholding iteration.

To compare the improvement of the expanding window algorithm over the global estimates, the result from the second and third global threshold were passed to the line-fitting algorithm. Results are shown in Figure 9.11 for the processing of MIAS image mdb025, with (a) showing the original image, (b) and (c) the results of global thresholding at the second and third threshold, and (d) the final straight-line estimate.

![Figure 9.11](image.png)

Figure 9.11: Original image (mdb025) in (a) with results from second (b) and third (c) global threshold shown as binary images. Passing (b) and (c) into the line fitting procedure produces the pectoral line estimate shown in (d).

**Automatic Voting on Pectoral Edge Estimates**

The line fitting procedure includes functionality to determine the best straight-line estimate by examining pixels on either side of the line. This line voting has several further applications, such as parameter tuning for the expanding window algorithms, results-based comparison of different estimation algorithms and as a bridging tool to combine the results of different estimation algorithms.

Both the MX and median-based expanding window algorithms depend on a value, $\Delta$, that corresponds to the size increment between successive windows. As the increment increases, errors due to interpolation when locating the inflection point will increase. As the increment decreases, the execution time increases, as does the probability of error from local variations in the regions. To compare the accuracy of results for different values of $\Delta$ the voting
algorithm can be used to determine if choice of $\Delta$ does cause accuracy problems and its results combined with an execution time analysis to determine the best algorithm relative to both speed and accuracy.

Comparing the accuracy of results is not confined to parameter tuning; results-based evaluation of different algorithms is also possible. To this end, the results of the presented algorithms, namely the MX and Median-based expanding window algorithms and the global approximation algorithm are evaluated to compare results. In addition, the results are compared to the results from the work of Kwok et al. [145]. In [145], an initial straight-line estimate of the pectoral muscle is produced, and this straight line is then allowed to curve to follow the pectoral edge more accurately.

The most powerful application of the pectoral estimate voting algorithm is to vote on pectoral edge estimates from different estimation functions, thus bridging the algorithms into a combined algorithm. Accuracy improvements in a combined algorithm can be measured by an increase in accuracy when compared to another algorithm, as compared to the accuracy when the individual algorithms being combined are compared to that algorithm.

One problem associated with use of the voting algorithm is misclassification. In certain images, a pectoral outline that is inferior to another can still meet the requirement of the voting algorithm better than the other estimates. Such misclassification can be checked by manual inspection of the results to determine the percentage of images misclassified and if there is a higher misclassification associated with any set of data.

The automatic evaluation of estimates from two sources is accomplished by a script that takes binary masks from both estimates and the original image. After comparison the pectoral outline of each estimate is superimposed over the original image and an icon added to the bottom right-hand image corner. The icons are shown in Figure 9.12 and specify whether the estimate is better (a), same (b), or worse (c) as the estimate being compared to, or indicate an error (d) if there is no pectoral muscle estimate.
Results and Discussion

Results are presented from testing on the entire database of 322 images. The individual algorithms all provided a degree of success with the straight-line fits producing visually acceptable results for a wide range of images in the MIAS database. As an example, three such images are shown in Figure 9.13 with the original image above one with the found pectoral outline superimposed. Figure 9.13(a) shows MIAS image mdb031 where a small amount of pectoral tissue is visualised (segmentation result from median-based expanding window algorithm using $\Delta$ of 5). Figure 9.13(b), showing image mdb110, has straight line edges internal to the pectoral muscle (segmentation result from global based algorithm). Figure 9.13(c) shows image mdb053 with a large amount of glandular tissue overlap (segmentation result from MX-based expanding window algorithm with $\Delta$ of 5). Figure 9.13(d) shows image mdb002 which is affected by tapes in the top part of the image (segmentation by median-based expanding window algorithms with $\Delta$ of 5).
Figure 9.13: Example cases where the pectoral muscle is outlined accurately using the various algorithms despite complications such as (a) small portion of pectorals visible, (b) straight line edges internal to pectoral muscle, (c) high glandular tissue overlap, and (d) presence of tapes over the pectoral muscle. Original image is shown above a processed image, oriented with pectoral muscle on left side.

Inaccurately fitted lines arise for a number of reasons, demonstrated in Figure 9.14. In images where the pectoral muscle boundary lies at various angles, like in MIAS image mdb035 shown in Figure 9.14(a), an inaccurate straight line fit results when the median-based algorithm with $\Delta$ of 5 is applied. Various kinds of noise can exist internal to the pectoral muscle breaking it up into various regions. These may result in the line being fitted to points inside the actual pectoral muscle. An example is shown in Figure 9.14(b) using image mdb062 with the global-based algorithm to determine the pectoral outline. In the image there is no strong outline for the pectoral muscle so an internal region is extracted. In image mdb123, shown in Figure 9.14(c), an obvious pectoral outline can be seen. However, an internal region also displays the properties of a straight-line boundary between two intensity regions and is extracted by the MX-based algorithm using a $\Delta$ of 5. For image mdb236 no pectoral muscle is included in the image, this can result in an inappropriately fitted line, shown in Figure 9.14(d), the result of the global-based algorithm. These bad pectoral
boundaries will be re-visited when evaluation of the pectoral line voting algorithm is discussed.

![Images of images showing inaccurate pectoral outlines caused by various factors.]

Figure 9.14: Inaccurate pectoral outlines caused by: (a) curved pectoral edge resulting in several possible angles, (b) variations internal to the pectoral muscle being detected, (c) straight line edge internal to the pectoral muscle present, and (d) no pectoral muscle present.

**Execution Time**

While evaluation of final pectoral outlines may have a subjective quantity, algorithm speed is a quantity that can be measured more objectively. The post-processing operation of line fitting is common to all the algorithms so any speed advantage of one over another is in determining the candidate thresholds. For both the median and MX-based expanding window algorithms speed depends on the increment, $\Delta$, that the window is expanded by, as discussed. The execution time taken to determine candidate thresholds is shown in the graph of Figure 9.15 for several values of $\Delta$. Execution time of the global thresholding algorithm is also shown for comparison. The global threshold does not depend on a $\Delta$ term, with three iterations of the minimum cross-entropy algorithm being taken in each case and execution speed the same.
Due to the low complexity of the global thresholding method, it is the fastest. It was noted already that the median value is faster to compute than the MX value, hence the median based expanding window algorithm is faster than that based on the MX value. The speed advantage of the median-based algorithm becomes more apparent as the $\Delta$ decreases. The lower the $\Delta$, the less practical the MX based algorithm becomes.

**Expanding Window Tuning**

To determine how performance is affected by the $\Delta$, results comparing accuracy for the three $\Delta$ values used in the speed comparison are shown in Table 1 for the MX-based algorithm and in Table 2 for the median-based expanding window algorithm, the evaluation performed using automatic voting. The voting resulted in measures of percentage of cases more accurate and percentage of cases less accurate for one $\Delta$ relative to another.
Table 1: Comparative performance of MX-based expanding window algorithm for different values of window expansion coefficient $\Delta$.

<table>
<thead>
<tr>
<th>MX-Based</th>
<th>Compared to results for:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta=1$</td>
</tr>
<tr>
<td>Percentage of estimates more accurate for:</td>
<td></td>
</tr>
<tr>
<td>$\Delta=1$</td>
<td>0%</td>
</tr>
<tr>
<td>$\Delta=5$</td>
<td>44.3%</td>
</tr>
<tr>
<td>$\Delta=10$</td>
<td>46.4%</td>
</tr>
</tbody>
</table>

Table 2: Performance comparison of Median-based expanding window algorithm for different $\Delta$ values.

<table>
<thead>
<tr>
<th>Median-Based</th>
<th>Compared to results for:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta=1$</td>
</tr>
<tr>
<td>Percentage of estimates more accurate for:</td>
<td></td>
</tr>
<tr>
<td>$\Delta=1$</td>
<td>0%</td>
</tr>
<tr>
<td>$\Delta=5$</td>
<td>46.4%</td>
</tr>
<tr>
<td>$\Delta=10$</td>
<td>45.5%</td>
</tr>
</tbody>
</table>

For some images, the same pectoral estimate was produced. For parameter values that produce very similar results only slight variations will tip the accuracy towards one or the other. If these are assumed to be random the percentage of estimates judged better for an algorithm relative to an algorithm with similar performance will be around the 50% point.

To read the performance tables and compare two algorithms, the table can be split along its diagonal, formed by the cells containing zero values resulting from the algorithm compared to itself. The two cells pertaining to the compared algorithms, one on each side of the diagonal, contain the percentage of estimates judged better for one algorithm relative to the other.
diagonal, can be compared. For example, comparing the MX algorithm when $\Delta$ equals 5 to when $\Delta$ equals 10, the table shows that in 40.6% of images a $\Delta$ of 5 produces a better result and in 47.7% of images a $\Delta$ of 10 produces a better result. These two numbers do not add up to 100%, meaning that in 11.7% of the cases the same approximation is produced by both algorithms. These figures indicate that the MX algorithm when $\Delta$ equals 10 outperforms the algorithm when $\Delta$ equals 5. This is in contrast to the results presented in [147] when a $\Delta$ of 5 produced better results using the simpler line voting algorithm with outlier removal performed only by eliminating values more than two standard deviations from the mean.

From the accuracy results of the MX algorithm, a $\Delta$ of 10 gives improvements over both the other options, while the median-based algorithm performs best with a $\Delta$ of 5, both of these are relatively fast when compared to the same algorithms when $\Delta$ equals 1, so if execution time is a factor these better performing algorithms may be used without referring to the slower results. When the accuracy of the voting algorithms was checked by a human non-expert, it was found that overall approximately 3.8% of images were misclassified, the algorithms giving preference to an inferior outline. The worst performance of the six comparison trials was 6.2% misclassification and best performance of 3.1% misclassification.

Although results are close to the 50% point that would indicate identical performance, for both algorithms each different value of $\Delta$ produced some pectoral estimates that were clearly better than those for other $\Delta$ values. To provide a best estimate from the expanding window algorithms to use in their comparison the voting procedure was used to combine the results from the three $\Delta$ values. Concerning speed, the main drawback with this is the number of lines that are fitted, as for the actual expanding window it only needs to be performed once using a $\Delta$ of 1 and the resulting function sub-sampled to generate data for other $\Delta$ values. The results presented from this point pertain, in the case of the expanding window algorithms, to the combination of results from $\Delta$ values one, five and ten.

**Algorithm Comparison**

To determine performance of one algorithm relative to the others the voting procedure was used again to calculate the overall improvement. Results are shown in Table 3 in the first
three rows/columns of data. This time the average number of misclassified images per whole database comparison was 4.6%.

Table 3: Algorithm accuracy comparison.

<table>
<thead>
<tr>
<th>Percentage of estimates more accurate for:</th>
<th>MX</th>
<th>Median</th>
<th>Global</th>
<th>Global + Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX</td>
<td>0%</td>
<td>47.4%</td>
<td>63.8%</td>
<td>32.8%</td>
</tr>
<tr>
<td>Median</td>
<td>39.3%</td>
<td>0%</td>
<td>59.4%</td>
<td>0%</td>
</tr>
<tr>
<td>Global</td>
<td>29.1%</td>
<td>33.4%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Global + Median</td>
<td>53.6%</td>
<td>33.4%</td>
<td>59.4%</td>
<td>0%</td>
</tr>
</tbody>
</table>

From the results, the MX shows better results than the Median methods with 47.4% of the images getting a better estimate from the MX-based algorithm, higher than the 39.3% of the Median-based algorithm. Both the MX and Median based algorithms produce on average a better fit than the global threshold based method having better estimates in 63.8% (MX) and 59.4% (Median) of cases while the Global threshold based method only does better in 29.1% (MX) and 33.4% (Median) of cases.

**Algorithm Bridging**

Combining pectoral estimates from several algorithms using the voting algorithm as a bridge results in better accuracy. The fourth data row/column in Table 3 shows results when the best estimates out of the Median-based and global threshold-based approaches are compared with the individual techniques. Since the automatically generated results are from the Global and Median-based approaches, the results can only be compared fairly to those of the independent MX-based algorithm. The results show that the Combined algorithm approach is more accurate than the MX-based in 53.6% of cases, while it is less accurate in only 32.8% of cases, a large improvement over the accuracy results from the comparison of each individual technique. Misclassification by the voting in this case is 2.8% of images with 6
images classified wrongly as being segmented better by the combined algorithms and 3 images where the MX-based algorithm was voted better although it produced a worse result. To see the effect of algorithm bridging, some of the images shown as examples where segmentation failed with some algorithm in Figure 9.14 are re-visited, with results shown in Figure 9.16. In the figure the pectoral approximations from combining the expanding window algorithms with the three $\Delta$ values are shown along with the approximation from the global-based method, the selected approximation out of the three is indicated with the ‘tick’ icon.

For the image in Figure 9.16(a), MIAS image mdb035, the bad approximation provided by the median-based expanding window when $\Delta$ equals five, shown previously in Figure 9.14(a) has already been eliminated by combining the result with the other median-based approximations. Now the median-based approximation is the best one, providing a reasonable straight-line estimate of the curved pectoral line while the approximations from
the MX-based and global approximations are tuned to a pectoral angle and location to fit well in part of the image while inaccurate in the other parts.

Figure 9.16(b), of image mdb062 shown to fail previously in Figure 9.14(b), shows an example where the voting algorithm fails to locate the best boundary. While the median-based approximation is closest to the actual pectoral boundary the fitted line lacks the properties of a step in intensity. Instead, the MX-based approximation is voted as best. Examining the image a small bright triangle fitting the pectoral muscle description has been outlined.

Two straight lines display the properties of the pectoral muscle boundary in image mdb123 in Figure 9.16(c), again the bad approximation that was shown in Figure 9.14(c) has been eliminated by combining the expanding window algorithms using different $\Delta$ values leaving three reasonable approximations to choose from, the one from the global-based algorithm chosen as the best approximation.

The case of the image where no pectoral muscle appears is not re-visited here since the voting algorithm results in no better approximation; instead another interesting image (mdb263) is shown in Figure 9.16(d). Out of the three choices, the median-based one is the only reasonable one, the global-based algorithm providing no approximation and the MX-based one outlining an edge internal to the pectoral muscle. The voting algorithm correctly chooses the median-based approximation.

Table 4: Percentage of cases where the Kwok et al. algorithms produced better results.

<table>
<thead>
<tr>
<th>Percentage of estimates more accurate</th>
<th>Compared to results of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MX</td>
</tr>
<tr>
<td>for: Kwok et al. Straight Line</td>
<td>52.6%</td>
</tr>
<tr>
<td>Kwok et al. Fitted Curve</td>
<td>69.3%</td>
</tr>
</tbody>
</table>
Table 5: Percentage of cases where the Kwok et al. algorithms produced worse results.

<table>
<thead>
<tr>
<th>Percentage of estimates more accurate</th>
<th>Compared to results of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kwok et al. Straight Line</td>
</tr>
<tr>
<td>for:</td>
<td></td>
</tr>
<tr>
<td>MX</td>
<td>47.1%</td>
</tr>
<tr>
<td>Median</td>
<td>42.7%</td>
</tr>
<tr>
<td>Global</td>
<td>32.5%</td>
</tr>
<tr>
<td>MX + Median + Global</td>
<td>60.1%</td>
</tr>
</tbody>
</table>

Results obtained by Kwok et al. in [145] were provided for comparison. Results are divided into two tables to avoid repeating information from previous tables. Table 4 shows the percentage of cases where the Kwok et al. algorithms outperform the ones presented in this chapter, while Table 5 shows the percentages where results from the three individual algorithms and the combination of all three are better relative to both the straight-line estimate and the fitted curves resulting in [145].

For the straight-line fit, the results show that while the Kwok et al. algorithm is better than the individual estimates, once the individual algorithms are combined, their straight-line estimate is more accurate. Misclassification by the voting algorithm in this case occurs on average in 5.8% of images, in 60% of these misclassifications correspond to the Kwok et al. estimate having a more accurate boundary.
When comparing the results with the fitted curve of Kwok et al, all the tested algorithms perform worse overall, offering better results only in some cases where the pectoral boundary shows minimal curvature or when the algorithm of Kwok et al. either fails to locate a pectoral muscle or results in a curve fitted to regions not part of the pectoral edge. Three examples where the combined algorithm produces a better fit than the curve are shown in Figure 9.17, performance being better for the line due to (a) the curve being fitted to glandular tissue, (b) no approximation by the fitted curve, and (c) the curve fitted to a boundary internal to the pectoral muscle. Figure 9.18 shows three examples where the curve fitting algorithm of Kwok et al. produces better results, due to (a) the straight line outlining a bright region internal to the pectoral muscle, (b) convexity of the pectoral muscle, and (c) concavity of the pectoral muscle (extremely concave in the example). The misclassification when comparing to the fitted curve rises to an average of 10.4% of images, weighted against the fitted curve, with 86% of the misclassified outlines favouring the combined algorithm where the fitted curve is actually more accurate. A misclassification example is shown in
Figure 9.18(a) where the fitted straight line is set against an internal edge at the bottom, while
the curve, although it runs closer to the real pectoral edge, does not seem to run along a
distinct intensity gradient at the bottom of the pectoral muscle.

![Image](image1.png)

(a) Comb. Kwok Curve (b) Comb. Kwok Curve

(c) Comb. Kwok Curve

Figure 9.18: Three examples where the straight line of the combined algorithm (Comb.) is worse than the fitted curve of Kwok et al. In (a), mdb003, the straight line fits into a bright
dge internal to the pectoral muscle (this edge is so distinct the
ing algorithm prefers it to the fitted curve), (b), mdb028, the
voting algorithm is better approximated
by a curve, and in (c), mdb111, the high degree of concavity of the
ectoral outline also leads to the curve being a better
proximation.

**Extension of the Voting Algorithm**

The voting algorithm has so far been used to choose from a number of pectoral estimates
originating from different algorithms. A closely related application is to use such a voting
algorithm to determine if refinements made to a pectoral estimate from one estimate
improve or worsen the estimate. As an initial investigation, the results from the Kwok et al.
[145] algorithm, being the initial straight-line estimate and its refined fitted curve were fed
into the voting algorithm. The raw results indicated that in 74.3% of the cases the refined
curve was better, but that in 23.8% of the cases the initial straight-line estimate was a better
approximation to the pectoral region (for the other images no approximation was made).
Upon manual inspection by non-experts it was determined that although 92.3% of cases were classified correctly, 31% of images where the straight-line was deemed better were incorrect. From most of the cases where the straight line was deemed better, visual inspection showed only slight degradation in the case of the fitted curve, although the procedure was useful in selecting the straight line approximation over the image in which fitting a curve resulted in no pectoral estimate.

As before, the main problem stems from either the intensity contours internal to the pectoral muscle or the subtle intensity differences towards the bottom of the pectoral muscle, where sometimes a boundary which cuts this region off completely represents a larger average intensity step than a boundary which includes the subtle region – by so lowering the average intensity step from one side of the fitted line to the other.

The results show that while the simple average intensity difference can be used to vote on straight-line estimates with some degree of accuracy, the more refined the pectoral line being voted on the more sophisticated this procedure has to be. Further investigation is needed in this area, a simple extension may be to divide the proposed pectoral outlines in half, and test each half independently.

**Conclusion**

A number of algorithms were presented in the chapter for delineating the edge of the pectoral muscle, two were local thresholding algorithms based on an expanding window. In addition, the pectoral estimates from global thresholding presented in Chapter 6 were also considered. For the expanding window approach, a modification of cross-entropy (MX) was introduced and the threshold provided by both it and the median was found to display useful properties when used with an image whose histogram approximated a bimodal distribution. When either MX or median values are taken from variable sized windows around the pectoral muscle as a function of window size, an inflection point occurs at intensities between two distinct intensity regions allowing them to be separated. This was introduced using a model bi-modal histogram, a model image of the pectoral muscle and adjacent region, and extended to an actual mammogram to extract the pectoral muscle.
A straight line was fitted to the output image from the thresholding algorithms and was also used to automate their evaluation by examining pixels on either side of the line to see which displays the properties of the pectoral muscle edge, which is a sharp step from bright to dark intensities.

Through the testing it was concluded that the MX algorithm was the most accurate, with 47.4% of images providing a better fit than from the median-based approach, while in the median-based approach which was the second most accurate, 39.3% of pectoral estimates were judged better than those from the MX algorithm, although the median algorithm has a speed advantage over the MX algorithm. The voting algorithm was also used to combine results from different thresholding algorithms, choosing the most appropriate. Results of this combination, along with original results were compared to the straight-line estimates and fitted curves of the algorithm of Kwok et al. [145]. For the MX algorithms alone, 47.1% of estimates were judged better in this comparison, while for a combination of MX, median, and global algorithms this rose to 60.1%. On the other side, the other algorithm was only better in 39.6% of cases when compared to the combination of MX, median, and global algorithms.

The fitted curve of Kwok et al. was judged superior in more cases than the combined algorithms, leading to the conclusion that the pectoral muscle is better represented by a fitted curve than a straight line. The next chapter on mammogram classification will however indicate that a straight line fit to the pectoral muscle is adequate in many cases.

**Future Work - Extensions of the Expanding Window Algorithm**

As an aside from the segmentation of the pectoral muscle, the expanding window algorithm can be used as a general local thresholding segmentation procedure for segmentation of intensity-separable regions. Some preliminary work in this direction is presented here in order to show again that the intensity properties of local regions of increasing size can be used to determine a threshold between two regions of separate intensities.

The main problem is finding a point from which to expand the window. The window location in the image will depend on the application, in the case of the pectoral muscle the
image corner was used to anchor the window. For an automatic application, the initial location must first be found by other means.

For a general application, semi-automatic segmentation is possible by letting the user select the window location. Such an algorithm has been implemented using the median-based expanding window due to its speed advantage. The algorithm has been integrated into the user interface presented in Chapter 6, a screen shot shown in Figure 9.19. The expanding window thresholding has a similar effect as the ‘magic wand’ selection feature present in a number of graphics programs. Pressing the ‘median magic wand’ button allows the user to select a spot within the image by double clicking. A square window is expanded, using the selection point as the centre and inflection points in the threshold versus window size function computed.

![Figure 9.19: The Median Magic Wand function implementing a median-based expanding window had been added to the graphical user interface from Chapter 6.](image)

For the case of the pectoral muscle, the intensity gradient perpendicular to the straight-line estimate was used to choose a threshold from those provided by the inflection points. To make the algorithm more general, another approach to choosing the best threshold was
taken. In the context of determining inflection points through the fitting of quadratic functions, fitted functions in adjacent windows can be represented as:

\[ y_n = a_n x^2 + b_n x + c_n \]
\[ y_{n+1} = a_{n+1} x^2 + b_{n+1} x + c_{n+1} \]

An inflection point exists where the sign of the \( a \) term changes between adjacent windows \( n \) and \( n+1 \). An indication of the “strength” of the inflection point found there is from the value of \( |a_{n+1} - a_n| \). Thresholds from the two strongest inflection points (where more than one existed) were selected for segmenting the image.

The inflection points provide a square area of interest to which thresholding is localised. Thresholding based on the inflection points is undertaken only within the area of interest and object boundaries from each threshold outlined.

The effect of using the technique in some regions of MIAS image mdb206 is shown in Figure 9.20, with original image in (a) and user-selected window centres show in (b). The outlines of the segmented objects in each case are shown (c-g). The red outline corresponds to regions extracted using the strongest inflection point to determine the threshold. The green line corresponds to objects using the second strongest inflection point. The area of interest in each case is indicated by the blue square. For the small bright regions (c)-(d) the first inflection point provides a more accurate outline, for the large bright region in (g), the second inflection point extracts it while the first only extracts a subset of pixels internal to it.
Figure 9.20: Original image (MIAS mdb206) shown in (a) with expanding window centres chosen by user in (b). Results of expanding window segmentation shown (c)-(g) with red outline corresponding to strongest inflection point, green second strongest, with blue rectangles outlining the area of interest.
The expanding window algorithm model is based on two homogenous intensity distributions present in the window. When the centre of the expanding window is placed in an area of a single relatively homogenous region, the behaviour is less predictable, as illustrated in Figure 9.21, with window centres shown in (a) on the original, again MIAS image mdb206. Sometimes minor deviations within the region will cause an inflection point, leading to noise being extracted as shown in (b). At other times the window will expand until other regions are included in it and separate them, shown in (c), where the darker tissue is separated from the pectoral muscle on one side and the glandular tissue on the other, and in (d) where the image background is separated from the breast and label. Post processing would be useful in such cases, either to limit the inflection point strength for an acceptable inflection point, or to apply some criteria to the result (i.e. roughness of extracted edges) to eliminate noise.
Figure 9.21: Algorithm performance when window centred in relatively large homogenous region, centres indicated in (a). Local variations are highlighted in (b) while in (c) and (d) the window expands to include adjacent regions to determine the threshold.
Chapter 10

CLASSIFICATION OF MAMMOGRAPHIC DENSITY AND DATA MINING USING HISTOGRAM DISTANCE MEASURES

Segmenting the Mammogram Database

While previous Chapters have concentrated on the segmentation of individual mammogram images, the aim of this Chapter is to bring together the segmentation methods from this thesis to isolate the breast tissue from the image background and pectoral muscle and ‘segment the database’ depending on the mammogram tissue class. Such segmentation can be used for classification of mammograms into risk groups, a problem approached by others previously. Another application is for retrieval of images for examination. Rather than randomly presenting the images to the radiologist, they can be served interactively according to human factors and a rating of reading difficulty or personal choice, allowing either ‘like’ or ‘different’ images to the current image to be served, taking advantage of the flexibility of soft copy reading.

Classification in this chapter is performed on the basis of histogram distance, a relatively new concept. After a review of current mammogram classification techniques, which are usually feature-based, existing histogram based image comparison methods are introduced. After discussing the inadequacy of some existing histogram comparison approaches, a histogram distance is introduced, showing improved performance against existing histogram comparison techniques and similar performance to a feature based classification technique.

While the performance validates the histogram-based distance measures, it also provides assurance on the performance of the segmentation algorithms presented in this thesis. The results presented in this chapter are reported from tests on the entire MIAS database, all with segmentation performed automatically and without user intervention using algorithms described in the earlier chapters of this thesis.
Introduction
Breasts can be classified into several types depending on mammographic appearance. The appearance of the mammographic parenchymal pattern was first associated with breast cancer risk by Wolfe [20], who separated parenchymal patterns into four classes. The breast classifications provided with the MIAS database include three types: fatty, with the least amount of glandular tissue, dense, with the most, and fatty-glandular type somewhere in-between. First, the classification of mammographic density is introduced, followed by an introduction to histogram distance measures. A new histogram distance measure, The Closest Point Distance is introduced, after which different segmentation algorithms from this thesis are brought together and classification algorithms evaluated on the segmented breast tissue.

Density Classification Application
The computerised assessment of breast tissue density has in the past been used to classify mammograms to determine risk, and combined with other risk factors such as age and family history. The level of glandular tissue also corresponds to a difficulty of reading, lesions and microcalcifications being more difficult to see when superimposed over the bright glandular tissue than over the dark fatty tissue. Since soft-copy provides the flexibility of having a number of mammograms digitally stored, rather than examining images randomly, the image difficulty may be used to serve the images in an order determined interactively depending on human and patient factors.

While studies on the benefit of factors that affect reading speed and accuracy need to be formalised, the work in this chapter presents an image database ordering method for mammograms. Rather than explicitly divide the database into classes or extract specific features, a distance measure between histograms of regions is used to order the database with reference to a given image. The same distance measures can also be used in the more conventional application of classification.

Previous Work on Automatic Type and Density Classification
Previous work to automatically classify mammograms according to some tissue type has mainly focused on finding a number of features with which the types may be discriminated.
In the work of Karssemeijer [81], the mammogram was segmented into breast tissue and pectoral muscle from which a number of features were extracted. For each mammogram, distance of tissue from the skin line was normalised to obtain invariance to breast size. Histograms of tissue at several distances from the skin-line were taken and the standard deviation and skewness of these histograms used as features. To derive more features, the pectoral muscle and breast tissue areas were modelled by probability density functions, the difference of the two distribution means and the overall difference of the distributions were then extracted as features. Classification was performed by the k-nearest neighbour classifier, the dataset consisted of 615 oblique mammograms, classified by a radiologist into four classes based on density. Overall, the agreement of the algorithm with the radiologists classification was 67%, rising to 80% when only “recent” (after 1991) mammograms were included due to improved imaging techniques.

The performance of Karssemeijer’s algorithm was evaluated on 265 images from the MIAS database by Blot and Zwiggelaar [126]. Using the three MIAS classes a success rate of 65% was reported with most misclassifications between dense-glandular and fatty-glandular. In the same work, Blot and Zwiggelaar used Karssemeijer’s algorithm on images consisting of reconstructed background texture [148] with an accuracy of 63%.

The idea of using pectoral muscle intensities as a reference to normalise tissue densities is not unique to Karssemeijer’s work, Matsubara et al. [83] divide the fibro-glandular area of the mammogram into three regions based on tissue density. The intensity of the densest region is normalised with reference to the pectoral muscle intensity and the ratio of the three regions used to classify the mammogram into one of four types. Results report a 91% agreement with radiologist classification while conceding that increasing the database size is necessary.

Among others, feature-based classification has been reported on by Bovis and Singh [84], using various texture-based features and an ANN classifier. Using 377 mammograms from the DDSM public database an accuracy of 77.3% is reported when the database is split into four classes and 96.7% when split into two classes. Blot and Zwiggelaar [150] have investigated the use of the percentage of glandular tissue, $h_{int}$ representation by Highnam et al. [151] for estimation of breast density. The $h_{int}$ method depends on determining the thickness of the compressed breast and the authors conclude that existing methods for this are not adequate for the task of estimation of breast density.
The difference in classification accuracy between algorithms is not only due to the accuracy of the image processing techniques involved. Some can be attributed to human inaccuracy when the mammograms are classified by radiologists as it has been established that a large variance can exist between radiologists [149]. This means that reported accuracies with respect to human-based classification on a non-public database should be treated with scepticism unless comparison is also made with the accuracy of other established algorithms.

**Histogram Based Image Comparison**

Image histograms provide “overall measures of the statistical content of images” [108], hence their suitability for image comparison and query-by-example retrieval [152], [153]. By measuring distances between histograms, applications such as ordering a set of images with respect to a reference image, or classification based on a set of typical class images is possible. An intensity histogram is independent of image orientation so no explicit image co-registration, apart from size and intensity scaling, is needed when comparing histograms reducing potential error sources. Some segmentation is however desirable to remove noise from contaminating the histogram.

Information provided by histograms has been used in several of the feature-based mammogram classifiers; the sub-region histogram standard deviation and skewness of Karssemeijer’s work is an example [81]. Zhou et al. [86] take a more flexible approach, the histogram of the entire breast region is extracted, and this region’s histogram is used for classification. The work in this chapter was performed before the paper by Zhou et al. was sighted. However, their idea of using the entire breast tissue histogram is essentially the same as here, although Zhou et al. take the feature-based route – extracting several shape based features and using a rule-based classifier to assign a breast density class. The work in this chapter takes a different approach to the feature-based classification, replacing the feature extraction and classification steps with a histogram distance measure.

**Histogram Distance Measures**

There are a number of distance measures for histogram comparison ranging in complexity. To show the theory behind some distances, two histograms \( F \) and \( G \) are defined:

\[
F = [f_1, f_2, \ldots, f_I], \quad G = [g_1, g_2, \ldots, g_J]
\] (35)
Each element $f$ and $g$ represents a pixel count in the bin labelled by the element’s subscript.

The most basic distance measures compare the number of pixels in identically labelled bins and sum the result to obtain an overall distance in the form of:

$$\text{Dist} = \sum_{i=1}^{I} \text{operation}(f_i, g_i)$$  \hspace{1cm} (36)

Common distances using this approach include the $L_p$ norm family, known as the Minkowski norms, that take the form:

$$L_p = \left( \sum_{i=1}^{I} |f_i - g_i|^p \right)^{1/p}, \ p \geq 1$$  \hspace{1cm} (37)

with the $L_1$ city block distance and $L_2$ Euclidean norm being the most common, the city block usually seen as a fast approximation to the Euclidean distance. Other distance measures applied in the form of (36) include the cross-entropy distance and Chi-square.

While these algorithms are adequate for some tasks they are limited since they only compare pixels between identically labelled bins, discarding any relationship to pixels in other bins of the histogram.

**Incorporating Inter Bin Relationships**

More sophisticated distance algorithms depend on the difference in pixel counts and also the bins they occur in. The simplest incorporate bin position by considering a cumulative histogram when determining distance. From the histogram definitions in Equation (35), the cumulative histograms can be defined as:

$$\text{FC} = [f_1, f_1 + f_2, \ldots, \sum_{i=1}^{I} f_i], \text{GC} = [g_1, g_1 + g_2, \ldots, \sum_{i=1}^{I} g_i]$$  \hspace{1cm} (38)

One measure that operates on cumulative histograms is the Kolmogorov-Smirnov (KS) distance, given as the maximum difference between the two cumulative distributions:

$$D_{KS} = \max |\text{FC} - \text{GC}|$$  \hspace{1cm} (39)
A similar distance is the Kuiper distance, given as Equation (40), which is more sensitive to the tails of distributions.

$$D_{Kuiper} = \max(FC - GC) + \max(GC - FC)$$  \hspace{1cm} (40)

The relationship between bin locations is more explicitly considered in the Quadratic Form distance [154]. Although colour histograms were considered, the algorithm can also be applied to greyscale distributions. The Quadratic Form distance is a weighted distance measure defined as:

$$D_{QF}^2 = (F - G)' A (F - G)$$  \hspace{1cm} (41)

In (41) A is a matrix with each element $a_{ij}$ representing a measure of similarity between histogram bins i and j, leaving a set of similarity parameters for each bin combination to be determined.

Rather than using an arbitrary similarity measure between bins, the Earth Movers Distance (EMD) of Rubner et al. [155] and the Minimum Difference of Pair Assignments of Cha and Srihari [156] treat the distance measure as an amount of work to transform one histogram into the other. Rubner et al. performed their work on both classical histograms with uniformly spaced bins and distributions known as signatures with bin positions to suit the image. Both colour and texture histograms and signatures were considered in their work. The EMD is given as the minimal cost to transform one distribution to the other, being the sum of costs to move each individual feature vector in the distribution. The term ‘feature vector’ is used to describe the bin position. In Rubner’s work on colour images this was a three dimensional vector in the CIE-LAB colour space. The number of pixels in each bin is seen as a weighting, or number of mass units that can be moved, rather than part of the vector. The solution is given in terms of solving the transportation problem where one histogram represents a set of producers, with the pixel mass representing product, and the other histogram a set of consumers, with the pixel mass representing a demand. Transporting the mass from each bin (producer) in one histogram to a bin (consumer) in the other histogram has an associated cost, corresponding to distance between the bins based on their location vectors. The transportation problem is the problem of meeting demand with product at the cheapest transportation cost; this cheapest cost corresponds to the Earth
Movers Distance. Rubner uses a variation of the Simplex algorithm to find the Earth Movers Distance. The transportation problem is however very intensive, for 8 bit greyscale histograms each producer can deliver to 256 consumers and there are 256 producers. In Rubner’s work on colour images complexity is reduced by using a ‘signature’ composed of only eight bins. For the task of image retrieval low-resolution histograms can be used without severe degradation of performance, as concluded by Brunelli and Mich [152], however they considered only the simple distances that use the difference in pixel levels of corresponding histogram bins, like the $L_p$ norms, and cumulative histograms with Kolmogorov-Smirnov and Kuiper distances. Since the EMD uses distances between bins it is unclear if it is more sensitive to low resolution histograms, but in either case there are limits to how low resolution can get before degradation occurs.

The Minimum Difference of Pair Assignments (MDPA) is defined by Cha and Srihari in [156] as:

Definition 1: If the two histograms are seen as two sets, then the Minimum Distance of Pair Assignments is “the best one-to-one assignment between two sets such that the sum of all differences between two individual elements in a pair is minimised”.

However the implementation given in [156] does not follow this definition, being instead a special case solution of the Earth Movers Distance for one-dimensional arrays (like intensity histograms) having the advantage of avoiding the complexity of the transportation problem, and providing a fast solution without the need for reduced resolution histograms. Based on the labelling of the bins, histograms can be divided into three types, an MDPA defined for each. The types are Nominal, where the bins are assigned arbitrary labels (eg. type of fish in a pond), Ordinal, where there is an order in the labels but no other numerical relationships besides ordering can be assumed, and Modulo, where the labels wrap around (eg. compass directions). For the application using intensity histograms, the ordinal application of MDPA will be examined.

The implementation of the MDPA for ordinal histograms treats each pixel in a bin as a separate element. Elements in one histogram are moved at a cost until it is transformed into the other histogram. The problem in this is that cost as defined in [156] corresponds to the number of bins that a pixel moves. Moving a single bin left or right corresponds to a cost of
1, irrespective of the number of pixels in any bin, also pixels from one bin in one histogram are distributed amongst the bins in the other histogram to make one histogram look like the other. Clearly this does not correspond to the claimed Minimum Distance Point Assignment algorithm in Definition 1, but a distance following the principles of the Earth Movers Distance, to which it will be shown to give identical results. The Minimum Distance Point Assignment, as defined in Definition 1 will be shown as a special case of a new histogram distance, introduced next.

Figure 10.1: A fatty-glandular type mammogram (mdb269) (a) with a fatty type mammogram (mdb309) in (b).

Figure 10.2: Histograms of mammograms from Figure 10.1 shown superimposed. The peak associated with the main tissue region can be differentiated in the two histograms by a difference in height as well as a difference in location along the intensity axis.
A New Distance Measure

Rather than having a histogram distance measure that only examines direction along the pixel number axis within each bin, or measures a distance in terms of the number of bins, the two can be combined. To see the motivation for this, consider the two mammogram types in Figure 10.1 and their histograms in Figure 10.2. The green histogram is from a fatty-glandular mammogram, while the superimposed blue histogram is from a fatty type mammogram. The peak corresponding to the main breast tissue is indicated for both histograms. As the level of glandular tissue decreases, the location and size of the main tissue peak changes, the peak moving towards the left of the histogram due to the lower tissue density and increases in height due to the lower variance caused by uniformity of tissue (seen from the texture in the images). This behaviour indicates that rather than being a one-dimensional distribution with fixed bins for different mammogram regions, the intensity bin axis of the histogram provides its own degree of freedom. Since intensity in the mammogram corresponds to tissue type, movement along the intensity bin axis represents change in tissue density while distances along the pixel count axis represent the proportion of a tissue type. The distance between a point on one histogram and a point corresponding to similar tissue in another histogram with respect to both axes is an indication of image similarity. A Euclidean distance measure can be taken between points on the two histograms as a combination of vertical and horizontal distances. Matching points on one histogram with points on the other to minimise the overall distance results in a Closest Point Distance measure.

The Closest Point Distance is similar to the Earth Movers Distance but rather than treating the number of pixels in a bin as a mass composed of individual units, the histogram height is regarded as a distance along an extra dimension in the feature vector. The difference between the Closest Point Distance and the Earth Movers Distance in terms of the transportation problem is shown in Figure 10.3. In the Earth Movers Distance (a), the bins in one histogram are producers (P1-P3) and bins in the other histogram consumers (C1-C3), the height of the histogram at each bin corresponds to the supply or demand of each bin. The cheapest way to deliver supply to demand is determined using distance between each bin (distance in horizontal direction only) as transportation cost. The distance of the top element to each bin is shown using arrows (this cost is the same for all elements in that bin). In the approach of the Closest Point Distance, shown in (b) for the same pair of histograms, the
top-most element in each bin is the producer or consumer, each with a capacity of 1, the transportation cost between producer and consumer includes distances along both axes of the histogram. The Closest Point Distance can be found as a solution to the transportation problem, using established algorithms such as the Simplex algorithm, much like in the EMD approach although a reduction in the resolution of the histogram is necessary to make it practical. Of note is the fact that the Closest Point Distance can lead to an unbalanced transportation problem, where supply may exceed demand or vice versa. For example if one image had pixels representing two intensities and another same sized image had pixels all belonging to a single intensity. If the first image is assigned as the producer, it has two producer bins with a total supply of two while the other image has one consumer with a demand of one. This problem is overcome using a standard solution of inserting dummy producers/consumers with zero transportation cost to balance supply with demand.

An approximation of the Closest Point distance that is less computationally intensive and practical even when using the full resolution histogram is to determine the distance of each point on one histogram to the closest point on the other. This approximation is non-linear, as several points on one histogram may be closest to the same point on the other, implying that the approximation is non-symmetric. Consider the simple example in Figure 10.4, with
histograms F and G. Here each histogram has two bins at a height of five with the other bins empty. As all bins are of the same height, distance between bins in the vertical directions are zero. The closest point to both blue histogram end points $f_1$ and $f_2$ is the red point at $g_1$, the total distance of blue histogram to red histogram using the approximation being 3. The closest point to both red points $g_1$ and $g_2$ is the blue point at $f_2$, the total distance from red histogram to blue histogram being 5, hence the approximation is non-symmetric.

![Histograms](image)

Figure 10.4: Two simple histograms superimposed to show the non-symmetry of the Closest Point Distance approximation.

What is actually desired when matching the points from the two histograms is to match points corresponding to similar regions. Determining the Closest Point Distance and its approximation can be seen as form of registration of the two histograms.

**Scaling – $L_2$ Norm to True MDPA distance**

Combining inter bin distances with distances along the pixel count axis elevates them above the class of ordinal data type. It must be determined how distances along the intensity bin axis are scaled with respect to distances along the pixel count axis. If distance between bins is weighted too high (approaching infinity), only bins 0 distance away will be matched, the final distance being the $L_2$ Euclidean distance from Equation (37). If the scaling is too low (approaching 0) the original description of the Minimum Distance Point Assignment is achieved. This is the case since when inter-bin distances are set to zero only the distance between bin levels is considered. Since each bin represents only one unit of supply/demand,
the solution to the transportation problem is a matching of bins from one histogram to the
other to minimise the difference in the levels of matched bins – fitting the definition of the
Minimum Distance Point Assignment in Definition 1, and referred henceforth as the True
Minimum Distance of Point Assignments.

Several scaling ratios are used in the evaluation to determine a range of results, with the
intensity bin axis scaled so that its range is equal to some multiple of the pixel count range for
the pair of histograms being compared. To avoid a scaling parameter, both axes may be
scaled based on their standard deviation, this is commonly used with distributions that
approximate a Gaussian. In this case however, the pixel count in most histograms are a
mixture of Gaussians while the intensity bins due to their nature are evenly spaced with each
bin appearing only once so the benefits of scaling based on standard deviation is not clear.

Non-Linearity
The non-linearity introduced by using the closest distance approximation results in the new
distance being non-metric, as the symmetry property and the triangle inequality are not met.
The reflexivity and non-negativity properties however are met by both the Closest Point
Distance and its approximation, as only the distance between a histogram and itself will be
equal to zero, corresponding points lying on top of each other, the distance being greater
than zero for a distance between a histogram and all other histograms.

Extracting a Histogram for Mammogram Comparison
The histogram distances will take into account whatever information is present in the
histogram without discrimination of its origin. To compare mammograms on the basis of
their breast tissue content it is therefore desirable to remove as much of the other regions as
possible, namely, the image background and the shadow from the pectoral muscle. The
poorly compressed near-skın fatty tissue is also discarded, its thickness being a function of
breast compression. The inner breast tissue is separated from the image background and
near-skın tissue by applying the minimum cross-entropy thresholding algorithm of Brink and
Pendock [101] to the image histogram as performed in Chapter 6. The algorithm is used in
preference to the skin-air interface segmentation algorithm in Chapter 8, as it is faster and an
accurate skin-line is not necessary. The pectoral muscle is extracted using the expanding
window algorithm from Chapter 9, the results produced by the combination of the Median-
based, Cross-Entropy Based and Global algorithms.
Figure 10.5: Examples of original images for each of the three breast types, with (a) dense-glandular, (b) fatty-glandular, and (c) fatty. Regions of interest from segmentation shown (d)-(f), and normalised region of interest histograms in (g)-(i).

Figure 10.5(a)-(c) shows an example of each mammogram type with (a) dense-glandular (mdb037), (b) fatty-glandular (mdb269), and (c) fatty (mdb309). The region of interest for each mammogram, determined from the segmentation described, is shown in Figure 10.5(d)-(f). The histogram of each region of interest is normalised so that the area under the curve sums to 1, making histogram comparisons independent of breast size. Normalisation of the intensity distribution along the bin axis is not performed as the machine settings and properties of the original films from which the images were scanned were not known. The normalised histograms of the region of interest are shown next to their image in Figure 10.5(g)-(i).

Algorithm Evaluation

Testing of several distance measures along with the new measure is performed using the MIAS classification under the assumption that images 'close' to a reference image will be of the same type. This type of testing is not perfect, as mammographic appearance covers a continuous spectrum and the risk of classification ambiguity when using a human classifier to
sort the mammograms into three set classes is always present. To see the results relative to any human classification, a benchmark can be set at the 65% accuracy level obtained by Blot and Zwiggelaar[148] on a subset of the MIAS database using Karssemeijer’s features, as mentioned previously. This comparative testing also implicitly measured the performance of the segmentation algorithms used to obtain the breast tissue, and thus provides further validation for the segmentation algorithms developed in this thesis. Testing is performed using the properties of ordered distance graphs and an image classification based on average histograms.

After an image is chosen the database can be ordered from closest to furthest image. Effectiveness can be seen by letting each image represent a point in a graph, colour coding each point based on image type, the graph of ordered images is expected to show clusters of colour. While examining order graphs provides a non-numeric overview of performance for a single image, several numerical measures of performance can be derived from order graphs, the advantage of these being the ability to average them to obtain an indication of accuracy for the whole database.

Examining the types of the closest images to the reference image in the order graph can be used to generate a confusion matrix for the database, resulting in a numerical indication of accuracy, although some overlap between the classes is expected due to human-based classification. An accuracy measure that can be gained from order graphs is the average distance in number of images that one image in a mammogram pair appears from the other. Since the breasts are viewed as one symmetric organ [9] the likelihood that the left breast for one case is more like the right breast of the same case than any other in the database is high.

While determining the dominant type within the closest images can be used to classify the image into a type another way is to compare the image histogram with a ‘typical’ histogram of each type and determine which one lies closest. To generate ‘typical’ class histograms, the histograms of images of the same type were averaged. Such classification produces a confusion matrix as in the closest images method with a speed advantage as a histogram is only compared against three histograms as opposed to those of the entire database. To avoid possible bias in the classification, the histogram of the reference image that is being classified is not included in the average histogram for its class.
Results

Ten distance measures are compared in Table 6, including the Closest Point Distance, determined with both the Transportation Problem using the Simplex algorithm and using the non-linear approximation, the scale is given as a ratio of x-axis range: y-axis range. The results from all algorithms show similar performance differences across the evaluation types.

Table 6: Algorithm performance using three performance measures.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average L-R Distance</th>
<th>5 Nearest Images</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>City Block (L₁)</td>
<td>40.36</td>
<td>62.24</td>
<td>62.42</td>
</tr>
<tr>
<td>Euclidean (L₂)</td>
<td>40.36</td>
<td>61.74</td>
<td>62.42</td>
</tr>
<tr>
<td>Cross-Entropy</td>
<td>40.30</td>
<td>61.93</td>
<td>62.42</td>
</tr>
<tr>
<td>Chi-Squared</td>
<td>46.70</td>
<td>61.93</td>
<td>62.42</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>46.70</td>
<td>58.45</td>
<td>55.59</td>
</tr>
<tr>
<td>Kuiper</td>
<td>37.30</td>
<td>60.50</td>
<td>62.73</td>
</tr>
<tr>
<td>Earth Movers Distance (13 bins)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDPA (Cha &amp; Srihari implementation)</td>
<td>41.11</td>
<td>60.99</td>
<td>57.45</td>
</tr>
<tr>
<td>Closest Point Distance (16 bins, 1:5 scale)</td>
<td>66.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closest Point Distance Approx. (1:7 scale)</td>
<td><strong>31.20</strong></td>
<td><strong>64.78</strong></td>
<td>65.53</td>
</tr>
<tr>
<td>Blot/Zwiggelaar using Karssemeijer features</td>
<td></td>
<td></td>
<td>65.09</td>
</tr>
</tbody>
</table>

To tune the distance scaling of the Closest Point Distance, the classification evaluation was used as this takes the least time, each image classification involving the comparison with three histograms. A graph of classification accuracy versus the intensity bin axis range as a multiple of the pixel count range of the average class histogram is shown in Figure 10.6(a) for the Closest Point Distance, solved using the Simplex algorithm with 9 bins, and in Figure 10.6(b) using the approximation on the original resolution histogram. Indicated on the figure is the classification accuracy obtained using the simple Euclidean distance (corresponding to accuracy as the bin axis multiple approaches infinity) and also the result obtained using the MDPA distance implementation in [156] (same as the EMD). It can be seen that as scaling increases, the accuracy reaches that of the Euclidean distance for both CPD and approximation. When distance along the intensity axis is scaled to zero, the True MDPA distance is obtained by the Closest Point Distance algorithm. It can be seen from the graph in Figure 10.6(a) that this does not correspond to the MDPA distance from the implementation in [156]. From the graph, two lower bounds exist on accuracy as a function of scaling, these are the accuracy of the Euclidean distance and the accuracy when scaling is zero (true MDPA for transportation algorithm). Also of note is that accuracy obtained by
normalising both axes so that one unit was equal to one standard deviation, resulted in classification accuracy of only 37.58%.

![Diagram](image)

Figure 10.6: Intensity axis range is transformed to a multiple of the pixel count axis range, the weighting affecting accuracy for the transportation problem based Closest Point Distance (a), and its approximation (b). As weighting gets higher accuracy approaches a limit equal to the 1 dimensional Euclidian distance, as weighting approaches zero, accuracy moves towards the True MDPA distance. The MDPA from [156] which is actually the EMD is shown.

The best results for the Closest Point Distance approximation were obtained with a scaling parameter of 7, however the scaling parameter leading to the most accurate Closest Point Distance solved using the transportation problem was also a function of the histogram resolution, the best results for the actual Closest Point distance obtained using 16 bins with a scaling parameter of 5.

![Table](image)

Table 7: Execution times to compare two histograms using the various algorithms.

<table>
<thead>
<tr>
<th>Method</th>
<th>Time for one histogram comparison (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>City Block (L₁)</td>
<td>1.00*10⁻⁴</td>
</tr>
<tr>
<td>Euclidean (L₂)</td>
<td>1.00*10⁻⁴</td>
</tr>
<tr>
<td>Cross-Entropy</td>
<td>0.0077</td>
</tr>
<tr>
<td>Chi-Squared</td>
<td>1.00*10⁻⁴</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.0055</td>
</tr>
<tr>
<td>Kuiper</td>
<td>0.0051</td>
</tr>
<tr>
<td>Earth Movers Distance (9 bins)</td>
<td>0.05</td>
</tr>
<tr>
<td>Minimum Distance Point Assignment</td>
<td>0.0028</td>
</tr>
<tr>
<td>Closest Point Distance (16 bins)</td>
<td>0.20</td>
</tr>
<tr>
<td>Closest Point Distance Approximation</td>
<td>0.0255</td>
</tr>
</tbody>
</table>
Due to their high complexity, both the Earth Movers Distance and the Closest Point Distance benefit from reduced resolution histograms. As an indication, the time to perform one histogram comparison for the Closest Point Distance and Earth Movers Distance when solved using the Simplex algorithm implemented under Matlab is shown in Figure 10.7 as a function of the number of histogram bins used. Even with a low number of bins, these are the most computationally intensive algorithms, as indicated in Table 7 which shows time taken for one histogram comparison for the algorithms reviewed. Results near that of the full resolution histogram can be obtained with relatively low-resolution histograms. The graph in Figure 10.8 shows Earth Mover Distance classification accuracy, as a function of the bin number showing stable results when resolution is greater than 17 bins. For the Closest Point distance accuracy depends also on the distance scaling. Identical classification results were obtained when the MDPA was used with the reduced resolution histograms showing that the MDPA is equivalent to the EMD.

![Figure 10.7: Execution time to compare two image histograms using the Closest Point Distance and the Earth Movers distance solved using the Simplex algorithm as a function of the number of bins used.](image-url)
Of further interest is the performance of the simple histogram distance compared to those using the cumulative histogram. While the Kolgomorov-Smirnov distance produces lower accuracy classification, the Kuiper distance outperforms the simple measures. Each of the three evaluation types is now examined in turn.

**Order Graphs**

Order graphs from several distance measures with the dense-glandular MIAS image mdb037 from Figure 10.5(a) as reference image are shown in Figure 10.9. Each bar in a graph represents one image, with the bar at position 0 the reference image. The length of each bar represents distance from the reference image and the bar’s colour indicates mammogram type. The graph is ordered to rank the image from closest to furthest from the reference image, clustering of colours indicates a relationship between type and histogram. Comparing the clusters based on algorithm type shows similar performance on visual inspection of images close to the reference image. Only the MDPA algorithm shows visibly worse performance for images furthest from the reference, the mixture of colours on the right side of the order graphs indicating confusion of classes. All algorithms show some degree of confusion. A sequence of ordered images taken using the order graph from the Closest Point Distance is shown in Figure 10.10. The distance from the reference image of each image is indicated, as is its type using the same colour coding as the graphs. The progression of images shows a change in appearance from dense to low density mammograms, exceptions include images ranked 165 and 242. These images are of the same type as the
reference image, however contain less of the low intensity tissue present in the reference image, hence judged dissimilar.

Figure 10.9: Order graphs for indicated distances using mdb037 of type dense-glandular as reference image. The graphs show a clustering of classes.
Figure 10.10: Closest Point distance: Images with distances (in number of images) from the reference image (mdb037) indicated, the reference image indicated as distance 0.
**Average Left-Right Distance**

The average distance between the reference image and its pair image shows again that the Closest Point Distance outperforms the other histogram distance measures. The distances are calibrated so that images adjacent to the reference image are at distance zero away from it. The use of average distances between images in a pair seems to indicate poor performance, the desired distance being zero as the pair should be the two most ‘like’ image in the database although the average distance for the best performing measure is 31.2 images (Closest Point Distance approximation). This high value for the mean distance arises due to the large tail in the closeness distribution, a graph of this distribution for the Closest Point Distance approximation shown in Figure 10.11. In this distribution there is a strong peak at zero distance, however because this is a one sided distribution, outliers at the far end move the average to 31.2.

![Figure 10.11: The mean inter-pair distance is high due to the large tail of the one-sided distributions, as seen from this graph for the Closest Point Distance Approximation whose average distance between pairs is 31.2.](image)

**Classification**

Both the five-nearest images, a simple form of classification, and the average histogram based classification show similar results. The advantage of using the average histograms is speed as only one histogram comparison is made for each class as opposed to comparing to the entire database. The average pixel-count normalised histograms for the breast tissue of the three MIAS classes are shown in Figure 10.12, although the actual histograms used varied with the reference image, as its histogram was removed from the average of its class to avoid possible bias. All histograms are similar in the low-intensity flat region but differ in the higher intensity peak in terms of its maximum amplitude and position, with less high intensity pixel represented in the fatty images than in the dense-glandular images.
Figure 10.12: Average histograms of the breast tissue region for the three MIAS classes used for image classification.

Classification performance is shown for the three classes in the form of confusion matrices in Table 8 for the six distance measures that the order graphs were shown for in Figure 10.9. The confusion matrices are given in terms of the percentage of images for each class (rounded to the nearest 1 percent). The confusion matrix obtained by Blot and Zwiggelaar using Karssemeijer’s features in [148] is also shown for comparison. However, it should be noted that these results were produced from a subset of 265 images of the MIAS database. Each row of the confusion matrix represents images of one class. The images of each class are distributed amongst the three cells of the row depending on how they were classified by
the algorithm, with the numbers indicating the percentage of images whose true class is of its row that were classified into its column. For example, in the case of the Cross-Entropy distance, out of the Fatty-Glandular images, 21% were classified as Dense-Glandular, 50% as Fatty-Glandular and 29% as Fatty. The results show that there is overlap between the classes with the Fatty-Glandular class confused between the two adjacent classes. The Dense-Glandular and Fatty classes are more reliably discriminated, the Closest Point Distance for example classifying only 7.14% of Dense-Glandular images as Fatty and 3.77% of Fatty images as Dense-Glandular.

Table 8: Confusion matrices generated by classification using 6 histogram distance algorithms, presented as class percentages rounded to nearest 1 percent. The confusion matrix from feature-based classification produced by Blot and Zwiggelaar is shown for comparison. Each data row represents images whose actual class is indicated in the left-most cell of that row. The type an image is classified as is indicated by the column.

<table>
<thead>
<tr>
<th>City Block</th>
<th>Cross-Entropy</th>
<th>Kuiper</th>
<th>MDPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>G</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>D 61</td>
<td>31</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>G 19</td>
<td>47</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>F 5</td>
<td>16</td>
<td>79</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CPD</th>
<th>CPD Approximation</th>
<th>Blot &amp; Zwiggelaar</th>
</tr>
</thead>
<tbody>
<tr>
<td>D 64</td>
<td>29</td>
<td>7</td>
</tr>
<tr>
<td>G 21</td>
<td>49</td>
<td>30</td>
</tr>
<tr>
<td>F 4</td>
<td>11</td>
<td>85</td>
</tr>
</tbody>
</table>

| D 63        | 28                | 10                |
| G 16        | 47                | 37                |
| F 4         | 9                 | 87                |

| D 56        | 36                | 8                 |
| G 17        | 58                | 25                |
| F 0         | 14                | 86                |

Overall, the Closest Point Distance and its approximation produced more accurate classification results than the other histogram distance measures and the feature-based classification using results from Blot and Zwiggelaar. From the confusion matrices, the Karssemeijer features are better at discriminating between Fatty-Glandular and Fatty images but not as good as some of the histogram-based approaches for discriminating Dense-Glandular from Fatty-Glandular images. Of note is the MDPA distances bias towards classifying images as Fatty, the Fatty-Glandular images being more likely to be classified as Fatty than any other type.
Conclusion

This chapter focused on the separability of mammograms into tissue classes, relating both to current applications for risk assessment and the possibility of using content based retrieval to serve images to a radiologist from a set of mammograms based on factors other than random order. The effectiveness of histogram-based image comparison and classification was investigated, comparing existing algorithms and introducing the Closest Point distance between histograms along with a fast approximation. Segmentation algorithms developed in this thesis were used to pre-process the images for classification, thus the performance of classification can also be interpreted as a measure of confidence in the segmentation algorithms.

The results indicate that it is possible to distinguish between mammogram classes on the basis of breast tissue histograms, results comparable to feature-based classification techniques which have been in the past the basis of such classification. Furthermore it was shown that for mammogram classification, the Closest Point Distance produces better performance, validated using several different performance measures on the MIAS database.

The successful classification results, which were obtained using the whole database from images pre-processed automatically without user intervention using segmentation algorithms developed in this thesis, shows that these segmentation algorithms are effective in the majority of cases.

Further Work

Of interest would be the application of the Closest Point Distance and its approximation to other image types, notably the colour images such as used in the work of Rubner et al. [155]. For colour images, the algorithm would be the same, but just as in the Earth Movers Distance the histogram bin locations would increase in dimension. Where Rubner’s histogram bins were located in the three dimensions of the CIE-LAB colour space, the Closest Point Distance, producer/consumer locations would be located in four-dimensional CIE-LAB/pixel count space.
CONCLUSION

In this thesis, several new methods for mammogram segmentation, based on pixel intensity, have been presented. Segmentation has been approached with a focus on hierarchical application, first using a simple image decomposition based on global thresholding and in each successive chapter refining the previous estimates and extracting more information using various algorithms ranging in complexity.

In each section, a comprehensive and critical review of existing methods was first undertaken to establish the progress of knowledge in that area and the current state of the art. After the analysis of existing work, new algorithms were proposed in several directions to segment the mammogram into its background, noise, breast and pectoral muscle regions. Chapters progressively build on the knowledge attained, with results of segmentation verified in the last chapter through accuracy testing on an application to classify mammographic density.

Summary of Research Outcomes
The segmentation process began with the segmentation of the image background and determination of the image orientation. Knowledge of image orientation simplified the segmentation of an accurate skin-air interface and also the pectoral muscle. The segmentation of the breast region and pectoral muscle in turn allowed the segmentation of the image database into different breast types and an evaluation of similarity between different breasts. The research outcomes from this thesis can be summarised as follows:

Orientation
Determining of orientation of the breast in the image has only been partially solved in the past, with algorithms limited to differentiating between left and right breast. In this thesis a new algorithm was developed which can also find breast orientation in rotated images and images upside down. The algorithm was successful in determining the orientation of all images in the MIAS database. Knowledge of image orientation is critical in many subsequent operations on mammograms, including determination of an accurate skin-air interface and
the pectoral muscle. In this respect the outcome presented here is useful both for algorithms developed in this thesis and also for algorithms in the literature that left orientation as a pre-processing step without addressing it.

**Background Segmentation**

The detailed analysis of minimum cross-entropy thresholding led to several algorithms to obtain an estimate of the image background, along with scanning artefacts and labels present on the image. Through interactive analysis of the different interpretations of minimum cross-entropy thresholding a global thresholding algorithm was chosen and coupled with a method for multi-level thresholding. Importantly, processing time was saved due to the finding that the less computationally intensive non-symmetric version of minimum cross-entropy was adequate to determine the threshold, rather than the full symmetric version. The simple background segmentation along with the orientation knowledge laid the first step in order for accurate determination of the breast border and pectoral muscle.

**Breast Border Segmentation**

The accurate segmentation of the breast border is a problem where solutions have been attempted before, attested to by the large literature review present in Chapter 8. An accurate solution has in the past been hard to obtain, sighted publications usually showing images with the determined boundary slightly inside the breast. A local thresholding technique and a polynomial fitting optimisation technique were proposed in this thesis. Both use the pre-processing steps of background noise removal and image orientation developed in this thesis. Also, both rely on the fact that the skin-line can be divided into a set of one-to-one functions, also determined in this thesis by examining the MIAS database.

The polynomial fitting technique, which resulted in more accurate skin-air delineation fits polynomials locally to sections of the skin-air interface obtained by thresholding. The threshold for the image and order of the polynomial fitted to the resulting boundary are changed so as to minimise the mean square error between the boundary pixels from thresholding and the fitted polynomial. The result is a very smooth skin-air boundary due to the inherent smoothness of the polynomial, although this same smoothness property leads to nipples in profile being cut off.
The entire MIAS database was segmented using the polynomial fitting algorithm and results compared to those from Chandrasekhar’s algorithm [89], which is used widely in the literature. Comparison was only subjective, performed by manual examination of images side by side with reference to the actual breast boundary as visible on the mammogram. Through this evaluation it was found that the polynomial-based algorithm produces better results in 29% of cases, while in 60% of cases results from both algorithms are acceptable.

**Pectoral Muscle Segmentation**

A new local thresholding algorithm developed in this thesis was used for the segmentation of the pectoral muscle. The orientation algorithm was needed to find the image corner closest to the pectoral muscle, after which the median intensity value, and that corresponding to a modified cross-entropy were examined as a function of window size. It was shown using a bi-modal pectoral muscle model, and confirmed on the MIAS database that an inflection point in the function indicates an intensity to separate the pectoral muscle from the adjacent region. The local thresholding was followed by straight-line fitting, with results from across the entire database compared using an automated algorithm to those obtained by Kwok et al. [145], who also used a straight line fitting algorithm and also a curve fitting algorithm. Results were mixed, the algorithm developed in this thesis showing better accuracy compared to the straight line fit of Kwok et al. but the fitted curve of Kwok et al. was judged the best. From this it can be concluded that the pectoral muscle may be better approximated by a fitted curve, although literature in the past that has used pectoral muscle location has usually relied on a straight-line approximation. The straight-line approximation was used in the subsequent step of mammogram classification to remove the pectoral muscle tissue from processing.

**Mammogram Classification**

The classification of mammograms into tissue density types was performed with accuracy results similar to those reported in literature. This work contributed to the thesis in a number of ways. Importantly, the output of segmentation algorithms developed could be used. Pre-processing of the image before classification that took place included orientation, background noise removal, pectoral muscle removal, and the extraction of the breast tissue. All processing was performed automatically using algorithms developed in this thesis. The resulting classification accuracy served as a benchmark for the segmentation to verify its accuracy.
A large contribution to knowledge, besides the field of mammogram segmentation and processing, was made through the analysis of existing histogram distance measures for the purpose of classification. While these have not in the past been used to classify mammographic density, they have been used in similar retrieve-by-example systems. Accuracy testing was performed on various existing methods by testing on the mammogram classification problem with a resulting ranking of algorithms. The idea of the Earth Movers Distance, the most sophisticated of the existing algorithms was re-interpreted and a new algorithm devised. This new algorithm, the Closest Point Distance outperformed the existing algorithms for mammogram classification.

**Suggestions for Future Work**

The work in this thesis covers several applications of mammogram segmentation; these in turn open possibilities for further extension. For example, the boundaries of extracted regions may be used as registration points for image comparison. Likewise, the generic algorithms introduced may find use in other fields besides digital mammography. The expanding window algorithm for the pectoral muscle may be used as a general local thresholding technique, and the closest point histogram distance may be applied to other image types, with colour and texture histograms added to the intensity information.

While the algorithms presented provide a sound basis for mammogram segmentation, their integration into any digital mammography system would require re-calibration and re-evaluation to tailor them to the specific imaging conditions and variations characteristic of the imaging hardware used. It is hoped that the algorithms and ideas presented in this thesis contribute not only to the field of digital mammography, but also find application in other image processing problems.
REFERENCES


246


GNU Image Manipulation Program (GIMP), www.gimp.org


