Statistical Models for Synthesising the Appearance of Mammographic Masses

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Medical and Human Sciences

2010

Michael Berks

School of Medicine

Contents

Word count: 79,938

List of Figures

Abstract

Breast cancer is the most common cancer in women worldwide. Many countries, including the UK, run screening programs in which x-ray mammograms of the breast are examined for signs of disease. The signs of breast cancer in mammograms may be very subtle, particularly in asymptomatic women. This makes reading mammograms a difficult task that is subject to human error. As a consequence it is vital readers are adequately trained and assessed throughout their career.

This thesis promotes the use of synthetic mammographic abnormalities as part of reader training and assessment. Using mammograms in which signs of breast cancer have been synthesised overcomes the problem of obtaining a sufficiently large volume of real data with known ground truth.

In particular, this thesis focuses on generating the appearance of malignant spiculated masses. Such masses are one the most frequent signs of breast cancer, but despite previous attempts, have yet to be synthesised successfully.

Statistical methods were used to model the appearance and location of a training set of real masses.

A global appearance model was used to encapsulate the variation in shape, texture and size of the central density present in each mass. Local texture models were constructed to describe the interaction between mass spicules and existing breast tissue in the region of the mass. Finally, a model describing the location of masses relative to breast shape was constructed. The models were used to synthesise the appearance of a malignant mass in an otherwise normal mammogram. By virtue of using generative statistical models, the synthesis process could be fully automated.

An observer study showed that masses generated by the method were identified as synthetic by expert readers at a rate significantly better than chance. However, this rate was similar to previous synthesis methods that required some level of manual input.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses

Acknowledgements

I would like to thank the following people:

My supervisor Dr Sue Astley, for her support, advice and unending positivity.

Chris Rose, Steve Caulkin and Chris Taylor for their invaluable advice at various stages of the project.

The readers at the Nightingale Breast Centre for their cooperation in experiments carried out during the project. Particular thanks to Caroline Boggis and Rumana Rahim for their help annotating images.

All the students in ISBE that have made it such a fun place to work. We might not have windows, but we'll always have cake!

Dad, Lyn, Liz and the boys for being the best family you could wish for.

Last but certainly not least, my fiancé Ali. Thank you for uncountable reasons. Without your support this thesis would not exist.

Dedication

This thesis is dedicated to the memory of my mum, Mary Berks.

Mary, a successful scientist in the field of biochemistry, died following a long battle with breast cancer, aged just 45 years old. She will always be an inspiration in everything I do, and I hope she would be proud of this work.

About the Author

Michael Berks received a First Class MSci in Mathematics and Computer Science from Imperial College London, graduating in June 2005. In January 2006 he began the research submitted to thesis in the Division of Imaging Science and Biomedical Engineering, at the University of Manchester University. During the course of this research, he published the following papers and abstracts related to this thesis.

- Michael Berks, Jennifer Diffey, Alan Hufton and Sue Astley. Feasibility and Acceptability of Stepwedge-Based Density Measurement. In Digital Mammography proc. of *8th International Workshop on Digital Mammography*, pages 355-361, Manchester, UK, 2006

- Michael Berks, Steve Caulkin, Rumana Rahim, Caroline Boggis and Sue Astley. Statistical Appearance Models of Mammographic Masses. In Digital Mammography proc. of *9th International Workshop on Digital Mammography*, pages 401-408, Arizona, USA, 2008

- Michael Berks, Chris Rose, Caroline Boggis and Sue Astley. Synthesising Abnormal Structures in Mammograms using Pyramid Decomposition. In Digital Mammography proc. of *9th International Workshop on Digital Mammography*, pages 143-150, Arizona, USA, 2008

- Michael Berks, Sue Astley, Steve Caulkin et al. Computer-based generation of masses in digital mammograms. An abstract presented as a digital poster at the *European Congress of Radiology*, Vienna, Austria, 2009

- Michael Berks, David Barbosa Da Silva, Caroline Boggis and Sue Astley. Evaluating the realism of synthetically generated mammographic lesions: an observer study. To appear in *Image Perception, Observer Performance, and Technology Assessment*, *SPIE Symposium on Medical Imaging*, San Diego, California, USA, 2010

- Jennifer Diffey, Alan Hufton, Michael Berks, et al. A stepwedge-based method for measuring breast density in the UK screening population: observer variability and comparison with human reading. To appear in *Physics of Medical Imaging*, *SPIE Symposium on Medical Imaging*, San Diego, California, USA, 2010

Chapter 1 : Introduction

1.1 Clinical Motivation

Breast cancer is the most common cancer in women worldwide. Each year approximately one million new cases are diagnosed [5]. In the UK alone, 11,990 women died as a result of breast cancer in 2007 [1].

To reduce mortality rates, many countries, including the UK, have adopted large-scale screening programs in which x-ray mammograms of the breast are examined for signs of disease. Such programs have been associated with a 35% reduction in mortality from breast cancer [2]. The current program in the UK invites all women between the age of 50 and 69 to screening once every three years. As a result, in the period from April 2007 to March 2008, mammograms were acquired for nearly 1.9 million women. The age range of women invited to screening is currently being expanded to cover all women aged 47 to 73. When this expansion is complete nationally, it will result in an additional 400,000 women to screen annually.

Thus every year, a vast number of mammograms must be examined for signs of disease. Whilst the last 20 years has seen great developments in computer algorithms that can automate this process, all mammograms in a clinical setting must be visually assessed by at least one human expert. Indeed because detecting signs of cancer in mammograms is a challenging task for even the most experienced breast radiologists, it is recommended best practice that all mammograms are examined independently by two expert readers. This process, known a double reading, has been shown to significantly improve the accuracy of detecting cancer [116].

However, double reading all mammograms acquired as part of a national screening program creates a large workload for individual readers. This workload, coupled with the complexity of detecting the subtlest signs of breast cancer, imposes great pressure on mammography readers. As a result, the way in which mammography readers are trained and evaluated throughout their career is of great importance, and it is primarily to this area we aim to contribute in this thesis.

In particular, we note that obtaining datasets comprising a sufficient range of mammographic abnormalities for teaching and assessment purposes may be problematic. We propose that synthesised mammographic lesions could be used in place of real data to overcome this problem. However if synthetic data is to be used, we must ensure that the lesions generated are indistinguishable from real examples. Moreover, for the full potential of synthetic data to be realised, we believe that any synthesis method must be able to run automatically so that very large datasets can be obtained without requiring significant human input.

In this thesis we develop a method for synthesising mammographic masses using a statistical modelling framework. Statistical models are constructed to describe the appearance of a real set of mammographic lesions so that by sampling from the model distributions, we can generate potentially infinite sets of synthesised lesions that match the appearance characteristics of our real data.

1.2 Thesis overview

In this section we provide a brief overview of the remaining chapters in this thesis.

Chapter 2 provides an overview of breast cancer, the imaging modality x-ray mammography and the national screening program currently implemented in the UK. This chapter sets out the clinical motivation for generating synthetic mammographic abnormalities that can be used in place of real data.

Chapter 3 presents a review of relevant literature from the field of computer-based mammography research. As a result of this review we conclude there has yet to be a method developed that satisfactorily synthesises malignant, spiculated breast masses. We note that partial synthesis (in which synthetic abnormalities are added to real normal mammograms) has proved more capable of generating realistic results than attempts to synthesise whole mammograms. Finally we identify a statistical appearance modelling framework we believe has the potential to generate synthetic masses that meet our requirements.

Chapter 4 includes a review of general methods in the field of image processing and analysis that will be useful for the methods we develop in the remainder of the thesis.

Chapter 5 describes the set of real data used throughout our work. These data comprise 540 mammograms supplied by the Nightingale Breast Centre, Manchester. Within the data are 101 malignant breast masses captured in either one or both mammographic projections. We describe the process by which these masses are annotated by a breast radiologist.

Chapter 6 presents a method in which each mammographic mass in our dataset is separated from the surrounding breast tissue. As a result of applying this process we effectively generate two new sets of image data: masses and mass backgrounds.

Chapter 7 describes the process of fitting a statistical appearance model to the set of masses. In particular we show how steps taken to optimise model fit have led to a model that generalises to unseen masses significantly better than a previously published model.

Chapter 8 includes a novel set of methods to model and synthesise the appearance of mass backgrounds. To obtain a rich set of descriptors of mammogram texture and structure, we apply the dual-tree complex wavelet transform (DT-CWT) to decompose our image data. We present a synthesis method in which DT-CWT coefficients of a normal region are modified, so that when the region is reconstructed it has the appearance of a real mass background.

Chapter 9 describes how the models of mass appearance and mass background appearance can be combined to produce a complete method for synthesising malignant mammographic masses in normal regions.

Chapter 10 presents a method for sampling a location within a normal mammogram at which a synthetic mass may be generated. The method incorporates an automatic segmentation of breast shape, so that given a normal mammogram, we can select a target location without user input. Coupled with the appearance synthesis method described in chapter 9, this produces a completely automated method for generating a synthetic breast mass in any normal mammogram.

Chapter 11 describes an observer study in which the realism of masses generated by

our method are assessed by expert mammography readers. Quantitative results of the study are given, in addition to qualitative feedback provided by the readers.

Chapter 12 concludes this thesis and provides a summary of the work we have presented. Possible areas for improving and extending the methods we have developed are discussed. The chapter finishes with a discussion on the contributions we believe our work has made to the field of mammography research.

Chapter 2 : Breast Cancer and Mammography

In this chapter we present background information on breast cancer and describe how the imaging modality of X-ray mammography can be used to screen for breast cancer.

Having discussed the process by which cancer may be detected in mammograms obtained as part of a large-scale screening process, we consider the pressures and workload this imposes on human readers of mammograms. In particular, we note the need for comprehensive training of readers and the need for efficient methods to evaluate their performance throughout their career. We describe the difficulties obtaining sufficient datasets with which to train and assess readers, thus motivating our goal to simulate signs of breast cancer in mammograms so that synthetic data may be used in place of real data.

2.1 Breast cancer

In this section we present a brief description of the adult female breast and the manner in which breast cancer develops.

2.1.1   Anatomy of the breast

The adult female breast consists of a mixture of parenchyma (functional tissue), fatty tissue and connective tissue [52]. The parenchyma is responsible for lactation and is composed of 15 to 20 lobes. Each lobe is a tree-like structure comprising a main duct that converges at the nipple and numerous branches that eventually terminate to lobules. The lobules produce and secrete milk which is drained through the ductal network to the nipple. The proportion of tissue types varies throughout a woman's life, with a general decrease in parenchyma and increase in fatty tissue with age, after full maturation. This process is referred to as the involution of the breast. In addition to the ductal network,

33

the breast contains a network of lymph vessels that drain lymphatic fluid away from the breast. The anatomy of the breast is depicted in Figure 2.1. To describe location, it is usual to split the breast along horizontal and vertical planes that bisect the nipple and run from the anterior to the posterior of the breast, thus forming the four quadrants labelled in Figure 2.1. The distribution of parenchymal tissue is not uniform across the breast, with the upper-outer quadrant containing the largest percentage.

Figure 2.1 The anatomy of a typical adult female breast
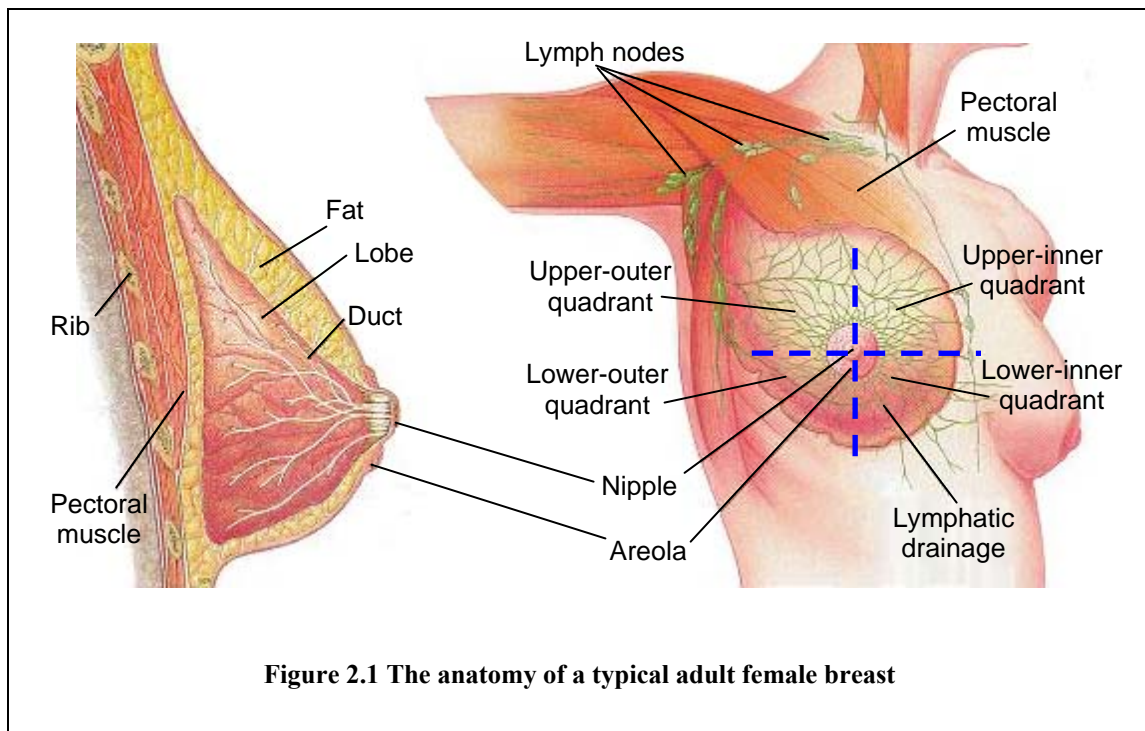
2.1.2   Breast cancer

Breast cancer is the most common cancer in the UK, and is the leading cause of deaths from cancer in women. In 2006, approximately 45,500 new cases were diagnosed, resulting in over 12,000 mortalities [1]. Although the disease can affect men, the vast majority of cases are in women. Incidences of breast cancer are very rare in women under the age of 30 years, with the rate rising steadily with age from then on. In the UK,

four in five new cases are diagnosed in women aged 50 and over [1].

The term cancer refers to a group of diseases that cause cells in the body to reproduce uncontrollably. Cancer cells divide rapidly and aggressively, and unless destroyed by medical intervention, result in death.

The following description of the progression of breast cancer is adapted from that given by Hayes [52]. Stage 1: Primary breast cancer begins as one or more cells of the parenchyma lose their normal regulation and proliferate aggressively, but remain confined by the duct or lobule membrane. At this stage the cancer is *in situ*. In the second stage, the cancerous cells break through the membrane of the duct or lobule, and invade the local tissue; the cancer has become invasive. Finally, in the third stage, the cancerous cells are transported via the lymph nodes or blood vessels, and may cause secondary cancers (metastases) throughout the body. However, breast cancer is an extremely heterogeneous disease, and the time at which the primary tumour develops from *in situ* to being invasive can occur at any time, from very early to very late in the cancer's lifetime.

If the primary tumour can be diagnosed and treated before it becomes invasive, there is a good chance localised treatment will successfully cure the disease. Invasive cancers that have metastasised require a systemic treatment of the body, which may ultimately prove ineffectual. Thus the earlier breast cancer is diagnosed the better the prognosis.

2.2 Mammography

In this section we describe how mammograms are acquired, consider the mammographic appearance of normal breast tissue and describe the appearance of

abnormalities that may be indicative of breast cancer. We describe the mammography screening program implemented in the UK and consider the pressure this puts on the human readers tasked with detecting breast cancer in mammograms. We conclude the chapter with a discussion of how reader training may be improved and describe the potential benefits of using synthesised mammographic abnormalities.

2.2.1   Mammogram acquisition

An x-ray mammogram is formed by compressing the breast between two radiolucent plates (see Figure 2.2 (a)). X-rays are directed through the breast and are attenuated by differing amounts, depending on the tissue they pass through. The x-rays are detected either by film or digitally, and a two-dimensional projection of the breast is generated. Three-dimensional information may be inferred by taking mammograms from different orientations, the two most common of which are the craniocaudal (CC) and mediolateral oblique (MLO) views. For a CC mammogram, the compression plates are placed horizontally above and below the breast, as shown by the red markers in Figure 2.2 (b). For an MLO mammogram, the plates are angled at approximately 45 degrees, with the lower end toward the centre of the chest, and the upper toward the shoulder, as depicted by the blue markers in Figure 2.2 (b). As a result of acquiring CC and MLO views of both breasts, a standard set of mammograms obtained during a single session consists of four mammograms. One such set is depicted in Figure 2.3. In general we will refer to the set of left and right, CC and MLO mammograms as a four-view mammogram case.

Figure 2.2 a) Diagram of mammogram acquisition b) position of the compression plates for a mediolateral oblique (MLO) mammogram in blue, and craniocaudal (CC) mammogram in red
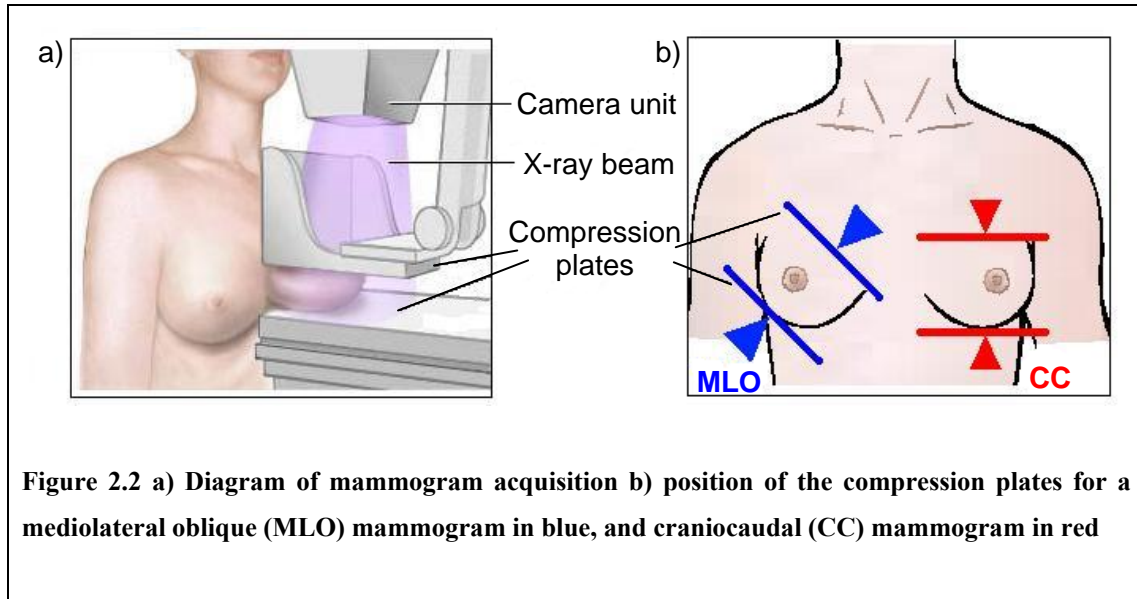
X-ray mammography is currently the only viable image modality for large-scale screening of asymptomatic women, meeting the key criteria of being low cost, delivering only a low dose of radiation, and providing very high spatial resolution in the image produced. This ensures that in attempting to detect breast cancer, we minimise the increase in risk of developing cancer associated with x-ray radiation. A high spatial resolution is required to detect the smallest abnormalities (such as microcalcifications discussed in section 2.2.4), that may be very early signs of cancer.

As noted above, mammograms may be acquired on film or as part of a fully digital system. Full-field digital mammography (FFDM) has been shown to improve the detection of cancer (particularly in younger women with denser breasts) [94, 110, 111], whilst reducing the cases in which repeat images must be acquired due to technical problems in the acquisition process. With additional advantages afforded by storing, transferring and retrieving data digitally, FFDM systems are replacing film mammography units throughout the UK, with the aim that every screening centre has

access to at least one FFDM system by 2010 [3].

However, at the time of commencing this project, the majority of screening mammograms currently acquired in the UK were acquired as films. To apply computer algorithms to film mammograms, the films are first digitised by dedicated mammography scanners. As discussed in chapter 5, the mammograms used throughout our project are digitised film mammograms. Thus in the next section we consider the relationship between the tissue structures in a breast and the appearance of a digitised mammogram.

2.2.2   Relating breast tissue and mammographic appearance

When viewed on a light box, the appearance of a film mammogram results from variations in density across the film. When a film is digitised, the scanner converts the range of densities into pixel grey levels. Thus in order to understand the appearance of different tissue types in a digitised mammogram, we need to examine how film density relates to the imaging process.

To begin we assume an (x, y, z) co-ordinate system such that during mammogram acquisition the film lies in the (x, y) plane and x-rays pass through the compressed breast along the z-axis, and thus perpendicularly to the film. In this system, the intensity of the x-ray beam after passing through any location (x, y) in the projection of breast can be written as

$$I\left(x, y\right) = I_0\left(x, y\right) \exp\left[-\int \mu\left(x, y, z\right) dz\right]$$

where I_0 is the incident intensity of the x-ray beam and $\mu(x, y, z)$ describes the

attenuation coefficient of tissue at the point (x, y, z) in the breast.

The exposure of the film at this location, $E(x, y)$ is computed as the sum of intensities imparted on the film during the acquisition process. Thus if $I(x, y, t)$ is the intensity at any given time, we can write

$$E(x, y) = \int I(x, y, t) dt = E_0 \exp\left[-\int \mu(x, y, z) dz\right]$$

Moreover, by assuming $\mu(x, y, z)$ is constant during image acquisition, we can write the intensity at any given time as

$$I(x, y, t) = I_0(x, y, t) \exp\left[-\int \mu(x, y, z) dz\right]$$

and so can rewrite the exposure of the film as

$$E(x, y) = E_0 \exp\left[-\int \mu(x, y, z) dz\right]$$

where E_0 is the integral of the incident intensity over time and is defined as the incident exposure.

It is the exposure of the film to the x-ray beam that changes the density of the film. The relationship of the change depends on the specific type of film-screen combination used. However, the film density will be directly proportional to the log of the exposure for a range of exposures. If the mammogram has been correctly acquired, the exposure of the film at any location in the projection of the breast will lie within this log-linear range, in which case we can write the film density at (x, y) as

$$D(x,y) = \alpha \log\left(E_0(x,y)\exp\left[-\int \mu(x,y,z)dz\right]\right) + \beta$$
$$= -\alpha\left(\int \mu(x,y,z)dz\right) + \alpha \log\left(E_0(x,y)\right) + \beta$$

(2.1)

where the values of the constants α and β depend of the film-screen combination.

Thus if a mammogram has been correctly acquired, the density of the film at any location in the breast has a linear relationship to the sum of superimposed tissue attenuation coefficients at that location. Moreover, if (as in the scanner we use) film densities are converted linearly into pixel grey levels, this relationship is passed on to the digitised mammograms. Finally, we note that in general, film densities are inverted during digitisation, so that high attenuation coefficients correspond to high grey levels and vice-versa.

We can now describe the appearance of particular breast tissue types in digitised mammograms.

2.2.3   Mammographic appearance of breast tissue

As described in section 1.1, the three main components of an adult female breast are fatty tissue, parenchyma and connective tissue. Their mammographic appearance is described below.

Fatty tissue: fat is largely composed of water, and so has a very low x-ray attenuation coefficient. As a result fatty regions of the breast appear as areas of low grey level in a mammogram and do not obscure denser tissue or tumours.

Parenchyma: the parenchyma is much denser than fat, and thus has a higher attenuation coefficient. As a result the superimposition of parenchymal tissue may obscure

malignant lesions present in the breast. For this reason mammography is less suitable for younger women, in whom the breast has a high percentage of parenchyma. The ducts of the parenchyma appear as bright curvilinear objects on the mammogram. The lobules may also be seen as small circular objects, although in older women, the lobules have usually diminished and are not detectable. In denser breasts the projection of the parenchyma may form a near unbroken high intensity region in the anterior of the breast, sometimes described as the fibro-glandular disc.

Connective tissue: ligaments supporting the breast may also be visible with a similar appearance to the parenchymal ducts.

Blood vessels and lymph nodes may be visible in mammograms. In the MLO view, the pectoral muscle appears as a region of denser tissue on the upper chest wall corner of the image.

The two-dimensional projection of complex three-dimensional tissue structures, combined with the high resolution of mammography, produces an image that has very few homogenous regions within the breast, and is rich in linear structures at multiple orientations.
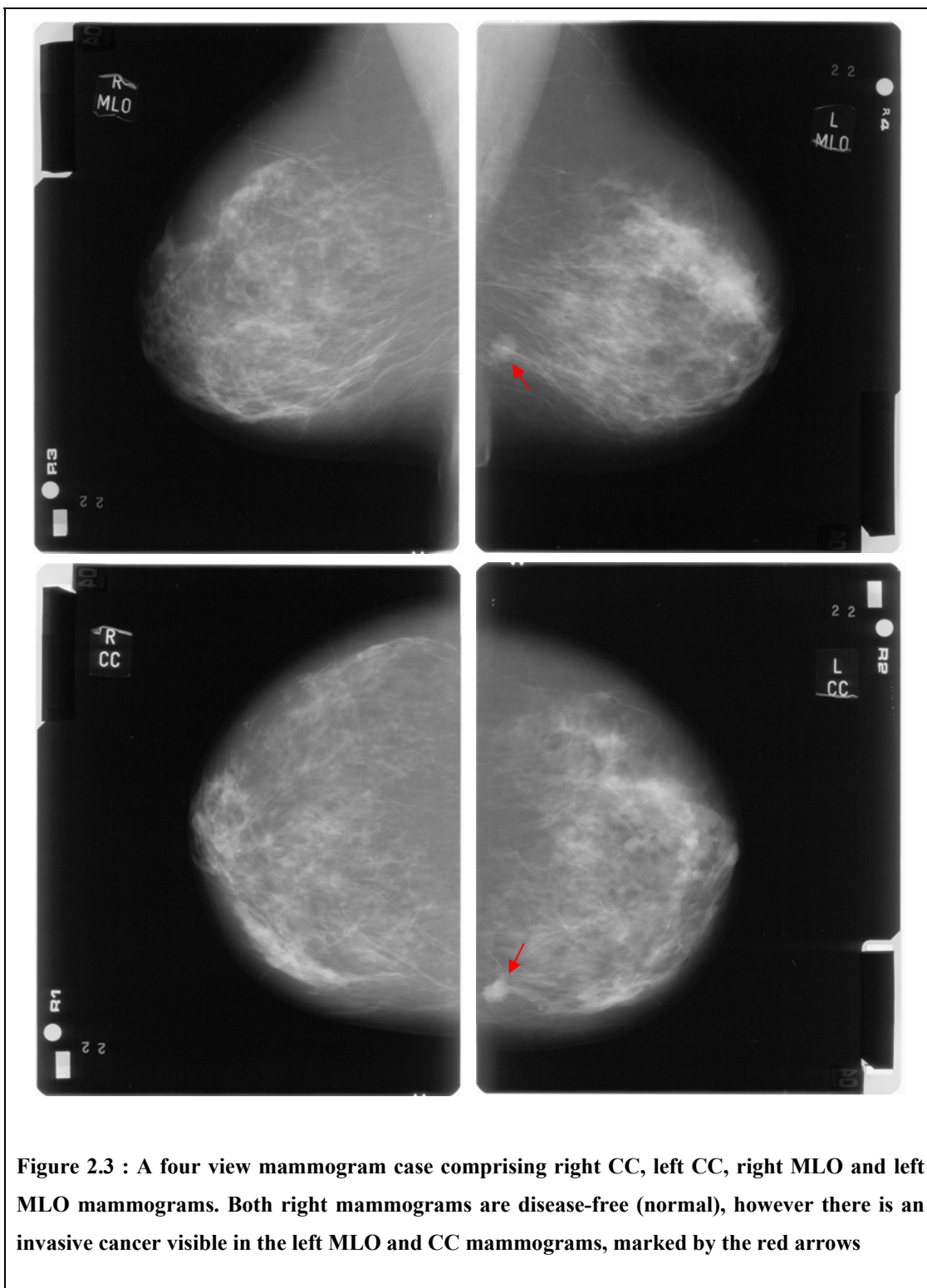
Figure 2.3 : A four view mammogram case comprising right CC, left CC, right MLO and left MLO mammograms. Both right mammograms are disease-free (normal), however there is an invasive cancer visible in the left MLO and CC mammograms, marked by the red arrows

Figure 2.3 depicts the CC and MLO mammograms of the right and left breasts from a four view mammogram case. In this instance the right breast is disease free, and thus

right MLO and CC mammograms are normal. However, there is a biopsy proven malignant mass in the left breast. The mammographic appearance of signs indicative of breast cancer, along with the appearance of other abnormalities in the breast, is described in the next section.

2.2.4   Mammographic appearance of breast abnormalities

In this section we describe the mammographic appearance of abnormalities in the breast.

Before proceeding we make two points. Firstly, not all abnormalities present in mammograms are signs of breast cancer. Many lesions are caused by benign processes such as cysts, or benign tumours. Secondly, there is no consistent universally applied system for classifying and reporting the appearance of lesions in mammograms. The American College of Radiology introduced the Breast Imaging Reporting and Data System (BIRADS) [14] aiming to standardise a lexicon of terms with which to report abnormalities in screening mammograms. However, whilst this has been formally adopted in the USA, it has not seen widespread take-up in the UK. The descriptions of lesions we give below are based on the classes that we believe are commonly used by UK mammogram readers, based both on our observations of their work in a clinical setting and through working with readers during our project. For reference purposes, the descriptions we give are combination of those given by Tabar and Dean [115], and those included in BIRADS [14].

We classify abnormalities into five groups: circumscribed lesions; spiculated (or stellate) lesions; calcifications; architectural distortions; and breast asymmetries. A description of the mammographic appearance of each class is given below.

Circumscribed lesions are characterised by their density, boundary, form and size. The density of a lesion is categorised into four groups: radiolucent (less dense than the surrounding parenchyma); low density radio-opaque (approximately the same density as parenchyma); high density radio-opaque (denser than parenchyma) and radiolucent and radio-opaque combined. The greater the density of a lesion, the higher its grey levels will be in a digitised mammogram. All radiolucent and combined density lesions are benign, as are most low density radio-opaque lesions.

The boundary of a circumscribed lesion may be indistinctly outlined or sharply outlined, possibly with an additional halo sign (a narrow radiolucent ring) or capsule (a narrow radio-opaque ring) surrounding the lesion boundary. Sharply outlined lesions, particularly those with halo signs or capsules are likely to be benign.

The lesions may have circular, ovular or lobular form and can vary size, from very small (1cm or less) to very large (greater than 5cm). A benign circumscribed lesion is shown in Figure 2.4 (a).

Spiculated lesions are radiating structures with ill-defined boundaries. The lesion centre may be a distinct solid tumour or contain oval or circular radiolucent areas. The radiating structures may have two basic forms: sharp, fine, dense lines (spicules) radiating in all directions, with the length increasing as the central mass increases or very fine, less dense spicules grouped together (we will refer to such groups of spicules as "sheaf-of-wheat" spicules). In general, spiculated lesions are highly likely to be malignant. Figure 2.4 (b) shows a malignant mass with an array of fine, dense spicules radiating from the centre of the mass.

Calcifications appear as small high intensity objects in a mammogram, usually caused by benign processes [115]. Microcalcifications are tiny (normally less than 1mm) high intensity spots that may be grouped into clusters and are usually malignant. The shape of individual microcalcifications may be granular or irregular. Although microcalcifications are generally much brighter than surrounding tissue, their tiny size may make them very hard to distinguish (being mistaken, for example, as image noise). As discussed in section 2.2.1, detecting microcalcification requires that mammograms are acquired at a very high spatial resolution. Figure 2.4 (c) depicts a malignant cluster of microcalcifications.

Architectural distortion refers to an abnormal pattern of parenchymal tissue. This pattern may appear as thin lines or spiculations radiating from a point, or as focal retraction at the edge of the parenchyma. Whilst architectural distortion is often associated with other signs of disease, it may occur when a focal mass is either not present or not visible. Architectural distortion may be due to previous trauma or surgery to a breast, however, if no such history is recorded distortions are suspicious for malignancy. Architectural distortions not associated with a visible mass may have a very subtle appearance in the breast, and can often be overlooked as normal patterns of parenchymal tissue. Indeed, a study by Burrell et al. [20] suggested that architectural distortion was the most commonly missed disease sign in interval cancers. Figure 2.4 (d) depicts a region of architectural distortion in which a focal mass is not visible.
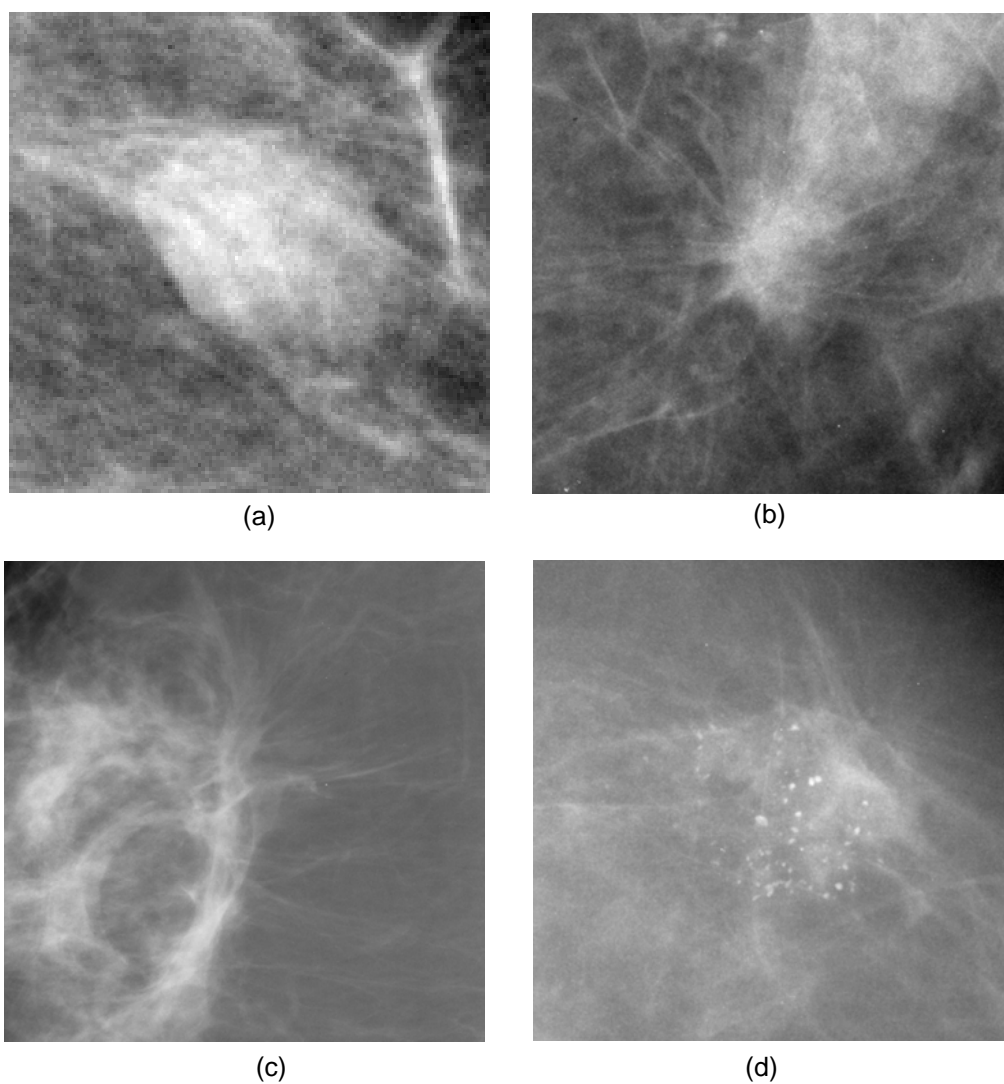
Figure 2.4: a: A circumscribed mass; b) a spiculated mass; c) a region of architectural distortion; d) a cluster of microcalcifications

Breast asymmetry, as the name suggests, refers to a difference in appearance between mammograms of the left and right breast. Global asymmetry refers to a general difference in appearance of breast tissue, for example one breast may have a much greater volume of parenchymal tissue than the other. Local areas of asymmetric tissue (sometimes described as focal asymmetries) may be caused by a mass or architectural

distortion. However, the mass or distortion may not be clearly visible due the superimposition of other breast tissue. Whether asymmetries are treated as suspicious depends on the presence of other signs of disease, and, if earlier mammograms are available, the extent to which an asymmetry has changed temporally. For reference, a large number of examples of asymmetrical breasts are presented in the Miller's PhD thesis [82].

2.2.5 Mammography screening

The earliest record of breast radiography dates back to 1913 and by the 1960s mammography was recognised as a useful tool with which to screen for breast cancer in asymptomatic women [15]. In the late 1980s the practice had developed to the point where several large scale national screening programs were established. These included: Sweden in 1986; Iceland in 1987; and the UK, Canada and Australia in 1988 [106]. In this section we focus on the screening program currently implemented in the UK, although more information on the screening policies of other countries can be found in the report by Shapiro et al. [106].

The National Health Service Breast Screening Program (NHSBSP) was established in 1988 in response to the recommendations of the Forrest report [46], with the aim of reducing breast cancer mortality rates in the UK. Breast cancer may be detectable in x-ray mammograms before palpable symptoms occur. Screening helps detect breast cancer at the earliest possible stage, with the latest figures showing approximately half of breast cancers in the UK in the 50 to 69 years age group are detected as a result of the NHSBSP. It is estimated that this saves approximately 1,400 lives in England in each year [2].

In 2005, the upper age limit in the program was extended from 65 to 70, so that all women between the ages of 50 and 69 are now invited for screening every three years. Under this scheme, in year April 2007 to March 2008 nearly 1.9 million women were screened in the UK, resulting in the detection of over 16,000 invasive breast cancers.

In 2009 a further extension of the age limits to include all women between the ages of 47 and 73 was trialled regionally. This latest increase in age range is due to be expanded nationally by the end of 2010, resulting in a further 400,000 women being invited for screening annually.

In the following section we describe the process by which the mammograms acquired during the NHSBSP are analysed for signs of breast cancer.

2.2.6   Reading mammograms

In this section we outline the process of reading mammograms, where reading is defined as the process in which a human observer visually assesses a mammogram for signs of disease. We discuss the current reading guidelines recommended in the NHSBSP and consider the workload this enforces on mammogram readers[1].

For each mammogram acquired in the screening program a reader must first decide if any abnormality is present. If the mammograms acquired in a single case are normal, then no further action need be taken. However, if any abnormalities are found, then a decision must be made on whether the abnormality is a benign finding, or whether it is

---

[1] As we note later in this section, in the UK mammogram readers are not confined to breast radiologists, and therefore we avoid using the term radiologist synonymously with mammography reader. However, it is current practice that at least one consultant radiologist assesses every screening mammogram.

indicative of disease. If any abnormality is suspected of being malignant, then the woman will be recalled for further investigation. This may include repeat imaging or a needle biopsy, to confirm a diagnosis. We note however, that for our work we are interested in the reading process up to the point at which a decision on the initial screening mammograms is made, and thus reviewing such further investigation is beyond the scope of this thesis.

When assessing mammograms, readers face two competing goals. Naturally they must aim to detect as many cancers as possible. However they must minimise the number of women without cancer that are unnecessarily recalled for further investigation. These false positive cases may cause undue stress to the women involved. Moreover, a high numbers of false positives across a population may lead to a reduction in invitation acceptance, to the detriment of the screening program. In statistical terminology, the goal of detecting true cancers is described as maximising sensitivity, whilst minimising false positives is known as maximising specificity.

The abnormalities described in section 2.2.4 may be very subtle in mammograms, particularly as many of the cancers present are at an early stage of development. As a result mammograms may be misdiagnosed, even by experienced expert readers. Therefore in the UK, it is recommended best practice that all mammograms are assessed independently by at least two expert readers. This process, known as double reading, has been shown to increase overall sensitivity without reducing specificity [116].

Given the number of mammograms acquired during the NHSBSP, double reading places a massive workload on human readers. Indeed individual readers in the UK typically read between 5,000 and 10,000 films a year (see section 11.5.1). Moreover this

workload is due to increase when the current extension in the age range of the NHSBSP is complete. Whilst recent studies have shown that combining computer-aided detection (CAD) algorithms with a single human reader may provide an alternative to double reading [49], such CAD systems are not yet widely used in the NHSBSP.

From the evidence outlined above, we see that reading mammograms is a visually demanding task – both in terms of the difficulty associated with detecting subtle signs of disease in individual mammograms and the sheer number of mammograms a reader must process. This places a great burden on how readers are trained and evaluated throughout their career. Training must be comprehensive to ensure readers are of sufficient expertise whilst being streamlined to ensure the staffing requirements imposed by double reading are maintained. Meanwhile qualified readers should be evaluated throughout their career to ensure they are performing at a sufficient standard. Currently, all UK readers are invited to take part in a bi-annual self-assessment program (PERFORMS) [32]. In this assessment, readers are presented with a set of films that have been specially selected to contain challenging cases. Ground truth for the films is obtained by consensus of an expert panel. As with training, this process must be streamlined to ensure such evaluation doesn't end up significantly adding to a reader's workload.

In the following section, we discuss how we think synthetic data can be used to improve methods by which mammogram readers are trained and evaluated.

2.2.7 Using synthetic lesions to train and evaluate mammogram readers

Trainee mammogram readers must become familiar with the appearance any

abnormality that could be found in a mammogram. Therefore it is essential they have access to a large set of mammograms containing a wide range of abnormalities. Moreover, to enable the provision of feedback and to evaluate progress, the ground truth of any mammogram must been known. That is, if there are one or more abnormalities present in a mammogram, then at the very least the location and type of abnormality must be available. In a learning environment, it may also be beneficial if more detailed information is available (for example, a precise demarcation of a lesion boundary).

Similarly, a dataset of mammograms containing abnormalities with known ground truth is required to quantify the performance of current readers. Moreover, in any assessment exercise (whether for trainee or qualified readers), a mammogram should only be seen once by any given reader. Once a mammogram has been seen, the extent to which a reader remembers the content of the image (whether consciously or subconsciously) may bias any further readings of the image. Therefore any assessment program must have access either to an extremely large dataset of suitably labelled mammograms or ensure that their current dataset is continually updated.

Obtaining sufficiently large sets of real data that meet the requirement outlined above may prove problematic. As described in section 2.2.5, huge numbers of mammograms are acquired each year. However because the mammograms are obtained from an asymptomatic screening population, the large majority are normal. For example, of the women screened in 2008 as part of the NHSBSP, only 9 in every 100 mammogram cases contained some form of abnormality that required further investigation, whilst cancers were detected in only 8 in every 1000 cases [4]. Thus simply obtaining a sufficient number of abnormal cases (and particularly cases with signs of cancer) for

training and assessment may be difficult.

Even if a large enough set of data can be acquired, it may not be possible to obtain the necessary ground truth for the data without significant input from an expert mammography reader. We propose that using synthetic mammographic abnormalities in place of real data would be an efficient solution to this problem.

By construction the ground truth of any synthesised abnormality is known. Moreover, once a method for synthesising data is established, new abnormalities can be generated to expand a dataset to any size and ensure that data are continually updated.

Of course, for synthetic data to be used clinically in training and assessing of radiologists, we must ensure that synthetic abnormalities share all the appearance characteristics of real abnormalities. In addition, we note that one of the major benefits we see in using synthetic abnormalities is the ability they give us to create very large sets of data. This advantage is mitigated if significant user input is required to generate each new synthetic example and therefore, from the outset, we propose that any method used to synthesise mammographic abnormalities should be able to run automatically to create large sets of data.

For reader training, we note that a method to automatically generate synthetic mammographic abnormalities could be incorporated in an intelligent tutoring system (ITS). Broadly speaking an ITS is a software system that interacts with a user, so over some period over training, a user is presented with problems specifically tailored to their needs. Such systems have been successfully employed to various learning tasks in medicine [11, 33, 41, 107]. In particular we note a system RadTutor [11], developed by

Azevedo and Lajoie, designed to train a user to differentiate between real and benign abnormalities in mammograms. Throughout training, RadTutor selects the most appropriate abnormality to display to a user, based on their previous performance. However, the system has only a limited supply of abnormalities from which it can choose. Therefore we envisage that a method capable of generating, on demand, an abnormality of some specified characteristic would be of great benefit to this form of system.

2.2.8   Using synthetic lesions to test mammography equipment

In addition to improving reader training and assessment as described in the previous section, synthetic breast lesions may be used evaluate the performance of mammography equipment. For example, both Saunders and Samei [101, 102] and Ruschin et al. [99] describe how synthetic abnormalities can be used we evaluating image quality in digital mammography systems.

Finally, where computer-aided detection algorithms (see section 3.1) are based on learning the characteristics of real mammographic lesions, we propose that synthesised lesions may be used to boost the size of training data, consequently increasing the performance the algorithm.

2.3 Summary

In this chapter we presented background information on mammography and breast cancer relevant to our project.

We gave a brief description of the anatomy of the female breast, and have described the process by which breast cancer develops. We introduced x-ray mammography, and in particular, considered how it is used as a screening tool to detect breast cancer at the

earliest possible stage. The mammographic appearance of breast tissue and breast abnormalities were described.

We outlined the screening program implemented in the UK (the NHSBSP) and highlighted the vast numbers of mammograms that are acquired annually. The process by which mammograms acquired in the NHSBSP are visually assessed for signs of breast cancer was described in section 2.2.6. We noted that reading mammograms was a complex and difficult task and that double reading all films acquired in the NHSBSP imposed a great workload on individual readers. As a consequence, we highlighted the importance of comprehensive yet efficient training and assessment of mammogram readers. In section 2.2.7, we discussed how synthetic data could be used to streamline the teaching of trainee readers and make optimal use of limited supplies of real data, thus providing the motivation for work presented in this thesis. Finally, in section 2.2.8, we noted other possible uses of synthetic lesions in mammography research.

As a result we state the following goal of our project: we aim to develop a method for synthesising the signs of breast cancer in mammograms such that

  i)   the synthesised lesions we generate are indistinguishable from real examples

  ii)   the method can run automatically to generate very large sets of data

In the following chapter we present a thorough review of previous attempts at synthesising mammographic lesions and discuss why we think there has yet to be method that satisfactorily meets the two requirements stated above.

Chapter 3 : Computer-Aided Mammography

In chapter 2 we presented our motivation for synthesising mammographic abnormalities. In this chapter we review background literature related to this area. The review is structured as follows: in section 3.1, we give a brief review of the ways in which computer-based research has aided the field of mammography. Section 3.2 then includes a thorough review of previous work attempting to synthesise breast lesions in mammograms. We discuss the limitations of existing synthesis methods, and highlight the areas in which we feel further development is required.

As a result of this review we identify other areas of computer-based mammography research that may be of benefit to our project. Specifically, we consider methods in which breast tissue structures are detected, modelled or classified (section 3.3.1); and methods in which the main breast shape is automatically segmented in mammograms (section 3.3.2).

3.1 A history of computer-aided research in mammography

As described in section 2.2.5, mammography has been used as a screening tool for breast cancer since the 1960s, with large-scale, national screening programmes established across the world from the late 1980s onwards. In parallel to the development of the imaging modality, research in the fields of computer science and image analysis has attempted to aid the detection and diagnosis of breast cancer in mammograms.

Within this work, the most prominent application of computer-based research has been the development of automated methods to detect signs of breast cancer. Indeed the first computer-aided detection (CAD) algorithm was published in 1967 by Winsberg et al.

[125]. Since then, research has developed to the point where in 1998 the first commercial CAD systems gained approval FDA approval to be used in a clinical setting in the United States [48]. There are now several commercial CAD systems being used clinically throughout the world. In general, such systems have been designed to provide prompts for a human reader. That is, the CAD system automatically flags locations in a mammogram that may be suspicious. The reader can then review these prompts before making a final decision on whether an abnormality requiring follow-up is present.

In screening programmes where reading by a single observer is the norm (for example, the United States), CAD systems may improve the performance of a single reader. Indeed, a recent study has estimated that CAD systems are now used in 25 to 30% of all mammograms read in the United States [74]. Meanwhile, in countries where double-reading is the normal practice, it has been proposed that single-reading with CAD could be used as an alternative to double reading, thus reducing workload. Recent large-scale clinical trials [49], have shown this may be feasible.

Related to computer-aided detection is computer-aided diagnosis. As the name implies, computer-aided diagnosis concerns making a decision once an abnormal finding has been detected. This may involve determining between a benign and an abnormal finding, or classifying a detected abnormality more precisely. Diagnosis algorithms may be incorporated within computer-aided detection systems (for example, commercial systems such as ImageChecker DMax by Hologic/R2 Technology estimate the likelihood of cancer for any abnormalities they have detected [58]) or used as independent tool by radiologists to help determine future management of a patient [60].

It should be noted that whilst methods have been developed to detect all signs of breast

cancer in mammograms, detection of masses and microcalcifications has proved more successful than that of the more subtle abnormalities such as architectural distortion. Commercial CAD systems usually target microcalcifications and masses, and a study by Baker et al. [13] showed that two of the most widely available commercial CAD systems identified less than half of the cases of architectural distortion in their test data. As a result, much recent research in CAD has focused on detecting abnormal patterns of breast tissue that may be indicative of disease regardless of whether a central mass is present [63, 88, 96]. A more detailed review of this work is given in section 3.3.1.

Aside from computer-aided detection and diagnosis, techniques in image processing and analysis have been used to aid a variety of other tasks in mammography. For example, as described in sections 2.2.7 and 2.2.8, synthetic mammographic abnormalities may be used in several areas, including reader training and assessment and the evaluation of mammography equipment. Additionally, computer analysis has proved beneficial in assessing breast density (that is, the amount of parenchymal tissue in the breast). Whilst not an immediate marker of disease, increased breast density has been shown as an important risk factor in developing cancer [19]. Automated methods have been developed to either classify mammograms into groups by density [92] or more recently, to provide a volumetric quantification of dense breast tissue as imaged in a mammogram [35, 55, 89]. Such measures can then be included in risk analysis models that may help determine the management of women at high risk of developing the disease.

Finally, we note that to enable top level tasks such as detecting disease or quantifying breast density, a large number of lower level image processing and analysis techniques

have been developed for mammograms. These include: methods to segment the main breast shape [16, 25, 44, 81, 85, 86, 126] and other global structures (for example, the pectoral muscle or fibro-glandular disc) [45, 62, 69]; methods to detect and/or classify local linear structures (for example, glands, ducts, vessels and spicules) [24, 63, 78, 88, 103, 121, 129, 130]; methods for registering one mammogram to another (for example, comparing left and right views of the breast to determine asymmetries or registering temporal pairs of mammograms to assess changes in parenchyma) [118]; and methods for analysing and classifying areas of mammographic texture [62, 91, 92, 98, 120].

Where we believe such methods have a direct relevance to the work in our project, these are reviewed in section 3.3. However, in the following section, we turn to the primary goal of our thesis and focus specifically on existing methods for synthesising mammographic abnormalities.

3.2 Synthesising signs of breast cancer

This section presents a thorough review of literature relating to the synthesis of mammographic abnormalities, and is intended to describe the current state-of-the-art in the field.

Synthesising microcalcifications

Kallergi et al. [61] described a method for synthesising microcalcifications with characteristics based on the BIRADS reporting system used by American radiologists. The number of microcalcifications in the cluster and the contrast between an individual calcification and the background grey levels were determined from a set of 100 real clusters (although how exactly these data were used to determine the features is not described). New clusters were then created by first selecting the number of

58

calcifications (a random number between 3 and 15) and then manually drawing the shapes of calcification boundaries in a binary grid, with a random contrast (in the range 0.04 to 0.20) assigned to each cluster. The positions in real mammograms at which synthetic clusters were superimposed were selected manually. The grey levels of the calcifications were then computed using the associated background grey levels and the calcification contrast values. Each cluster was smoothed with a randomly chosen Gaussian template before being inserted onto the background. Realistic results were claimed, however no statistical evaluation of these results was given.

Lado et al. [70] also synthesised microcalcification clusters and inserted these onto real mammographic backgrounds, but used a method more closely based on real data. The process first automatically detected microcalcification clusters in a set of real mammograms. Features of the cluster (size, shape, density) and individual microcalcifications (size, shape, grey levels) were extracted to create a set of model parameters. New microcalcifications clusters were generated by randomly sampling from the set of features, and were manually inserted into real mammographic backgrounds. To evaluate the method, two radiologists and one physicist rated the realism of synthetically generated clusters. ROC curves were fitted to the responses and the area-under-curve (AUC) calculated. A mean AUC of 0.54 ±0.03 suggested there was no significant difference in appearance between real and synthesised clusters.

Nappi et al. [84], synthesised microcalcifications by generating three-dimensional models of calcification clusters and then projecting the clusters onto mammographic backgrounds using a simulated model of x-ray acquisition. Individual microcalcifications were created using one of three basis models: irregular, ovoid and

elongated. A three-dimensional model of the breast duct network was generated using branching trees, and the individual microcalcifications were placed at the nodes of network simulating the way in which microcalcifications are believed to form in real breasts. Results were shown in the form of images, however no quantitative analyses of the realism of synthetically generated clusters were reported.

Synthesising mass lesions

An alternative method for generating synthetic mammograms from a three-dimensional model was presented by Bliznakova [17]. Whilst the method was designed to synthesise masses and microcalcifications, the main focus of the work was generating three-dimensional models of the main breast structures, and projecting these to generate a full two-dimensional mammogram. However, both lesions and the synthetic mammograms in which they were contained were easily distinguishable from real examples.

Highnam et al. [54] used the h_{int} representation of mammograms to generate synthetic masses. In the h_{int} representation, the x-ray acquisition process (and its reverse) is modelled to estimate the proportion of interesting parenchymal tissue and fatty tissue depicted by each pixel in a mammogram. Partially synthesised mammograms were generated by converting a virtual three-dimensional mass into the h_{int} model, adding it to the h_{int} model of a real mammogram at a user designated position, and applying the x-ray acquisition model to generate the two-dimensional image. The virtual three-dimensional masses were either generated using geometric functions (although exactly how is not described) or by obtaining the h_{int} representation of a real mammogram containing a mass, and estimating the contribution made by the mass to the h_{int} values. Whilst images of synthetic lesions were shown, no quantitative evaluation of the

method was presented.

Skiadopoulos et al. [112] generated synthesised circumscribed lesions by transforming local grey levels in real mammographic backgrounds. The user selected the size, shape and position of the lesion, which was then generated by transforming the local background grey levels with respect to a set of basis functions. This approach has the advantage that the grey levels of the abnormality implicitly match the local background grey levels, overcoming a common problem when synthetic abnormalities are generated prior to being inserted into a real background. The results suggest the synthetic abnormalities generated were some of the most realistic in the literature. Six observers attempted to distinguish between real and synthetic lesions. ROC curves were fitted to the observer responses and a mean AUC of 0.55 ± 0.03 suggested synthetic lesions could not be identified at a rate significantly better than chance.

Ruschin et al. [99], simulated a range of mammographic abnormalities. Benign lesions, malignant lesions without spicules, malignant lesions with spicules and microcalcification clusters were all synthesised and inserted onto real mammographic backgrounds. Benign lesions were synthesised using the method described by Skiadopoulos et al. reviewed above. Malignant masses were synthesised using a random-walk algorithm to produce an initial irregular shape. Morphological erosion and dilation were applied to the shape to produce a continuous border. A linearly decreasing grey level profile was added to the modified shape to produce a blurred border, completing the synthetic mass. Stellate masses were synthesised by first generating a malignant mass using the method described above. Individual spicules were then created by generating a super-ellipse (that is, a shape described by the equation

61

$(x/A)^n + (y/B)^n = 1$, where $0 < n < 1$) of constant grey level before bisecting each ellipse along the shorter axis and discarding one half. As with masses, the remaining shapes were blurred to achieve a smooth drop-off at the border. A set of between 10 and 20 individual spicules were sampled, randomly oriented and centred on the centre of a simulated mass. Each abnormality was positioned in a normal mammogram and had its grey level contrast manually adjusted to match the appearance of local breast tissue. Visually realistic results were claimed, although no quantitative evaluation was used to justify this.

Saunders and Samei [102] generated synthetic benign, malignant, and spiculated malignant masses and inserted these into real mammogram regions. A set of 24 masses were extracted from the Digital Database for Screening Mammography (DDSM) [53] and used to build a model of mass appearance. Models of mass shape and intensity were based around an ellipse fitted to each mass outline. For each mass, the average grey level along concentric elliptical profiles extending from the centre of the mass to the border was computed, producing a trace of average grey levels. A modified sigmoid curve was fitted to the trace of each mass, to obtain a finite parameterisation. The parameters for each mass within a particular class (benign or malignant) were then averaged across the training set to form a model of mass intensity. To complement the texture model, shape information was extracted for each mass by measuring the deviation of the border from the fitted ellipse. The normalised power spectrum was then computed for each deviation profile, and the power spectra averaged for both benign and malignant masses. For spiculated masses, spicule appearance was modelled as a function of intensity versus length by computing average grey level profiles along a

straight line fitted to the length of each spicule, and then linearly regressing intensity against length. To simulate a new benign or malignant mass, a random sample of ellipse parameters (that is, major axis, axis ratio and orientation) were selected from those in the training set to generate a new ellipse. A new border deviation profile was generated by randomly sampling Gaussian noise, transforming in to the Fourier domain, filtering by the appropriate benign or malignant power spectra before transforming back to the spatial domain and scaling to match the ellipse. The deviation profile was used to modify the shape of the sampled ellipse into a mass border. For spiculated masses, triangles were then added to the mass shape although details on how the triangles were selected and positioned are not given. Finally, the appropriate benign or malignant texture model was used to generate an array of grey levels that could be superimposed on a real normal mammogram. The realism of both benign and malignant synthetic masses was evaluated in an observer study involving a single mammography reader. In a comparison against real masses, the AUC of a ROC curve fitted to the reader's responses was 0.59±0.08 for benign lesions and 0.61±0.07 for malignant masses. This suggests the reader was able to identify malignant masses at a rate significantly better than chance.

In a further paper, Saunders et al. reported the results of an extended study [101]. In the latter study, the models of benign and malignant mass appearance were constructed from an expanded set of 152 masses drawn from the DDSM. The models were constructed in the same manner as described in the earlier paper, however modelling and synthesis of spiculations were not described in the latter paper and we presume these were omitted from the study. The latter study included three mammography readers, for which average AUC values of 0.68±0.07 and 0.65±0.07 were reported for

benign and malignant masses respectively. The latter study also included a method of synthesising microcalcification clusters, for which an average of AUC of 0.62 ± 0.07 was reported.

Caulkin [23] generated synthetic masses and spiculated masses that were superimposed on real normal mammograms. A set of real mammograms, each containing a mass, was used as training data. The boundary of each lesion, any associated spicules, the pectoral muscle and breast skin-air boundary were annotated by a radiologist in each mammogram. To obtain mass-only grey levels, the local tissue background was subtracted from each lesion by estimating what the grey level of pixels in the mass region would have been if the mass were not present. A principal component based appearance model (see statistical models of shape and appearance, described in section 4.2) was fitted to describe the variation in shape, texture and size of the masses. Similarly, a principal component based statistical model was fitted to describe the shape, width, brightness and length of spicules in the training data. Because both the mass and spicule models were constructed as generative statistical models, new examples could be created by randomly sampling from the model distributions. A probability distribution was constructed to model the position of lesions within the breast. However, whilst a method for positioning synthetic lesions was proposed, this was not used to generate the set of synthetic masses that were quantitatively evaluated. In the evaluation, 15 breast radiologists attempted to distinguish between real and synthetic masses. A mean AUC value of 0.69 ± 0.13 was reported, suggesting the readers were able to identify synthetic masses at a rate significantly better than chance.

Other earlier work on synthesising mammographic abnormalities includes Lefebvre et

al. [72], who described a similar, but less advanced, simulation model of microcalcifications to the work of Lado et al. Claridge and Richter [26] described a simple method for synthesising mammographic masses. Finally, Smith [113] generated shape models to synthesise masses in a method similar to the initial steps of Caulkin's method.

Synthesising other abnormalities

Other than the methods for synthesising microcalcifications and mass lesions described in the preceding two sections, we note that to the best of our knowledge, there has been no previously published work in which other signs of breast disease such as architectural distortion or global asymmetry have been synthesised.

3.2.1   Discussion

In evaluating the work the described above, we focus on how the approaches taken to synthesise mammographic abnormalities could be adapted to meet the criteria stated in section 2.2.7. This requires that the abnormalities not only are realistic, but that it will be possible to automate the generation of new synthetic abnormalities, without significant user input.

A common limitation in many of the methods described above was the manner in which the appearance of a synthetically generated abnormality was controlled by manual adjustment of parameters [17, 54, 61, 84, 99, 112]. We have two problems with such methods: firstly because synthetic appearance parameters are manually adjusted, the extent to which a new abnormality is representative of the real population depends on the subjective judgement of the user implementing the method. Thus without re-evaluating every set of abnormalities generated, we cannot be sure such abnormalities

can be used in place of real data. Secondly, because manual input is required, the methods cannot be automated to produce large sets of synthetic abnormalities. Thus the full potential of using synthetic data cannot be realised.

Exceptions to methods requiring a high level of manual input were Lado et al.'s method for synthesising microcalcification clusters [70] and both Saunders et al. [101, 102] and Caulkin's [23] methods for synthesising breast masses. In each case, an attempt was made to parameterise the appearance of a training set of real data, and subsequently generate new lesions based on the learnt parameterisations. In the case of Lado's method, synthesis was limited to randomly sampling parameters from the training data and so only a finite number of unique clusters could be generated (although the final mammographic appearance would differ if the same cluster were superimposed on different normal backgrounds). Meanwhile both Samei et al. and Caulkin used a more general method in which a model of appearance was learnt so that vectors sampled from a random Gaussian distribution could be modified to reconstruct unique synthetic masses.

Abnormalities generated by all three methods were evaluated quantitatively. The observer study implemented by Lado indicated that the synthetically generated microcalcification clusters were realistic enough to be mistaken for real data by mammography readers. However, neither of the methods for synthesising malignant breast masses were able to produce synthetic lesions of sufficient realism. Thus despite the fact that mass lesions are one of the most common signs of breast cancer in mammograms, there has yet to be a method developed that both permits automatic mass generation and produces lesions of sufficient quality to be used in place of real data.

Given the clinical motivation for generating synthetic mammographic abnormalities, as discussed in section 2.2.7, this would suggest there is a particular need for developing a new method that can successfully simulate the appearance of malignant breast masses in mammograms.

To determine how we might develop a new method for generating synthetic malignant (and particularly spiculated) mammographic masses, we first note that methods in which a synthetic mass is generated and added to a real normal mammogram [23, 101, 102, 112] produced far more realistic results than those methods that attempted to synthesise whole mammograms [17, 84]. Thus at the broadest level, we intend to modify the appearance of an existing mammogram to produce a final synthetically generated breast mass, rather than attempting to synthesise whole mammograms. To obtain further details on how we might proceed, we present a more detailed comparison of the methods for synthesising malignant mammographic masses described by Caulkin [23], and Saunders et al. [101, 102].

First we consider the way in which the methods attempted to parameterise the appearance of a central mass. Of the two methods, we believe the principal component model used by Caulkin offers a richer set of descriptors to describe the variation of a global population of masses than the parameterised intensity and deviation model used by Saunders et al.. In particular we note that:

i) In Saunders et al.'s model the grey level texture of masses was described by just five parameters averaged across the set, whilst Caulkin's model included 37 modes of texture variation. This suggests a greater variation in mass texture can be described by Caulkin's model.

ii) Mass appearance in the model used by Saunders et al. is strongly constrained to the initially fitted ellipse. Both intensity change (texture) and shape variation were considered with respect to the ellipse and, as discussed by the authors, the model was not well suited to describe the appearance of masses whose shapes were non-elliptical. The model used by Caulkin does not enforce such constraints. Instead the variation allowed in the model is learnt from the way in which a real set of data varies in appearance. Providing the training data comprises sufficient variation to represent a global population, the model should be flexible enough to learn these changes without requiring that all masses conscribe to a particular shape.

iii) In Caulkin's model, the step in which the parameters of shape, texture and size are combined into a complete appearance model allows for a correlation between these components to be learnt. Such correlations may be an important part of real data – for example, we might expect that larger (and therefore more developed) masses have both a greater intensity (that is, higher grey levels) than smaller masses, and have developed a more irregular shape, than smaller, newer masses. In the synthesis method described by Saunders the size, intensity and shape variation were all sampled independently of one another, and thus such correlations cannot be captured in the model.

For these reasons, we think that the model described by Caulkin provides a good base for describing the variation in mammographic appearance of the central density in a malignant breast mass. As a result, in the next chapter we take a broader look at statistical appearance models, and in particular the principal component models used in the active appearance model framework. Later, in chapter 7, we propose a method for

optimising the fit of an appearance model so that it generalises to describe new masses with significantly lower error than the model fitting method used by Caulkin.

Aside from describing the appearance of the central density in a mass, we now consider how Caulkin and Saunders et al. synthesised mass spicules. A direct comparison of the methods is difficult because very few details on how spicules were modelled and synthesised were provided by Saunders (indeed, as noted above, a method for simulating spicules appeared to be omitted in an extended version of the initial paper [101]). However, when discussing his work, Caulkin noted that the appearance of spicules was the least convincing aspect of the synthetic masses generated by his method. In fact, we don't believe that a method in which individual spicules are modelled separately prior to being added to a central density will ever be sufficient to capture the true appearance of a real stellate mass. In particular we propose that such a method:

i)   Cannot adequately describe the appearance of clusters of very fine spicules (see the mammographic appearance of sheaf-of-wheat spicules described in section 2.2.4).

ii)   Does not account for the interactions between the central density, spicules and the surrounding breast tissue.

These limitations (and particularly point (ii)) are discussed further in section 8.2. In addition, we note that both limitations were raised as areas in which future work would be necessary by Saunders et al. in a discussion of their work, although further details on how these issues could be addressed were not given [101]. Because the literature on

synthesising abnormalities contains few details on how patterns of breast tissue are affected by the presence of a malignant mass, we review a wider range of mammography in which this subject is considered later in this chapter (see section 3.3.1).

We also note that whilst we do not specifically intend to synthesise architectural distortions in which a central mass is not present, by considering the way in which a spiculated mass affects local patterns of breast tissue we hope our work will go some way towards showing how patterns of architectural distortion could be simulated. As noted above, we are not aware of any previous methods that have attempted to synthesise such abnormalities.

Finally, we observe that none of the methods reviewed used a method for automatically selecting a location for synthesised lesions in a target mammogram. Whilst Caulkin [23] described a method for modelling the locations of masses in a set of training data with respect to a global breast shape, this method was not extended to permit the realistic sampling of new locations in a normal mammogram (this is discussed further in section 10.1). Therefore a method that allows synthetically generated lesions (regardless of lesion type) to be automatically positioned would be of benefit - even to methods that have already successfully simulated abnormal appearance given a manually selected location (such as the method described by Lado for synthesising microcalcifications). We note that to automatically locate a synthetic mass in a whole mammogram, it will be necessary to have a method for automatically determining the main breast area in mammograms. Thus methods of segmenting the breast (and pectoral muscle in MLO mammograms) are reviewed in section 3.3.2.

3.2.2   Summary

In section 3.2, we have presented a detailed review of existing methods for synthesising the appearance of abnormalities in mammograms. As a result of this review we have concluded that there is a specific need for developing a method in which realistic malignant, spiculated breast masses are synthetically generated. In particular we suggest that any new method must consider the way in which a mass affects the appearance of local breast tissue. We also propose that a method for automatically selecting a location within a whole normal mammogram should be developed. In the following section, we review work from a broader spectrum of mammography research than that specifically looking at lesion synthesis. For each of the topics listed, we are particularly interested in how the work reviewed may be of benefit to a synthesis method that overcomes the limitations described above.

3.3 Related areas of mammography research

In this section we review a selection of work in which techniques from the field of image processing and analysis have been applied to mammograms. The reviews are structured to match the two key areas in which we intend to develop a method for synthesising malignant mammographic masses, as discussed in the previous section. Thus in section 3.3.1 we consider various methods in which the mammographic appearance of breast tissue is detected, modelled and classified. Meanwhile in section 3.3.2, we review methods for automatically segmenting the main breast shape in mammograms.

3.3.1   Detecting and modelling breast tissue structures

In this section we review work in which breast tissue structures (and the patterns they

form) are detected, modelled and classified. Because the appearance of spicules may closely match the appearance of some normal tissue structures in the breast, we include the detection and classification of spicules in the appropriate sub-sections below.

Detecting linear structures

Because mammograms are formed by the projection of a complex network of ducts in the breast, they are usually rich in linear structures at a variety of scales and orientations. Thus a natural first task when working with the appearance of breast tissue in mammograms is to explicitly detect linear structures. We note that because many of the structures in the breast do not form straight lines, they are commonly referred to as curvi-linear structures (CLS). A review of work detecting such structures in mammograms is given below.

Early work in this area was undertaken by Cerneaz and Brady [24]. By considering the physical characteristics of CLS in addition to a model of both film acquisition and film digitisation, a kernel width was selected from which to estimate the second derivative of the image surface in a mammogram. Strong features within the derivative were selected as candidate CLS pixels. A series of morphological operators was applied to the binary map of candidate pixels (thinning, separating, reconnecting segments lying on the same line) to produce a final set of CLS pixels.

Another early CLS detection method was described by Parr et al. [87]. In this method, the line operator originally developed by Dixon and Taylor [36], was applied across multiple scales to obtain a measure of line strength, orientation and scale at each pixel in a mammogram. Non-maximal suppression was applied to obtain a final set of CLS

pixels.

A further method using multiscale directional second order Gaussian derivatives was developed by Karssemeijer and te Brake [63]. Again, the maximum filter responses across scale were used to compute the line strength, orientation and scale at each pixel in a mammogram. Pixels of negative contrast or low strength were discarded. Additionally the image gradient was estimated using a first derivative Gaussian filter, and any coarse scale pixels with an orientation closely aligned to the gradient normal were rejected on the basis that the magnitude response at such pixels was likely to be caused by a ripple due to a strong image gradient in the vicinity.

All of the above methods, and two further line detection methods (using a variant of the line operator comprising radial, not rectangular, bins; and applying directional morphological opening) were compared by Zwiggelaar et al. [129]. Each method was used to detect synthetic lines of varying contrast and profile that had been added to a selection of real mammographic backgrounds. It was found that the line operator was most successful, although it should be noted the comparison was essentially over the operators used to measure an initial line strength, orientation and scale at each pixel and not over subsequent processing steps.

We now review three more recent CLS detection algorithms. In each case, local image phase is calculated and use to infer structural information in the image. The technical details of this concept are discussed in more detail in the following chapter, in section 4.3.1.

Schenk and Brady [103] use a set of steerable complex filters [47] at multiple scales to compute local energy, orientation, phase and an approximate width for each pixel in a mammogram. A set of candidate pixels are selected by hysteresis thresholding energy before using local phase to remove all pixels apart from those with structural shape corresponding to an edge or ridge (that is, non-pure features and valleys are ignored). Within this candidate set, CLS are determined by searching for parallel sets of ridge pixels bordered by step-up and step-down edges. The search is performed by looking for suitable edge pixels in the neighbourhood of each candidate ridge pixel, given the ridge pixel's approximate width. The exact width of a CLS at each point can then be determined by the normal distance between the two edges. Finally the set of CLS is skeletonised, so that in a given region, the number of intersecting linear structures can be computed. This information is used to improve the specificity of a calcification detector by rejecting high responses that were most likely caused by the intersection of two or more CLS in the immediate vicinity of the response.

Contemporaneously to Schenk and Brady's method, McLoughlin et al. [78] produced a similar CLS detection scheme to reduce the false positives of a microcalcification detector. In their algorithm, a mammogram is convolved with odd/even pairs of log-Gabor filters at six orientations and five scales. At each pixel, a local energy for each orientation is defined by computing the vector sum of phase-magnitude pairs across scale. Linear features are expected to produce a magnitude peak in one direction only, whilst round features such as microcalcifications produce a strong response at all orientations. This characteristic is used to specify a line measure that enhances linear features and suppresses round features. The resulting line strength image is further processed using a relaxation labelling scheme, and the final CLS map used to reject

pixels from consideration in a subsequent attempt to detect microcalcifications.

Work by Wai et al. [121] built on the two methods described above. Again a measure based on the congruency of phase across scales was used to construct a local energy measure from which CLS could be detected. However, rather than filtering with multiple oriented complex filters at each scale, the monogenic signal scale space [42] was used to efficiently compute image amplitude, phase and orientation at each pixel. In all three of these methods, using local phase helped infer structural information and allowed features to be constructed that were more robust to changes in contrast and image noise.

Finally, we note that Rangayyan and Ayres [96] describe a method for detecting CLS in mammograms using a bank of 180 oriented Gabor filters of fixed scale and elongation. At each pixel the maximum filter response and the associated orientation are recorded. NMS is applied to the magnitudes to determine an initial set of candidate pixels. Finally, as in Karssemeijer's method described above, pixels closely aligned with the image gradient are rejected. This method appears somewhat more simplistic than the recent state-of-the-art methods described above (for example, it uses filters at only one scale). However, it was subsequently used to search for architectural distortion patterns (see below) and thus is included here for completeness.

Detecting abnormal patterns of linear structures

In this section we review work in which patterns of linear structures are modelled and classified to detect whether an abnormality may be present in a mammogram. In each of the three methods described, an explicit detection of CLS is initially performed using one of the algorithms reviewed in the previous subsection. These are the algorithms

attributed to Parr [87]; Karssemeijer and te Brake [63]; and Rangayyan [96] respectively. The remaining details of each method are given below.

Parr et al. [88] developed a method for detecting patterns of linear structures associated with a spiculated mass. The method uses a training set of mammograms each of which contains an annotated spiculated mass. For a given pixel, the local pattern of orientations is parameterised by extracting a 512x512 region about the pixel and concatenating the orientation associated with the maximum CLS response within each 32x32 block of the region. To build a statistical model of lesion patterns, an orientation pattern is extracted from pixels within the border of each mass in the training data. Factor analysis is applied to the resulting set of vectors in order to form a compact representation of the data. To detect suspicious patterns in unseen mammograms, orientation patterns are extracted in a regular grid across the breast region. By computing the factor scores of each pattern using the lesion model, the likelihood of the pattern being associated with a spiculated mass can be computed. Applying leave-one-out testing of the mammograms in the training data, Parr et al. reported a detection rate of approximately 80% sensitivity for a fixed specificity of 0.5 false positives per image. It was noted that lesions with a diameter as low as 1.7mm were detected. Further results for this method, in combination with an algorithm for detecting mass densities, were reported by Zwiggelaar et al. [130].

Where Parr et al.'s method aimed to learn a distribution of orientations from real training data, Karssemeijer and te Brake [63] describe an algorithm for detecting stellate distortions using an explicitly prescribed pattern of orientations. The aim is to detect regions of a mammogram in which an abnormal number of CLS are oriented toward

some focal point. For each pixel within the breast region of a mammogram, two features are defined that characterise the proportion of CLS pixels within radial segments of a local circular neighbourhood that are approximately oriented towards the centre. By analysing whether this proportion is statistically greater than the proportion that might be achieved by chance, the overall likelihood of having a stellate distortion centred at that pixel is computed. Karssemeijer and te Brake reported a detection sensitivity of approximately 90% with a corresponding specificity of approximately 1 false positive per image.

Rangayyan and Ayres [96] describe a method using phase portraits to determine regions of architectural distortion. Three classes of phase portraits are pre-determined: nodes, in which the orientation radiates away from a central point; spirals; and saddles. The orientation of CLS pixels within a sliding analysis window are used to compute the best matching phase portrait. The focus of the best matching portrait is used to increment a vote for that portrait class at the focus location. As a result a map of votes for each phase portrait is constructed throughout the mammogram. Peaks within the node map are used to determine sites of possible architectural distortion (although in early work peaks in all three portrait classes were considered [10]). However, whilst a detection sensitivity of 84% was obtained, this came at the price of 7.8 false positives per image. Clearly this is a significantly larger rate of false positives than either of the two methods described above, although it should be noted that in general architectural distortions are harder to detect than distortions explicitly associated with a mass.

Whilst we are not directly concerned with detecting abnormal patterns of breast tissue, the work in this subsection is interesting in that it provides information on the

characteristics we expect to find in the region surrounding malignant masses. Moreover, the method developed by Parr et al. [88], suggests that a compact representation of oriented structure patterns can be constructed for a training set of mammographic data. We propose that the framework of the model described by Parr et al. could be adapted to allow new patterns of orientation to be sampled. This may provide a template from which to base the distortion of breast tissue in mass regions.

Classifying and modelling general mammogram appearance

We conclude our review of work related to breast tissue appearance by considering methods in which the general appearance of mammograms are modelled and classified without assuming the importance of oriented linear structures in the breast.

Karssemeijer [62] presents a method for characterising the distribution of parenchymal patterns in whole mammograms. This is achieved by computing histograms of pixel grey levels in concentric regions moving from the skin-air boundary of the breast to the centre of the chest wall. For each region, the standard deviation and skewness of the histogram are computed and used to form a compact representation of tissue distribution in the breast. By computing this representation for a training set of mammograms grouped by density, a classification scheme can be employed to determine the density class of unseen mammograms. The representation of mammographic appearance described is too coarse for our needs, however the method provides useful context for the work of Petroudi [92] and Rose [98] described below.

Petroudi et al. [92] also describe a method for classifying mammograms by density given a statistical analysis of their texture. Each mammogram in a training set of data is initially filtered using the MR8 filter back [119]. The MR8 bank comprises 38 filters: 6

oriented edge and 6 oriented bar filters across 3 scales, a Gaussian filter and a Laplacian of Gaussian filter. The maximum filter response across orientation is recorded for the bar and edge filters at each scale, and thus a vector of 8 responses is stored at each pixel. The training mammograms are split by their BIRADS density class, and the set of all MR8 pixel responses aggregated within each class. For each set of data, k-means clustering [76] is used to learn 10 cluster centres (termed textons [73]). Every pixel within each training mammograms is replaced by its closest matching texton (regardless of what class the texton was learned for), and a histogram of textons computed for each mammogram. The set of histograms is used as a model of texture within each density class. Thus for an unseen mammogram, the MR8 filter is applied and used to compute a texton histogram. The closest matching histogram of those learned from the training data is computed, and density class associated with the closest matching histogram assigned to the unseen mammogram.

In this method, the texton histogram effectively parameterises the appearance of each mammogram in a compact vector, and the training set of all such vectors provide a model of whole mammographic appearance. This model is not generative. Whilst statistics of the texton vectors could be learned across the training data and used to sample new instantiations, appearance cannot then be reconstructed from a texton vector. In computing a histogram the spatial arrangement of textons is lost. Moreover, because the MR8 filtering scheme is not invertible, even with a spatial map of textons, full mammographic appearance cannot be recovered. Of course this is not a problem in the task the model is designed for. The results reported by Petroudi indicate 91% and 94% of low density and high density mammograms were correctly classified. This indicates that the model provides a successful analysis tool for classifying density,

suggesting the texton vectors provide a good discriminative model of mammographic appearance. However, such a model cannot be used towards our goal of synthesising mammographic appearance.

For this reason, we now consider a model developed by Rose [98] that uses similar appearance characteristics, but in a generative framework that enables synthesis. Like Petroudi et al.'s method, mammograms in a training set are decomposed across scale and orientation. However, an invertible decomposition, the steerable pyramid [109], is used. A principal component based appearance model (see statistical models of appearance in section 4.2 of chapter 4) is fitted to the data to describe the combined variation in breast shape and the distribution of coarse scale pyramid coefficients. This model encapsulates an approximate global appearance of mammograms. To describe detailed tissue appearance, at each pixel a vector comprising the pyramid coefficients across scale and orientation is extracted. The distribution of all such vectors sampled from the training data is modelled using a Gaussian mixture model (fitted using k-means clustering or an expectation maximisation algorithm). A training set of normal mammograms was used to build a model of normal appearance. The model can also be used to synthesise a new mammograms as follows: first the global model distribution is randomly sampled to generate a breast shape and a set of coarse scale pyramid coefficients. Next, a set of detailing coefficients can be sampled for each pixel by conditionally sampling the local model, given the coarse scale coefficients already sampled from the global model. Finally, the set of pyramid coefficients can be inverted to reconstruct a fully synthesised mammogram.

Whilst we do not intend to synthesise whole mammograms, this method has been

influential on our work in two ways. Firstly, it describes an explicit separation between coarse and fine scale mammographic appearance – later we seek to exploit this by maintaining the low-frequency components of a real normal mammogram whilst modifying the high-frequency structures. Secondly, it suggests that a generative model of decomposed coefficients can be used to reconstruct synthetic mammogram texture. Again this is something we will use when modelling the appearance of breast tissue in the region of a malignant mass. However, we conclude our review by noting limitations of the work. Firstly, when sampling the detailing coefficients, spatial correlations between coefficients within the same scale are lost (this is in contrast to the pixel based synthesis algorithm also developed by Rose [98] and discussed in section 4.4). Secondly, a real valued steerable pyramid was used. As noted in section 4.3.2, the coefficients of a real-valued decimating decomposition suffer from shift dependence. This issue, and the problems it may cause for model building are discussed further in sections 4.3.2 and 8.3.1.

Summary

Throughout section 3.3.1, we have reviewed work to contextualise and inform our attempt to synthesise the appearance of breast tissue and mass spicules in the region around a central density. The key points we summarise from this review are:

i) Decomposing mammographic appearance across scale and orientation has proved vital in detecting linear structures in breast tissue. Moreover recent developments have shown an estimation of local phase provides added structural information and helps stabilise against noise [78, 103, 121].

ii) A statistical model of abnormal orientation patterns in mammograms has been

developed [88]. Whilst not strictly generative, the model uses a framework that may be adapted to sample new abnormal orientation patterns.

iii) A generative model of normal mammographic appearance has been proposed [98] using the statistical properties of coefficients localised in scale and orientation.

In the next section, we turn to methods for segmenting the breast shape and other global structures in a mammogram.

3.3.2   Breast segmentation algorithms

As noted in section 3.2.1, one aim of our project is to develop a method for automatically selecting a location at which to synthesise a breast lesion in a normal mammogram. To achieve this, we require a method for automatically segmenting the breast so that a set of possible locations can be determined.

Of course, segmenting the breast region in a mammogram is a natural first step in a host of other tasks and thus numerous methods have been proposed in the literature, as reviewed in the following section.

Segmenting the breast

An early method for segmenting the breast in mammograms was developed by Bick et al. [16]. In this method noise is removed from a mammogram by median filtering. The local range of each pixel (that is, the difference between its grey level and the minimum grey level is some local region) is calculated, and together with a global histogram of grey levels, used to determine pixels belonging to the breast. Such pixels are defined as having a local range above some threshold, whilst lying between the peaks in the histogram associated with background and off-film pixels. The region of breast pixels is

modified by morphological connectivity operators to remove small artefacts, until a contiguous region is formed. The outer contour of this region is calculated and used as the final breast-air border.

Chandrasekhar and Attikiouzel [25] describe a method for segmenting the breast region in a mammogram in which the cumulative grey level histogram of the mammogram is used to compute an initial estimate of the background region. A measure based on the local variance of the mammogram background is used to characterise the image, and direct a sequence of steps that refine the initial border. The final breast border is computed from the set union of binary images generated by fitting polynomial surfaces of varying degrees to thresholded regions of the original mammograms.

Ojala et al. [85] described a method for obtaining an initial segmentation of the breast using a process similar to that described by Bick et al. (that is, thresholding based on histogram analysis followed by morphological operators to obtain a contiguous breast region). In addition, three further methods were compared for refining the initial segmentation given a set of control points spaced along the breast-air contour. In the first, a Fourier transform is computed and used to remove high-frequency components from the boundary. In the second, an active contour algorithm [64] is applied to determine the final border. Finally the third method uses a B-spline interpolant [117] to calculate a smooth border between control points. Of these methods the active contour method was deemed most successful.

Ferrari et al. also use an approach in which an initial segmentation is refined using an active contour model. However, additional constraints were placed on the active contour to improve the fitting process [44]. A similar approach was also employed by Wirth and

Stapinska [126].

Finally, an alternative method was developed by Pan et al. [86] in which the monogenic signal (see section 4.3.1) is used to infer orientation and structure information across multiple scales in a mammogram. Using thresholds based on complex phase and orientation at a coarse scale, an initial breast region is formed. The final skin-air boundary is determined using a local amplitude based measure similar to that used by Wai et al. [121] when detecting CLS (see section 3.3.1). The method is designed to remove the dependence on the grey level histogram used in the segmentation algorithm methods described above. As a result the authors propose it should be more robust to changes in intensity range or contrast, and thus may be more flexible when segmenting the breast in mammograms acquired from different imaging systems.

Segmenting the pectoral muscle

In addition to defining the main breast border at the skin-air boundary, it is often necessary to delineate the edge of the pectoral muscle in MLO mammograms. As discussed in section 2.2.3, the pectoral muscle generally appears as an approximately triangular region in the upper corner of the chest wall side of the MLO view. In a mammogram the pectoral muscle usually has a higher grey level than tissue in the breast and therefore may confound methods that analyse the intensity of tissue. In addition, the position of the pectoral muscle provides a useful anatomical reference point in MLO mammograms.

A simple approach for fitting a straight line to the pectoral muscle is described by Karssemeijer et al. [62]. In this method, a region in the upper corner of a mammogram in which the muscle is located is extracted, and filtered using a Sobel edge detector to

enhance the edge of the muscle. A Hough transform [37] is then computed to detect the most suitable straight line, given constraints on the angle of the line. This initial approach is adapted by both Kwok et al. [69] and Ferrari et al. [45] to obtain a more precise delineation of the pectoral muscle that was not confined to a straight line.

3.3.3   Segmenting the fibroglandular disc

Having segmented the main breast region, it may be useful to extract more specific regions within the breast, such as the fibroglandular disc. For example, the shape of the fibroglandular disc between left and right breasts may be used as an asymmetry measure, examined for local distortions or selected as a region within which malignant masses are most likely to be found. With regards to a synthesis method, the fibroglandualr disc could be used to influence the location of synthetic masses in a target normal mammogram. However, as we discuss in section 10.3.4, this is not an option we pursue further. For completeness, we note that attempts to segment the fibroglandular disc include work by Byng [21], Ferrari et al. [43], Saha et al. [100] and Zhou et al. [128]. For further details on this work the reader is directed to the respective papers.

3.4 Summary

In this chapter we have reviewed a selection of work in which techniques from the field of image analysis and processing have been applied to mammograms. In section 3.1, we presented a brief history of this general field of work and discussed the main top level tasks that researchers have attempted to solve. Such tasks include the automated detection and diagnosis of mammographic abnormalities, and the quantification of dense tissue within the breast.

This general synopsis was followed with a detailed review of existing methods for synthesising breast abnormalities in mammograms. This review was intended to show the current state-of-the-art in the literature and highlight the areas in which we believe further work would be beneficial. In particular, we noted that there has yet to be a method in which malignant, spiculated masses have been synthesised to a standard that they could be used in place of real data when training or assessing mammogram readers. As a result, we identified the development of such a method to be the key goal of our thesis.

More specifically, we proposed two areas in which further work was required: firstly, a new method must take into account the way in which a mass distorts surrounding tissue; secondly, the new approach should include a method for placing a synthesised mass within a normal mammogram without requiring a manual selection of the location. Given these goals, we identified various other areas of mammography research that would be relevant to our project.

In the next chapter, we continue our review of background literature. However, rather than focusing our search on methods that have previously been applied to mammograms, we consider more general techniques from the field of image processing and analysis.

Chapter 4 : Image Analysis and Processing Techniques

4.1 Introduction

In this chapter we present a review of various methods and techniques from the general field of image analysis and image processing. This review serves two purposes: firstly, it provides further context for much of the computer-aided mammography research described in the previous chapter. Secondly, the review provides technical details on specific methods and tools that are used and developed in later chapters of this thesis.

We concentrate on three areas of research: statistical models of shape and appearance (and in particular, the models used in the active appearance model framework developed by Cootes et al.); wavelets and other multiscale transforms of images; and texture synthesis algorithms.

4.2 Statistical models of shape and appearance

In section 3.2, we described the method used by Caulkin [23] to model the appearance of the central density in a training set of mammographic masses. In this method, a principal component based statistical appearance model was used to describe the variation in shape, texture and scale of the training data. In a comparison with existing mass synthesis methods, we proposed that this was a sensible base from which to develop our own synthesis method. Thus in this section, we give a general review of the field of research that led to the development of the statistical model used by Caulkin, and consider the current work within this area.

Statistical models provide an efficient representation of the variability in shape and appearance of objects in image data. For example, assume we are working with some

class of object (say a human face, an organ in a medical image, or, as in our work, a mammographic mass) and have a way of parameterising the appearance of each instance of the object in an image. If we have training data containing many instances of the object, we can construct a statistical model that encodes the probability with which a particular parameterisation is likely to occur. Thus the model describes the variation in appearance of the object across the set of training data. Moreover, if the training data contains sufficient variation to be representative of a global population of the object, then a good statistical model will provide both a general and specific representation of object appearance. That is, it will be *general* in that the parameterisation of any valid instance of the object will be deemed as likely under the model distributions. It will be *specific* in that parameterisations that are likely under the model distributions only describe the appearance of valid instances of the object. Thus a model with good specificity will not represent the appearance of image objects of a different class.

Further, if the appearance of an object can be reconstructed from its parameterisation, then the model fitted to the data is generative. Generative models have several advantages. For example, as used in this thesis, they can generate new unique instances of an object that can be used as training data for other applications. Generative models can also be used to search for instances of an object in unseen images. Indeed this is the area which has seen most development of statistical appearance models, and a review of this work is given below.

4.2.1 Models of shape

We start by focusing on statistical models of object shape. Early work in this field included a method by Goodall [50] in which a set of corresponding landmark points was

used to describe the shape of each instance of an object in a training set of data. The shapes were then rotated, scaled and translated to minimise the sum of squares distances between corresponding points on different shapes. This process, known as Procrustes Analysis [50], aligned the shapes so that mean shape and covariances between landmark points could be computed. These summary statistics were then used to assess the differences between sets of shapes.

Cootes et al. [29] expand this idea by applying principal component analysis (PCA) to the deviation of each shape from the mean shape. In appendix A.1, a detailed description of principal component analysis applied to any arbitrary dataset is given. Here we note that as a result of applying PCA, a compact representation is established in which each shape in the training data is parameterised by its coordinate along a finite number of principal modes of variation (where the number of modes is typically much smaller than the number of landmark points originally used to describe each shape). By construction, the parameters along each mode are linearly independent. By sampling new parameterisations subject to the variance associated with each mode, valid instances of an object's shape can be reconstructed from the model. Such models are known as point distribution models (PDM).

 As a further development to PDMs, active shape models (ASM) were developed by Cootes et al. [30] so that instances of an object shape can be located in a new image. The idea is that given a set of training data and a PDM describing the shape of objects in the data, the local grey level information at each landmark point can be used to determine the optimal instantiation of the PDM in an unseen image.

This is achieved by extracting grey level gradient profiles normal to the boundary of

each landmark point for all images in the training data. PCA is then applied to compute a separate statistical model of the variation in grey level gradients for each landmark. Thus given a new image and some initialisation of landmark points, an iterative search can be performed in which the positions of individual points are optimised subject to the constraint that the set of all landmarks must be a valid instance of shape under the PDM. The optimal position of an individual point is defined as the position at which the grey level gradient profile extracted normal to its current position minimises the Mahalanobis distance to the appropriate model mean.

ASMs have been shown to be useful in segmenting object shape in numerous applications. However, whilst ASMs use local grey level information when searching for an object, the full appearance of an object – that is, the distribution of pixel intensities throughout its shape – cannot be reconstructed from the model. Below we describe how shape models can be extended to include a complete description of object appearance.

4.2.2   Models of combined appearance

To develop a statistical model that includes a more complete representation of an object's appearance, Lanitis et al. [71] extract a texture vector from each object to describe its full appearance. To ensure that there is a suitable correspondence between elements of the texture vectors across different objects, the texture vectors are extracted from shape-normalised copies of each object. That is, having constructed a PDM to learn object shape, each image is warped so that the landmarks within that image assume the mean shape of the PDM. PCA can then be applied to the set of shape-normalised texture vectors to build a compact statistical model of texture. As a result,

any object in the training data can be described by its parameterisation in both the shape and texture models.

However, Edwards et al. [38] note that there may be further correlations between the shape and texture parameterisations of an object. Therefore if shape and texture are treated independently, combinations of shape and texture could arise that are not parameterisations of valid instances of the object. In this case, the model would lack specificity. To overcome this, Edwards et al. apply an additional level of PCA to concatenated vectors of shape and texture model parameters. The principal modes in the resulting model describe the combined variation in shape and grey level texture of objects across the training data. By sampling parameters from the combined model, vectors of shape and texture parameters can be reconstructed which in turn can be used to reconstruct the image object. Thus the combined appearance model provides a fully generative, compact representation of the data.

Cootes et al. [27] further refine the process of building a combined appearance model, and, in the active appearance model (AAM) framework, develop an efficient method for using combined appearance models to perform segmentation. The key to the AAM algorithm is learning *a priori* how changes to model parameters affect the extent to which the synthetic object, as reconstructed from the model, matches the real object in an image. This relationship is learnt using the training data, so that when attempting to fit the model in a new image, by considering the residual error between the image and the current model reconstruction it is possible to choose the best way of adjusting the model parameters during each iteration. Because a consistent relationship between changes in model parameters and model fit can be learnt only for small changes, AAM

91

searches require a good initialisation of landmark points to be successful. To improve the capture range of the model, statistical appearance models can be learnt for multiple resolutions of images in the training data. The models are then fitted using an AAM search from coarse to fine resolution, with the best fit at each resolution projected up to the next finer scale.

The AAM algorithm has been used successfully to segment a wide variety of image data [12, 27, 83]. As noted above, however, in our project we are more interested in the manner in which a combined statistical appearance model describes a set of training data than how the search algorithm is constructed. With regard to this, we note that one of the issues addressed by Cootes et al. [27] is how the individual texture and shape vectors should be scaled to account for the difference in metrics between the two components. In the earlier work by Edwards et al. [38], it was proposed that shape and texture were scaled so that the overall variance contributed by the two components was equal. However, Cootes uses a more sophisticated method in which the magnitude of the change caused to the shape-normalised texture vectors by varying parameters in the shape model is calculated. This allows weights to be selected so that a unit change along each shape mode causes a unit change in the resulting texture vector (by construction, a unit change along any texture mode also causes a unit change in texture). As we show in chapter 7, this issue is crucial in fitting a combined statistical model of appearance to mammographic masses.

4.2.3   Further work on statistical appearance models

In the previous section we described the basic construction of a statistical appearance model that encapsulates the combined variation in shape and texture of a training set of

image objects. Below we consider some further extensions of this work.

Automatically selecting correspondence points

The first step in building PDMs (and consequently full combined appearance models) is to place a set of corresponding landmark points on each image object in the training data. Doing this manually can be very time consuming and is subject to human error and interpretation. The problem becomes even harder when marking points in three-dimensional volumes (see extension of model into 3-D below). As a result there has been significant research aimed at automating the process of obtaining correspondences between images.

Given a set of continuous shape boundaries of a particular object (obtained for example, by automatically segmenting an object across a set of training images), Davies et al. [34], developed a method for obtaining optimal correspondences between the shapes. This is achieved by constructing an objective function termed the minimum description length (MDL). The idea of the MDL function is that it quantifies the amount of data that would need to be transmitted to fully describe the training set given some set of parametric statistical models. The better a model fits a set of training data, the less information that needs to be transmitted to encode the data. However, the cost of transmitting the model modes is also included in the function. Thus a model that minimises the MDL of a set of data provides the best compromise between model complexity and model fit. In the formulation by Davies et al., a set of landmark points placed on the boundary of each shape are iteratively refined until the MDL of the shape data and the PDM built from the current set of landmarks is minimised. The final positions of the sample points on each shape are used as optimal landmark points from

which to build further statistical models.

This idea was extended by Cootes et al. [31] to determine optimal correspondences in a set of images, without requiring a predetermined set of shape boundaries. In this method a set of initial sample points is placed on a grid in each image across a training set of data. The sample points are iteratively refined using piecewise affine deformations until the MDL of the image data and a full combined appearance model (built using correspondence defined by the sample points) is minimised. The method effectively simultaneously registers and models the images. This idea is further extended to include an explicit segmentation of parts of the image in work by Petrovic et al. [93].

Modelling non-linear variation

In the statistical models described in sections 4.2.1 and 4.2.2, it is assumed that parameters along separate modes can be varied independently of one another. PCA removes linear dependencies from a set of input data, and thus if the data contain no non-linear correlations, the parameters in the transformed principal component model space will be fully independent. Thus the distribution of the model parameters can be described by a single multivariate Gaussian distribution with zero mean and diagonal covariance matrix.

However, if non-linearities exist in the input data, then a Gaussian distribution may not be sufficient to describe the distribution of parameters in the model space. As a result, parameterisations deemed valid under the assumption of a Gaussian distribution may in fact reconstruct non-valid instances of the object. That is, the model would lack specificity. In a search framework such as the AAM, noise in the data could lead to illegal parameter combinations being selected, thus reducing the robustness of the

algorithm. Moreover, when generating synthetic data, as in Caulkin's work [23] and as we intend to do, then illegal instances of an object could be created.

Various attempts have been made to allow for non-linear variations in a statistical appearance model. These include the use of Gaussian mixture models by Cootes and Taylor [28]; the use of kernel PCA by Romdhani et al. [97]; and, more recently, the use of Gaussian process latent variable models by Huang et al. [59].

Modelling three dimensional image data

We conclude this section by noting that although the descriptions of statistical models discussed thus far refer to modelling two-dimensional image data, the models can be extended into three dimensions. Mitchell et al. [83] described the first fully complete adaption of the AAM using three-dimensional image data. However, throughout the remainder of this thesis, only two-dimensional image data are considered.

4.3 Wavelets and other multiscale transforms

In section 3.3, we reviewed work in which the appearance of breast tissue structures in mammograms were detected, classified and modelled. A common theme through many of the methods was the use of some form of multiscale filtering to obtain structural information localised in scale and orientation [63, 78, 103, 121, 129, 130]. Because mammograms are rich in linear structures of different sizes and orientations this kind of approach makes sense.

A large number of multiscale image transforms have been developed in the field of image (and signal) analysis and processing. Indeed, as we saw in section 3.3, considering just a subset of applications in mammography research we noted the use of

the MR8 filter bank [92]; Gabor filters and log-Gabor filters [78, 96, 103]; the monogenic signal [86, 121]; the multiscale line operator [88, 130] and the steerable pyramid [98]. However the goal of these transforms is essentially the same. That is, given a location in an image and some set of frequency scales, we want to know whether a structure is present at the location and if so, at what scale and orientation. If the decomposition allows it, information on the type of structure present (for example an edge, a ridge etc.) may also be beneficial.

In the following sections we describe three multiscale transforms: the monogenic signal; the steerable pyramid; and the dual-tree complex wavelet. As noted above, the first two transforms have already been applied to mammograms. We propose that the dual-tree complex wavelet has great potential in analysing, modelling and synthesising breast tissue structures in mammograms.

4.3.1   The monogenic signal

In the field of signal processing, the analytic signal is constructed from a real-valued 1-D signal by combining the signal with its Hilbert transform. In the frequency domain, this has the effect of removing negative frequency components, leaving only positive frequencies. As a result, in the time domain the analytic signal represents the original signal as a phase-magnitude pair at each point. This representation provides more information than the real values contained in the original signal alone. The magnitude encodes the energy (amplitude) of the signal at any point, whilst the phase contains local structure information.

Moreover, if the analytic signal is combined with a bandpass filter to separate the signal across frequency scales, then phase and amplitude measures can be computed at each

scale for a given time point. Given such a decomposition, the idea of phase-congruency was proposed as a feature detector by Kovesi [68]. At a particular location in a signal, phase congruency measures the extent to which phase across scales match. This is based on the idea that features such as pure steps or impulses are self-similar across scale, and thus have a constant phase response.

If we now consider an image, and take the projection of the image at some fixed orientation, we obtain a 1-D signal to which the concept of phase, and consequently phase congruency apply. Thus phase describes the local cross-sectional structure of the image surface at some location, orientation and scale, and phase-congruency provides a measure of feature strength at that location. Because phase-congruency uses information from all scales and does not rely solely on magnitude, it should be less affected by noise and has been shown to be a robust feature selector [68].

However, computing phase in the scenario above requires selection of an orientation at which to project the image at each location. To achieve this, Kovesi [68] filtered the image using a bank oriented log-Gabor filters at each scale – a method that was also employed by Schenk [103] in the mammogram CLS detection algorithm reviewed in section 3.3.1. However, this approach is intrinsically computationally expensive and coefficients generated have a high degree of redundancy.

To overcome this, Felsberg and Sommer [42] introduced the monogenic signal as a direct 2-D extension of the analytic signal. In the monogenic signal, the Riesz transform [114] is applied to an image. This produces a vector triple at each location in the image from which local amplitude, orientation and phase can be calculated. In this formulation, if the orientation and phase computed at some location are ϕ and φ

respectively, then a 1-D analytic signal in the direction of ϕ would be expected to produce a local phase of φ. As with the analytic signal, bandpass filters can be applied with the monogenic signal to compute amplitude, orientation and phase across multiple scales (Mellor and Brady [79, 80] discuss the optimal choice of filters to use). Using this approach, local phase (and phase congruency) can be computed and has been shown to be a robust measure of image features in several tasks [42, 79, 80, 86, 121]. Because phase-based measures computed from the monogenic signal are invariant to changes in contrast and intensity, they are ideally suited to applications such as cross modality registration. This was demonstrated by Mellor and Brady in their work registering MR and ultrasound volumes of the brain [80]. It may also be advantageous in mammography where different imaging equipment and different physiological factors (breast size and density) can produce images with greatly varying contrast ranges.

In this section we have looked at the benefits the monogenic signal yields as a tool for analysing medical images. However, we conclude this section by considering a disadvantage of the monogenic signal when modelling breast tissue in mammograms. By its construction, the monogenic signal only encodes information on cross-sectional shape at one orientation for any given location and scale. Thus it can only fully represent the embedding of intrinsically 1-D features (that is, lines and edges) in a 2-D signal. In a mammogram, features of a greater dimensionality may arise when one or more linear structures cross or branch, or for round objects such as calcifications. To completely represent such structures it is necessary to be able to infer cross-sectional shape in multiple orientations at a given location. We note that recent work by Wietzke and Sommer [124], proposes a further extension of the monogenic signal into a complete 2-D analytic signal, however, in the following two sections we consider an

alternative approach to obtaining structural information at multiple orientations for each location in an image.

4.3.2   The steerable pyramid

In the initial formulation of phase congruency for images described in the previous section, multiple oriented complex Gabor filters were applied to images at a range of scales. Whilst this approach is significantly less efficient than applying the monogenic signal, it does enable structural information to be captured at multiple directions for any location in any image. With this in mind, we describe the construction of the steerable pyramid in this section. This wavelet-like transform combines the concept of filtering an image with a bank of oriented filters with the efficiency gained by downsampling low frequency components in the image signal.

The complex steerable filter was introduced by Freeman and Adelson [47], and extended into a multiscale decomposition by Simoncelli and Freeman to form the steerable pyramid [109]. In the steerable pyramid, the number of orientations at which to filter the image can be chosen (hence its steerability). To form the decomposition, an image is initially low-pass and high-pass filtered. The following steps are then applied recursively to the low-pass image:

1. Apply a complex band-pass filter at the selected number of orientations

2. Low-pass filter

3. Down-sample the new low-pass image by a factor of two in each direction

The result is pyramidal structure of complex valued subbands localised in scale and orientation. The decomposition shares many of the properties of a standard real-

valued, discrete decimating wavelet transform, with the added advantages of directional selectivity, shift invariance (see below) and the computation of complex phase from which structural information can be inferred. Like a wavelet transform, the steerable pyramid is invertible: the filters are self-inverting and thus image reconstruction is achieved by reversing the decomposition process.

Despite the factor of two down-sampling, for an original $m \times n$ image, each orientated subband in the L-th level of the pyramid has $\left\lceil \dfrac{m}{2^{L\text{-}1}} \right\rceil \times \left\lceil \dfrac{n}{2^{L\text{-}1}} \right\rceil$ complex coefficients. As a result, the decomposition is redundant by a factor of $\dfrac{8k}{3}:1$, where k is the selected number of orientations.

The pyramid can be constructed using non-complex filters (as used, for example, in the model of mammograms constructed by Rose [98] we reviewed in section 3.3.1), in which case the redundancy is reduced by a factor of two. However, not only does this remove the possibility of using phase to infer structural information, the resulting real-valued coefficients will not be shift-invariant.

What is shift invariance?

In any decimating multiscale decomposition (that is, one in which the number of coefficients is down-sampled between each frequency level) the coefficients at each level form a sampling grid with respect to the original image (or signal in 1–D). Small shifts in the image prior to decomposition change the relative position of this grid. However the distribution of energy (that is, the magnitudes of the coefficients) through the levels of the decomposition should not change. They should be shift invariant.

Using complex filters, the magnitudes of pyramid coefficients will be shift invariant, whilst shifts in the decimation grid produce a global offset in coefficient phase that is cancelled out when local phase measures are computed. However, the real valued steerable pyramid, along with standard real discrete wavelet transforms (DWT), does not have this property. In this case the coefficients are shift dependent. This is because the filters do not separate positive and negative frequency components in the pyramid (analogous to the frequency spectrum of a real valued signal prior to computing the analytic signal). The negative frequency aliasing terms that cause shift dependence are cancelled out when the pyramid is inverted, and thus do not affect the reconstruction of the image. However, they prevent meaningful analysis of individual pyramid coefficients, and, as discussed later in section 8.3.1, would confound the models of mammographic appearance we construct.

As a result, we can either accept the $\frac{8k}{3}$:1 redundancy inherent in the complex valued steerable pyramid, or, as we discuss in the next section, consider a different complex wavelet transform.

4.3.3   The dual-tree complex wavelet transform

The dual-tree complex wavelet transform (DT-CWT) was developed by Kingsbury [65]. Like the complex steerable pyramid it was designed to overcome the two main disadvantages of discrete wavelets as a tool in image processing, namely the lack of shift invariance and poor directional selectivity.

The dual-tree complex wavelet transform is formed by computing two parallel real discrete wavelet transforms (trees) with a 90º phase difference. Combining the responses of the two trees as the real and imaginary part of a complex wavelet

produces coefficients with magnitudes that are shift invariant. The filters used in each tree have been the subject of careful development work to ensure the resulting dual-tree of wavelets has the desired properties. The key features of the filters are that they:

i)   allow perfect reconstruction

ii)   are linear phase

iii)   are odd-length in one tree and even-length in the other

iv)   are near orthogonal

v)   are composed of rational coefficients

Further details on the design of such filters can be found in the article by Selesnick et al. [105].

The DT-CWT extends into two-dimensions by filtering along columns then rows. Like a real DWT, wavelet coefficients at each scale are produced from the high-pass/low-pass (hi-lo), lo-hi and hi-hi filter combinations, whilst the lo-lo band is passed to the next level. However in the DT-CWT, each filter also has a complex conjugate, thus there are 12 possible hi-lo, lo-hi and hi-hi combinations at each scale. The complex nature of the filters allows for each filter and its conjugate to select either positive or negative frequencies in the horizontal and vertical directions. As a result the 12 filter responses can be manipulated to produce 6 directional subbands of complex wavelet coefficients at each scale. By construction, these subbands are aligned at approximately ±15º, ±45º and ±75º. The filters required to reconstruct the image are simply time reversed copies of the analysis filters. Thus the DT-CWT can be efficiently inverted by

separable filtering along rows and columns of the decomposed subbands.

Unlike the steerable pyramid, the number of oriented subbands cannot be changed, and their directions are fixed. However, for an original $m \times n$ image, each orientated subband in the L-th level of a DT-CWT has $\left\lceil \dfrac{m}{2^L} \right\rceil \times \left\lceil \dfrac{n}{2^L} \right\rceil$ complex coefficients. As a result the transform has a redundancy of just 4:1, which is lower than any steerable pyramid containing two or more orientations (and a factor of four smaller for an equivalent number of six orientations). Moreover, because the dual-tree is computed using separable filters, it is possible to move between the wavelet domain and the image domain with greater speed and efficiency than the steerable pyramid.

That said, as noted by Kingsbury [66] and Anderson [6], because the DT-CWT partitions the frequency domain with rectangular co-ordinates, the six subbands at any frequency level are not equidistant from the origin (the ±45º subbands formed from the hi-hi filter combinations will be a factor of approximately 1.34 further from the origin than the four bands formed from hi-lo combinations). This can be remedied by weighting the coefficients in the ±45º subbands. Alternatively, the ±45º subbands could be replaced by band-pass filters designed to match the radius of peak response of the remaining subbands, as recommended by Kingsbury [66]. Applying the extra set of filters adds to the computational cost of the DT-CWT and results in a loss of perfect reconstruction (although the effect should be small). However, such a step may be necessary if the intention is to use DT-CWT coefficients to compute rotationally invariant features in an image. In contrast, the steerable pyramid partitions the frequency domain with respect to radial co-ordinates and the peak response frequencies

of all subbands are equidistant from the origin.

With regards to its use in image processing and analysis applications, we note that the DT-CWT has yet to be applied to mammograms in any published literature. However, work by Anderson et al. [6-8] has shown how measures analogous to phase congruency can be computed using the DT-CWT. As we discuss later in chapter 8, we intend to take advantage of these features in the DT-CWT when we attempt to model the appearance of breast tissue in the region around mammographic masses.

4.4 Texture synthesis

We complete our review of general techniques in image processing with a review of texture synthesis algorithms. Whilst we have already stated our intention to synthesise the central density in mammographic masses using a statistical appearance model similar to those discussed in section 4.2.2, we have not yet determined how we might synthesise mass spicules and the distortion of breast tissue in the region of a mass. In section 3.3.1, we reviewed a method by Rose [98] for modelling and synthesising the appearance of whole mammograms. Work towards the formation of this model was influenced by techniques from the field of texture synthesis in images. Thus in this section, we review some key texture synthesis papers including further details of a statistical modelling approach developed by Rose [98] to synthesise local patches of mammographic texture.

The first algorithm we consider is a non-parametric sampling method proposed by Efros and Leung [40]. In the algorithm, a source texture I_s is used to generate matching texture in a target image I_t. The target image can be blank, in which case the algorithm must be initialised by selecting a random patch of pixels from the source texture to

act as a seed. Alternatively, the target may be a patch of unfilled pixels in an image, in which case it may be possible to use the remainder of the image as the source texture. In either case, the algorithm is controlled by fixing some patch of size κ (κ must be odd and greater than 1), and proceeds as follows. First the set of all unfilled boundary pixels are obtained (that is, unfilled pixels 8-connected to a filled pixel – if I_t was initially blank, these will be the pixels neighbouring the randomly chosen seed). One of these pixels, say $p = (x, y)$ is randomly selected, and the $\kappa \times \kappa$ patch surrounding p extracted:

$$W_t(\pm i, \pm j) = I_t(x \pm i, y \pm j), \quad i, j = 0, 1, \ldots, \kappa - \frac{1}{2}$$

Of course some of the elements in W_t are extracted from as yet unfilled pixels, and thus have no valid value (in particular $W_t(0,0)$ will always be blank). Next the set of all $\kappa \times \kappa$ patches in I_s (the exemplars) are compared to W_t. In this comparison, a distance measure is computed (for example, the sum of squares difference, weighted by a centred Gaussian kernel to give more importance to pixels closer to p than at the edge of the patch) using only those elements of W_t that have valid values. The set of exemplars with a distance measure below some preselected tolerance are considered as possible candidates, from which one exemplar is chosen at random. If W_s is the selected exemplar, then the value of the central pixel in W_s is used to populate p, that is $I_t(x, y) = W_s(0,0)$. This process is repeated until all pixels in I_t are filled.

Despite its simplicity, Efros and Leung's algorithm has been shown to be one of the most effective and flexible texture synthesis algorithms in the literature. Moreover, it can be seen as the basic template for local patch-based texture synthesis, and, as a

result has been extended in a myriad of ways. These include: extensions to work with colour images [9, 39, 122, 123, 127]; using more sophisticated methods than an exhaustive search to select the best exemplar [122, 123]; synthesising at multiple resolutions from coarse to fine [123]; adding patches rather than single pixels to the target image [9, 39, 127]; explicitly parameterising the feature space of exemplars [98]; constraining the sampling by image features [127].

Of these extensions, we consider three in more detail. Firstly, Efros and Freeman [39] extend the algorithm described above by adding patches rather than single pixels to the target image I_t. At first glance this is a trivial extension – rather than simply placing the centre pixel of the selected exemplar in the target image ($I_t(x, y) = W_s(0,0)$), all the values of the exemplar corresponding to unfilled pixels are used ($I_t(x \pm i, y \pm j) = W_s(\pm i, \pm j)$ for all $i, j = 0,1,...,\kappa - \frac{1}{2}$ and $I_t(x \pm i, y \pm j)$ unfilled). However this leaves discontinuities at the boundary of the patches that give a tiled appearance to the generated texture. This is solved by computing the optimal cut in the overlapping region between I_t and W_s, subject to an error function defined on the overlapping region. Thus in the overlapping region, some of the pixels will be maintained from I_t, whilst the remainder are substituted from W_s. Efros and Freeman term this algorithm image quilting, and like the original Efros and Leung algorithm, it has been shown to produce realistic results for a diverse range of textures. The advantage of image quilting over pixel-by-pixel synthesis is twofold: firstly, it is computationally faster as many pixels are added to the target texture during each iteration. Secondly, it is more stable – a risk in pixel-by-pixel synthesis is that the algorithm can fall into a part of the search space not well populated by exemplars, in

which case pixel values not representative of the sample texture are sampled. These pixels in turn are used to constrain the next set of unfilled pixels, locking the algorithm into the wrong part of the search space and consequently generating a target image that does not match the properties of the sample texture.

The second extension of the original Efros and Leung algorithm we consider was developed by Rose [98]. In this work, rather than simply storing all exemplars from a sample texture, the set of exemplars are treated as vectors populating a multi-dimensional feature space. The probability density of this space is modelled using a mixture of multivariate Gaussian distributions (fitted either by using k-means clustering or by using the expectation maximisation algorithm). During synthesis, the filled pixel values in the target patch W_t are used to condition the distribution. Thus a sample patch W_s can be generated from the existing values in W_t and a random sample from the marginalised distribution. As above, W_s can be used to add either a single pixel or a whole patch to I_t. In the latter case, by virtue of the conditional sampling, the pixels at overlap boundary of W_s should match the existing pixels in I_t, and thus the quilting procedure used in Efros and Leung's later algorithm is not required.

The advantage of this approach is that because the model building stage need only be performed once, increasing the number of feature vectors (for example, by using a larger sample texture or a training set of images) does not increase the computation required during synthesis. This makes the algorithm proposed by Rose more suited to synthesising texture given a large training set of image data than exemplar based methods such as Efros and Leung's algorithm. Because of this, we intend to develop an extension of this algorithm to model and synthesise mass regions (see section 8.6).

However, it should be noted that for the synthesis to be successful, it is vital that enough training data is used to adequately learn a model distribution with strong specificity. If not, then there is an even greater danger than in the original Efros and Leung algorithm of sampling texture not representative of source texture. Once one unlikely pixel is sampled, this may force future conditioning of the model to the edges of the distribution space, from which increasingly extreme sample patches may be generated. The issue of suitably populating the model feature space in our work is discussed further in section 8.6.6.

Returning to non-parametric sampling, Wu and Yu [127] present a method in which in addition to sampling exemplar patches, a feature map is used to constrain the synthesis. This feature map is composed of binary elements, specifying either strong lines or edges in the target image. During synthesis, a suitable exemplar W_s is selected with respect to the match of pixel values in the overlap region (as in Efros and Freeman's algorithm) and also to the match of features in W_s to the existing feature map. Once W_s is selected, the features within W_s are deformed to be continuous extensions of the existing feature map. This deformation is then used to warp the pixel intensities in W_s, and this is then added to the image. With regards to our work, this form of hybrid synthesis in which binary features are used to control local texture patches may provide a way of simultaneously generating spicules whilst deforming the appearance of the surrounding breast tissue. The possibility of using such hybrid models is discussed further in section 8.7.

We conclude our review of texture synthesis research by examining an algorithm of markedly different properties to those described above. Portilla and Simoncelli [95]

describe a method in which a set of parameters are used to define the appearance of a sample texture. To synthesise a new texture, a random field of Gaussian noise is sampled. This random field is then iteratively modified so that it matches each of the set of sample texture parameters in turn. Although convergence of this process is not guaranteed, the authors report that it has never failed to converge over a test set of many hundreds of input textures. The final synthesised texture should match the complete set of parameters that describe the sample. To parameterise a sample texture, a set of statistics based on its complex steerable pyramid decomposition (see section 4.3.2) are computed. These include marginal statistics of the lowpass images at each scale in the pyramid; autocorrelations of the lowpass images; cross-correlation of magnitudes in each of the oriented subbands; and cross-correlation of an interscale measure of coefficient phase. The last statistic is particularly interesting, as it uses the idea of complex phase representing image structure as discussed in section 4.3, and is very similar to the inter-level coefficient (ILP) transform of dual-tree complex wavelet introduced by Anderson et al.[7] (the ILP transform is discussed in more detail in section 8.4.3). However, by taking the expectation of each statistic across each image, we do not think the parameter that captures local structure is optimally used by Portilla and Simoncelli. Where the algorithm is of particular interest, is the idea that by modifying the coefficients in a complex valued decomposition of an image to match the statistics of some model (in this case the decomposition of a sample image), the resulting reconstructed image has matching appearance properties to the sample image. Whilst we will use a different framework (including a different choice of statistical constraints and the use of a model learnt from training data as opposed to a single image), this is an idea we seek to exploit when synthesising the appearance of breast

tissue in the region of masses.

4.5 Summary

In this chapter we have provided a review of statistical appearance models; a review of wavelet transforms and other multiscale image transforms; and finally, a review of texture synthesis techniques.

Whilst these topics may appear somewhat diverse they are linked by a common theme. The statistical models discussed in section 4.1 provide a global description of appearance. That is, given a sample of parameters from a model, the complete set of grey level intensities belonging to an image object are specified. We propose this form of model is suitable for describing the appearance of the central density in mammographic masses. However, such global models cannot cope with the heterogeneity of appearance present in the breast tissue surrounding a mass. In such regions there is a complex array of local structures. The multiscale image decompositions we reviewed in section 4.3 provide us with tools that can better describe such structures. Finally, the texture synthesis methods reviewed in section 4.4 describe appearance as series of local interactions. As we discuss further in chapter 8, we intend to combine multiscale descriptors of local structure with a texture synthesis algorithm for combining local structure patches to produce a realistic simulation of breast tissue in the surrounding region of a mass. Coupled with the global model of central density appearance, this would achieve a complete method for synthesising malignant breast masses in mammograms.

We have now completed our review of background material and can proceed to describe in detail how we have developed the models of mass and breast tissue appearance

outlined above.

Because our method is based on learning the appearance characteristics of real data, the first step in our method is to collate a suitable set of real mammographic masses on which to train our models. This process is described in the following chapter.

Chapter 5 : Data

5.1 Data Acquisition

The mammograms used throughout the project were provided by the Nightingale Breast Centre, University Hospital of South Manchester. The original mammogram films were acquired during 2004 as part of the UK national breast screening program. Sets of mammograms acquired at the 2004 screening appointment consisting of four films – the left and right CC and MLO views, were selected if they contained at least one film with an abnormality that was subsequently biopsy-proven to be malignant. The cases were selected in a sequential order, however, as no patient data was used in the selection criteria, the cases were assumed to be a random sample from the screening population. In total 540 films from 142 cases were used.

5.1.1   Film digitisation

The films were digitised on a Vidar scanner, at a resolution of 40μm per pixel and a grey level range of 12 bits (4096 grey levels). The bit depth was later reduced to 8-bit resulting in mammograms with a grey level range of 0-255. The scanner generates pixel grey levels that have an approximately linear relationship to the film density. As discussed in section 2.2.2, this imparts a linear relationship between grey level in the digitised mammograms and the sum of tissue attenuation coefficients superimposed at any location. This linear relationship is important to the work we present in the following chapter.

All digitized mammograms were fully anonymised, and no patient data was used at any stage in the project. On this basis ethics approval was not required for the work

presented in the remainder of the thesis.

5.1.2 Identifying abnormalities

Once the films were digitised, a consultant radiologist from the Nightingale Breast Centre confirmed the type of abnormality present in each mammogram and its approximate location of within the image. Table 5.1 gives the number of abnormalities present in the set of mammograms, separated by view. Note that in several mammograms more than one type of abnormality was present.

Table 5.1: Number of abnormal image signs present in the mammograms included in the dataset

	RCC	RML	LCC	LML
Mass	43	46	44	48
Individual cluster of microcalcifications	14	14	42	40
Architectural distortion or abnormal asymmetry	2	2	1	1
No abnormality present	76	74	64	63

In total there were 181 separate masses in 159 different mammograms from 87 cases. Where an MLO mammogram had multiple separate masses that were also visible in the corresponding CC view, the masses were identified and labelled by the radiologist such that we knew which masses corresponded in the two views.

5.2 Annotating masses

The methods we present in the remainder of this thesis require an accurate annotation of the border of each mass. Obtaining an annotation of sufficient accuracy is a non-trivial task – as described in section 2.2.4, breast masses, and particularly malignant masses, often have ill-defined borders; in addition a mass border may be obscured by superimposed dense tissue in the breast. For these reasons, annotation should ideally be carried out by expert mammogram readers.

Previous solutions to this problem include having readers annotate each breast lesion by hand on acetate overlaid on the mammogram film before scanning in the annotation and aligning with the digitised mammogram. Whilst this method allowed readers to make an accurate demarcation of the lesion border on the acetate, errors were introduced in the process of aligning the scanned images. Alternatively, readers have been asked to annotate directly on digitised mammograms using proprietary image software (for example, the polygon tool in Microsoft Paint). However, such software is not designed for the specific needs of mammogram annotation and may suffer from a lack of suitable features (for example, zoom and contrast control, or editing of previously defined borders) or require a greater degree of training for the reader.

To overcome these limitations, we designed and implemented a mammogram annotation tool using the graphical user-interface toolbox in Matlab®. The software allowed the user to demarcate the border of each mass in the dataset, mark further features of interest associated with the mass (such as spicules), and locate the position of the nipple in the mammogram. In addition the user could review and edit any previously saved annotations.

To enable accurate annotations, the software allowed the user to zoom within a region of a mammogram to any magnification, whilst simultaneously viewing copies of the region displayed at both the original grey-scale contrast and a contrast in which the dynamic range was maximised. To aid the annotation of spicules, we experimented with using an additional view in which lines in a region were enhanced using the line operator [36] (discussed previously in section 3.3.1). However, after consultation with the breast radiologist that subsequently used the software to annotate the masses in our data, this feature was not deemed useful. Instead, an additional view in which a region was displayed using a colour scheme that further enhanced visual contrast was included. Further details of the software are given in appendix B.

Using the software the 181 masses in the dataset were annotated by a breast radiologist from Nightingale Breast Centre. Figure 5.1 shows two such examples, one with and one without spicules marked.

Whilst annotating the masses, the radiologist was asked to rate how well the border was defined on a scale from 1 to 5, with 1 being the least well defined. Note that this rating should not be seen as equivalent to the way a radiologist might score a lesion boundary to determine lesion type, but rather a general measure of how accurately the radiologist felt they could mark the limits of a mass using the software. Thus rather than being solely a measure of the mass, the rating took into account other factors, such as dense parenchymal tissue obscuring the border.
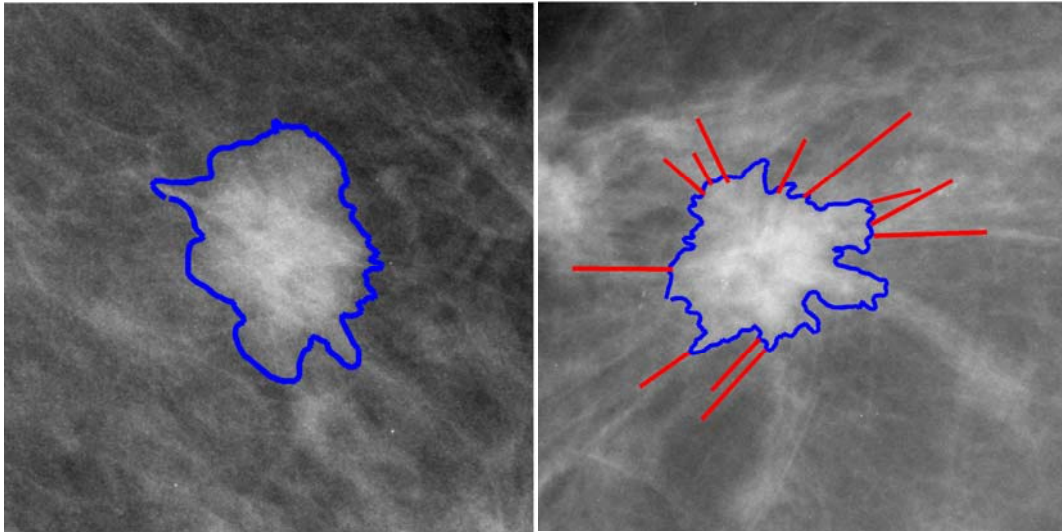
Figure 5.1: Two masses from the dataset with the border, as marked by a breast radiologist, shown in blue. On the right hand mass, several spicules are also marked in red

Of the 181 masses annotated, two were assigned a rating of 1 by the radiologist. In conjunction with the radiologist it was decided these masses were not suitable for inclusion in any further models. In both cases, the rating of 1 was assigned because a large portion of the mass was obscured by dense parenchymal tissue, and thus attempting to delineate the mass border was impossible. It was deemed that this was a feature of the breast as opposed to a particular feature of the two masses. As such we do not believe that excluding the two masses from subsequent datasets biased the models we develop.

5.3 Selecting a dataset of independent masses

As a result of discarding two masses following expert annotation, the dataset comprised 179 mammographic masses that were projections of 101 breast masses from 86 separate cases.

We use these masses to form a training set from which to build the statistical models of appearance described in the remainder of this thesis. However, we note that the CC and MLO mammographic projections of the same mass are likely to have highly correlated appearance properties, especially as the two projection angles are not orthogonal. In particular, the size, average grey level and boundary properties of a single physical mass projected in the two views will be closely related.

Further, not all of the 101 masses are visible in both CC and MLO views. Thus, if a training set was formed using all 179 mammographic masses, then the set would contain a bias towards the appearance properties of masses that were visible in both views. To prevent this bias, where a mass was visible in both CC and MLO views, only one of the views was included in the training set. In these cases, the mammogram view to be included was randomly selected. Where a breast contained more than one mass, a single projection of each separate mass was included. Other ways of selecting a training set given the data that were available are discussed in the following section.

As a result of the selection process outlined above, the final training dataset comprised 101 masses, of which 31 were from left MLO mammograms, 20 from left CC mammograms, 27 were from right MLO mammograms and 20 were from right CC mammograms.

For computational efficiency, in every mammogram containing a mass, a rectangular region containing the mass was extracted. Each region was selected to contain the mass in its entirety (including any spicules marked by the radiologist) with an additional buffer of 200 pixels (~80mm). Thus from here until chapter 9 we refer to our training data as comprising mammogram regions. In chapter 10 we return to considering full

mammograms when we discuss how to automatically locate synthesised masses.

5.4 Discussion

In section 5.2, we described how masses in our data were annotated by a breast radiologist. We developed software to allow the radiologist to demarcate the border of each mass as accurately as possible. However, this task involves a subjective judgement and, despite the radiologist's expertise, may be prone to human error – particularly for masses with an ill-defined border. Because of this, it may have been advantageous to have had masses annotated by multiple experts. This would allow us to compute the variability between experts and set a measure on the confidence we had in the annotations as ground truth for the data.

However, because we attempt to model the variation in appearance across whole mass regions, our method is less sensitive to inaccuracies in the annotated borders than if we modelled only the image data contained within each mass border. As we show in the next chapter, pixel grey levels in each mass region are split into data used to model central mass densities and data used to model the surrounding tissue. Thus varying the annotated border of a mass may change how grey levels in the region are split between the two datasets, but does not affect the overall content of data modelled.

In section 5.3, we described how following annotation, we selected a subset of masses that would be used to train the models of appearance we build in subsequent chapters. Treating the mammographic projection of each physical mass as a separate image object, we randomly selected a single CC or MLO projection of each mass to be included in the training set. We now consider whether this was the best method for

choosing a dataset.

Firstly, we might instead have used all 179 mammographic masses. However as discussed in section 5.3, the appearance characteristics contained in this set, and consequently learnt by the models, would be biased towards the appearance of masses that were visible in both CC and MLO views. As a result, synthetic data sampled from the models would also be biased towards these appearance characteristics. Thus whilst using all the mammographic masses may provide a richer set of data from which to build statistical models of appearance, we could not trust synthetic data sampled from the models to be representative of the global population of masses. This would limit the use of such synthetic data, and thus we do not think using all 179 mammographic masses is a sensible solution.

An alternative solution would be to use both MLO and CC projections, but include only those masses that were visible in both views. There are 78 masses visible in both views and so such a set would comprise 156 training examples. However because the appearance characteristics of each CC/MLO pair are strongly correlated, this set contains less variation in appearance than the set generated using a single view of all 101 masses. That is, in discarding 23 masses because they were not visible in both views, we have lost more variation than we gain by adding the extra projection of the remaining 78 masses.

A further point to consider is whether it is valid to build a model from masses using a mix of CC and MLO views. In building statistical appearance models we are assuming there is some underlying distribution that describes how a mass can appear in a mammogram. By including masses from both views, we are making the assumption that

this underlying distribution does not differ between the two. If we consider the physical 3-D location of a mass within a breast, then this assumption is equivalent to saying that the alignment of a mass (and any local structures associated with mass) with respect to the x-ray source for one view, is as likely to occur as an equivalent alignment with respect to the other. Whilst we acknowledge the theory that masses grow in ducts radiating from the nipple [52], we are not aware of any data in the literature showing a constraint on the specific rotational alignment of a mass. Thus we believe it is valid to build appearance models from a dataset containing a mix of mammogram views. This may however be an area for further research.

Under the assumption that it is valid to build appearance models from a dataset containing both MLO and CC views, the selection criteria described in section 5.3 for constructing a training set maximises the variability in mass appearance without introducing bias into the data.

Finally, we may consider whether we should try and learn the appearance relationship between the projections of a physical mass in two mammogram views. Learning this relationship would be necessary to generate complete cases of synthetic data (as opposed to individual mammograms containing a synthetic abnormality). However, an attempt at a building such a model is beyond the scope of this thesis, although would again be an interesting area for further research. This is discussed further in section 12.2.2.

5.5 Summary

In this chapter we have described the origin of the mammograms used throughout the project. We have described the process by which the masses present in the

mammograms were identified and annotated by a breast radiologist.

The method for selecting a suitable training set of masses from all those available was discussed, with the conclusion that for our models choosing a single mammogram projection of each separate physical mass was most appropriate. This ensured the assumption that the objects in the training set were independent examples drawn from the global population distribution holds true, whilst maximising the number of examples in our training set.

In the following chapter we describe the pre-processing steps required to convert mammogram regions containing masses into regions from which appearance models can successfully be built.

Chapter 6 : Separating a Mass from the Background

Following our review of mammography literature, we determined that we would attempt to synthesise malignant mammographic masses by modifying the appearance of real normal mammograms (see section 3.2.1).

Recalling the mammographic appearance of breast lesions described in chapter 2, we observe that the central density of a malignant mass appears as an area of higher grey level than the neighbouring breast tissue. This occurs because the cancerous cells comprising a mass have higher x-ray attenuation coefficients than normal breast tissue. In simulating the presence of a mass in a normal mammogram, we are effectively simulating this pattern of increased attenuation.

In the previous chapter we described how we have constructed a set of real masses from which we will attempt to learn a model of mass attenuation. To do this, we make use of the fact that there is a linear relationship between pixel grey level in the digitised mammograms and the sum of attenuation coefficients corresponding to superimposed tissue (see sections 2.2.2 and 5.1.1). As a result, an increase in x-ray attenuation due to the presence of a mass causes a proportional increase in pixel grey level. If we can

i) model the pattern of increased grey level for each mass in our training data

ii) synthesise new patterns of increased grey level that match the appearance of the real data included in the model;

then adding the synthesised grey levels to a normal mammogram is equivalent to simulating a pattern of increased attenuation corresponding to the presence of a mass in

the breast[2]. Thus, our appearance model of image grey levels corresponds to a physical model of mass attenuation.

In this chapter we describe how we compute the increased pixel grey level corresponding to the central density of each mass in our training data. We refer to this process as separating a mass and describe the two sets of data we form as a result of this process. Finally, we discuss how this motivates the work we present in chapters 7 and 8.
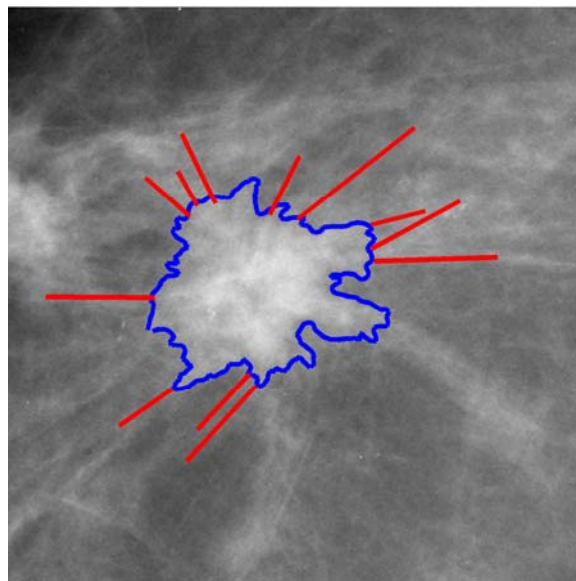
Figure 6.1: A mass region from the dataset. The mass border (as annotated by a radiologist) is marked in blue. Several (although not all), spicules have been marked in red

6.1 Introduction

We now consider the appearance of a mammogram region from our dataset, as for example, shown in Figure 6.1. The principal feature of the region is the area of increased intensity due to x-ray attenuation through the mass, causing a blob-like object

[2] Assuming the normal mammogram is also obtained using the same film-screen and digital scanner. Further details of this limitation are given in the discussion in section 6.6.

in the centre of the region, marked by the blue border. In addition there are spicules associated with the mass, some of which have been labelled in red. Finally the region contains the projection of all the normal breast tissue in the area (some of which may have been distorted by the presence of the mass) that produces the characteristic textured appearance of mammograms.

Our aim in this chapter is to represent the grey level of every pixel in the region as being the sum of two sources of attenuation:

i)   the central mass density, which we assume is a contiguous object

ii)   the combination of mass spicules (and any other abnormal structures such as microcalcifications) and all other normal breast tissue in the region

Figure 6.2 shows a schematic of this idea. Throughout the remainder of this chapter we will refer to the first source as *mass attenuation* and the second source as *background attenuation*.

Figure 6.2: Schematic of the process for separating the central density in a mammographic mass. Green lines represent breast parenchyma, red lines represent mass spicules.
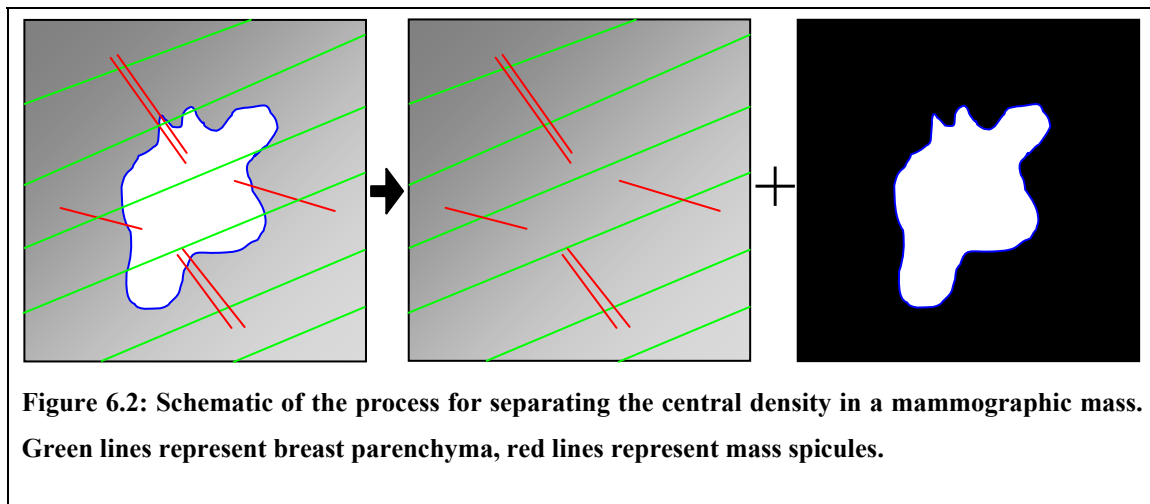
At first it may seem odd to include mass spicules in the background attenuation, as opposed to the mass attenuation. However we make this choice for three reasons.

Firstly, the central mass can be considered as a single object, of which we have exactly one in each region of the training set. This allows us to model the population of masses using the statistical appearance model presented in the next chapter. However, there is no consistent set of structures associated with a mass across all training examples. For example, some masses are highly spiculated, whilst others have few or no spicules; some masses may contain clusters of microcalcifications, whilst others do not. Thus attempting to include these abnormal structures in the model of mass appearance would prevent us from finding a consistent set of correspondences across each object in the training data, thus confounding the model.

Secondly there is a practical consideration: whilst the increased image intensity corresponding to the central density of a mass can be successfully separated from the background (as we show in the remainder of this chapter); the fine-scale, low contrast characteristics of many spicules make them very difficult to separate from the background. This process would be further complicated by the similarity in appearance between mass spicules and some normal linear structures in the region.

Finally, and related to the point above, mass spicules are inextricably linked with the normal breast tissue. Indeed, as described in our review in section 3.2.1, failing to account for the interaction between a mass and its surrounding background was one of the principal shortcomings of previous attempts to synthesise mammographic masses.

Overcoming this failure is one of the key aims of this project. It is between mass spicules and the existing normal breast tissue that the majority of this interaction takes place. Separating the central mass, whilst leaving any spicules produces a set of training regions from which we can learn to model (and synthesise) this interaction.

6.2 Representing a region as the sum of mass and background

To proceed with separating the regions in our dataset, we first formalise our definition of how mass attenuation and background attenuation combine to form pixel grey levels in each region.

Assume we have a $X \times Y$ pixel mammogram region containing a single mass. We can define the grey levels at each pixel throughout the region by the function:

$$R(x, y), \quad for \quad x = 1, ..., X, \quad y = 1, ..., Y$$

Because the relationship between x-ray attenuation and image grey level is linear, using equation (2.1) from chapter 2, we can rewrite the grey level at any pixel in the region as

$$R(x, y) = \alpha \left(\int \mu(x, y, z) dz \right) + \alpha \log \left(E_0(x, y) \right) + \beta$$

where $\int \mu(x, y, z) dz$ is the sum of superimposed tissue attenuation coefficients, $E_0(x, y)$ is the incident exposure of the film to the x-ray beam and α and β are constants that depend on the film-screen combination and scanner used to obtain the digitised mammogram. We can then separate the sum of superimposed tissue so that

$$R(x, y) = \alpha A_M(x, y) + \alpha A_B(x, y) + \alpha \log \left(E_0(x, y) \right) + \beta$$

where A_M and A_B represent mass attenuation and background attenuation respectively.

Now let the annotated mass border obtained in section 5.2 define the set of pixels that have an increase in grey level attributed to mass attenuation. We can label this set of pixels

$$M = \{(x, y)\}$$

so that

$$A_M(x, y) = 0 \quad \forall(x, y) \notin M$$

Finally, we can rewrite the pixel grey level as

$$R(x, y) = R_M(x, y) + R_B(x, y), \quad R_M(x, y) = 0 \quad \forall(x, y) \notin M$$

where $R_M(x, y) = \alpha A_M(x, y)$ is the increased grey level due to the presence of the mass and $R_B(x, y)$ represents a background grey level corresponding to the sum of breast tissue attenuation coefficients and the incident exposure at (x, y).

Applying this representation across a region, we have effectively separated the region into two new images R_M and R_B. We define these as the *mass image* and the *background image* respectively.

Having described a region in this manner, the task for the remainder of this chapter can be defined as finding the best estimates for the unknown background image grey levels (that is, $R_B(x, y)$, $\forall(x, y) \in M$) for each region in training data. In the next section, we look at an initial method for estimating R_B using thin-plate spline interpolation.

6.3 Estimating background images using thin-plate spline interpolation

Our first attempt at estimating the background image in a region is based on the method developed by Caulkin [23]. The idea is to use a thin-plate spline interpolant to estimate the unknown $R_B(x, y)$ at each pixel within the mass border, based on the pixel grey levels sampled from a set of locations in concentric rings outside the mass border.

This is equivalent to setting the image surface as an infinitely extending thin-plate of metal, with known heights at each of the sampled locations. The resulting interpolant gives the height at any other location such that the total energy of the thin-plate is minimised. Formally, if we define our set of sample locations as L, then we seek the function $f(x, y)$ that minimises

$$E_f = \int\int \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy$$

Given the constraint that

$$f(x, y) = R(x, y) \quad \forall (x, y) \in L$$

Solutions to this system were published by Bookstein [18], and the solution we use is given in appendix A.2. Having determined $f(x, y)$ we can then set

$$R_B(x, y) = f(x, y) \quad \forall (x, y) \in M$$
$$R_B(x, y) = R(x, y) \quad \forall (x, y) \notin M$$

However, one problem with this approach is that the interpolating function $f(x, y)$ is susceptible to large fluctuations if any of the grey levels at the sample points are unusually high or low. In a mammogram, and especially in the region of a malignant

mass, there is the particular problem of sampling a high grey level from a pixel lying on a structure such as a spicule, duct or microcalcification, or even simply a pixel of high intensity speckle. To overcome this, we first smooth the region with a Gaussian kernel of standard deviation σ to obtain

$$R' = G\left(0, \sigma^2\right) * R$$

and then reset the constraint

$$f(x, y) = R'(x, y) \quad \forall (x, y) \in L$$

It remains to specify the set of sample locations L. Clearly there are many ways of selecting L, however the method employed by Caulkin uses three parameters to specify a set of sample points that lie in a bounded concentric ring of the annotated mass border. The parameters are described below, and diagrammatically in Figure 6.3.

- n_1 defines the distance in pixels from the mass border to the first ring of sample locations

- n_2 defines the width of the ring from which sample points are selected.

- d defines the distance between each sample point in the vertical and horizontal directions.

In the next section, we describe how suitable values for n_1, n_1, d and σ can be determined experimentally.
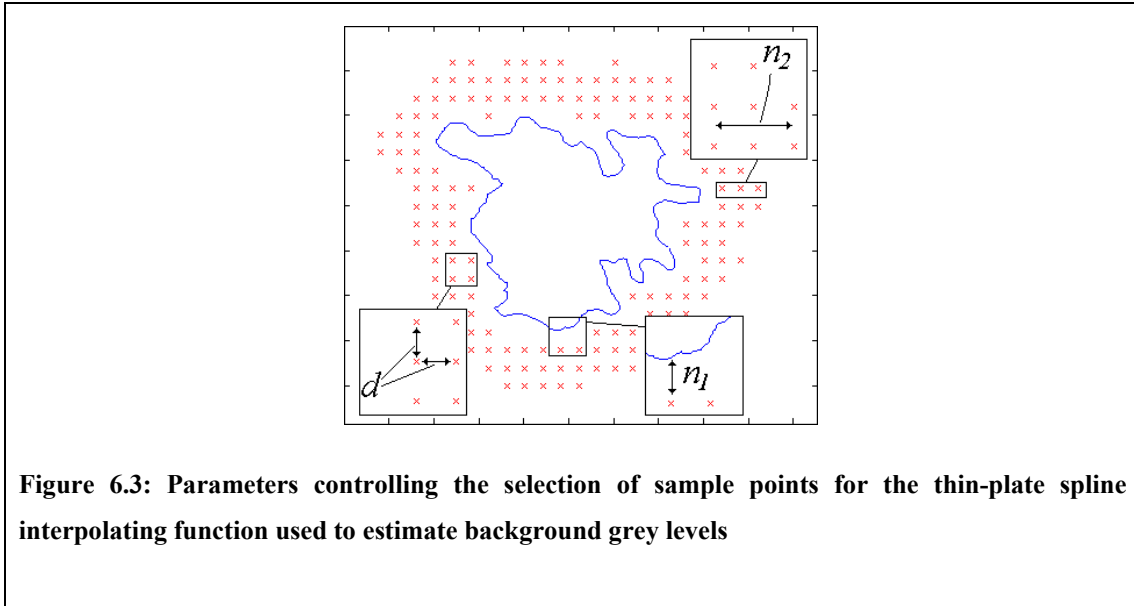
Figure 6.3: Parameters controlling the selection of sample points for the thin-plate spline interpolating function used to estimate background grey levels

6.3.1 Experimentally selecting parameters

To determine values for n_1, n_2, d and σ experimentally, we used regions for which the true values in R_B are known.

To obtain such regions, we added masses that had already been separated to normal mammographic background, thus generating an approximate simulation of the masses in our training set. Of course this required making an initial separation of the masses in our data, which in turn requires choosing an initial set of parameter values.

To choose an initial set of parameters, we observed that previously Caulkin had used values of $n_1 = 20$, $n_2 = 20$, $d = 5$ and $\sigma = 5$. Next we noted that the mammograms we are using were digitised at a spatial resolution of 40μm per pixel compared to 100μm in Caulkin's work. As all four parameters are directly related to pixel distances, we scaled each value accordingly to obtain $n_1 = 50$, $n_2 = 50$, $d = 12$ and $\sigma = 12$. We then separated the mass intensities in each of our 101 training examples using the thin-plate

spline method outlined above. From this set, we randomly selected 50 separated masses, each of which was superimposed on a normal mammogram randomly selected from our data. In doing so, we effectively obtained 50 mass regions for which we knew true values for M, R_M and R_B. Thus for each of these test regions, we can compute the error of an estimated background $\widehat{R_B}$ as $e = \left| R_B - \widehat{R_B} \right|_{RMS}$. Taking the mean of all such errors across the test regions provides a measure of performance for a given parameter set. As a result we were able to systematically test different combinations of n_1, n_2, d and σ.

Firstly, all combinations of $n_1 = 0,10,...,100$ and $\sigma = 5,10,...,35$ were tested with $n_2 = 20$ and $d = 10$. The mean error using each of these parameter combinations are shown in Figure 6.4. It was observed that a combination of $n_1 = 60$ and $\sigma = 25$ produced the lowest errors.

Figure 6.4: Background estimation errors associated with combinations of n_1 and σ ; n_2 = 20 and d = 10 were fixed
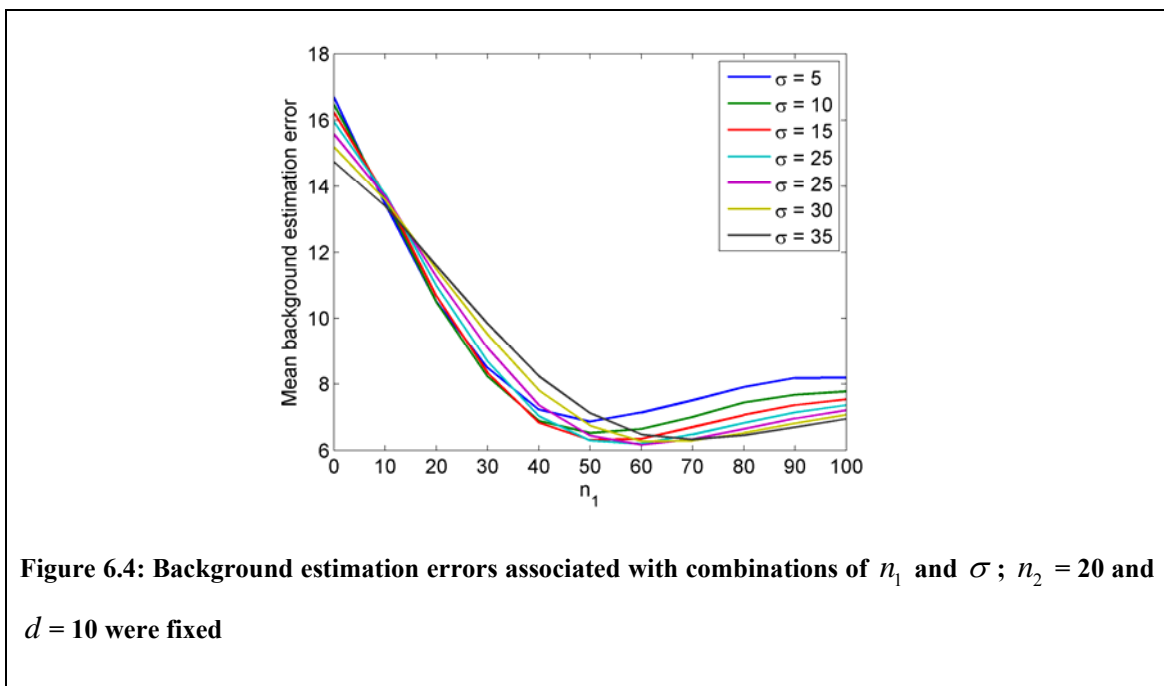
Next, values of $n_2 = 10, 20, ..., 50$ were tested for fixed values of $n_1 = 60$, $\sigma = 25$ and $d = 10$. It was observed that varying n_2 made negligible difference to the mean error, and thus a value of $n_2 = 10$ was selected for computational efficiency. Finally, for fixed $n_1 = 60$, $\sigma = 25$ and $n_2 = 10$ values of $d = 5, 10, 20, ..., 50$ were tested. In this case, $d = 10$ returned the lowest mean error, although there was only a small increase in error associated with the other values.

From the results, it is clear the most important choices of parameter are the combination of n_1 and σ. After smoothing the region, the increased intensity corresponding to the central density will be dispersed outside of the mass border. If sample points are selected too close to the mass border, the resulting interpolating function will return intensities higher than desired. Therefore the greater the degree of smoothing, the larger n_1 should be. However, the further the first set of sample points are from the region to be estimated, the less accurate the interpolated estimates are likely to be. Thus there must be a compromise between selecting a small n_1 and applying a large degree of smoothing.

Because of this interaction, adding masses to our test regions was crucial to make a realistic judgement of the best (n_1, σ) pair to use. If we had not added the masses, then the parameters selected would have been biased towards high σ and low n_1 combinations that in the presence of a real lesion would cause high mass intensities to be blurred into the region of sample points.
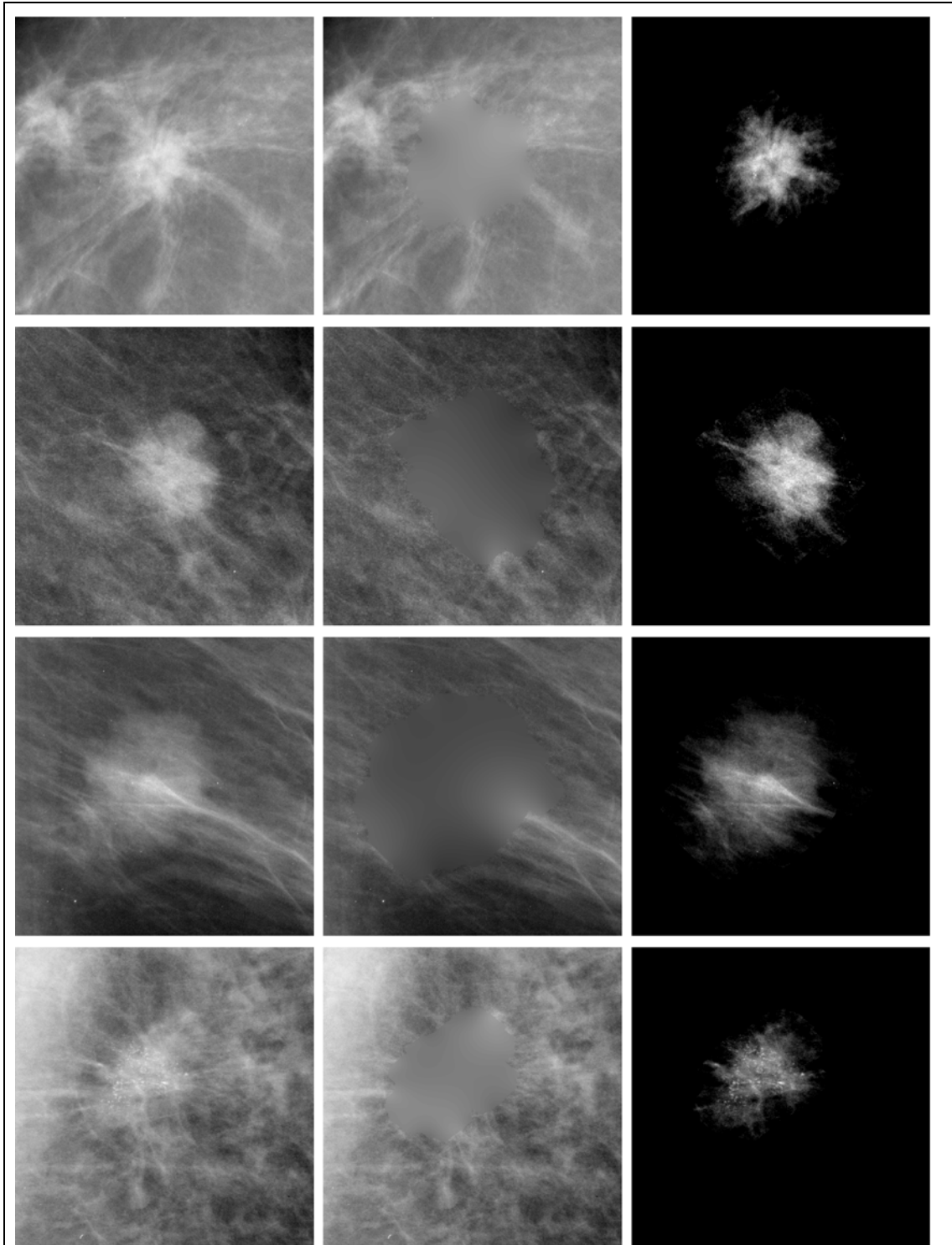
Figure 6.5: Estimating background intensities using the thin-plate spline method for four regions from the dataset. In each row, the left column: original region R, centre: estimated background intensities R_B, right: separated mass R_M

6.3.2   Evaluating background estimation in real masses

Having selected parameters of $n_1 = 60$, $n_2 = 20$, $d = 10$ with which to define sample locations, and Gaussian kernel with a standard deviation of $\sigma = 25$ to smooth each region, we used a thin-plate spline interpolant to estimate values for R_B and R_M for each real mass. Figure 6.5 shows four such regions from our data. In the left-hand column the original regions are shown. The centre column displays the estimated R_B, whilst the right-hand column shows the resulting separated masses, R_M.

From the nature of the problem, it is hard to evaluate how well we have estimated R_B. However a simple qualitative test is that the estimated region should have the appearance of real tissue in a mammogram. Looking at the regions in Figure 6.5 this clearly isn't the case: whilst we have achieved a good average approximation of the intensities, the estimated regions are much smoother than the real appearance of breast tissue.

By construction, the thin-plate spline interpolating function is only suitable for predicting a surface that comprises low frequency components. As a result, the estimates for R_B in the region M contain no high frequency structures such as individual ducts or mass spicules. Consequently these structures are instead included with the mass intensities R_M. Unlike the ideal separation of intensities represented in Figure 6.2, the separation we have achieved is summarised by the schematic in Figure 6.6.
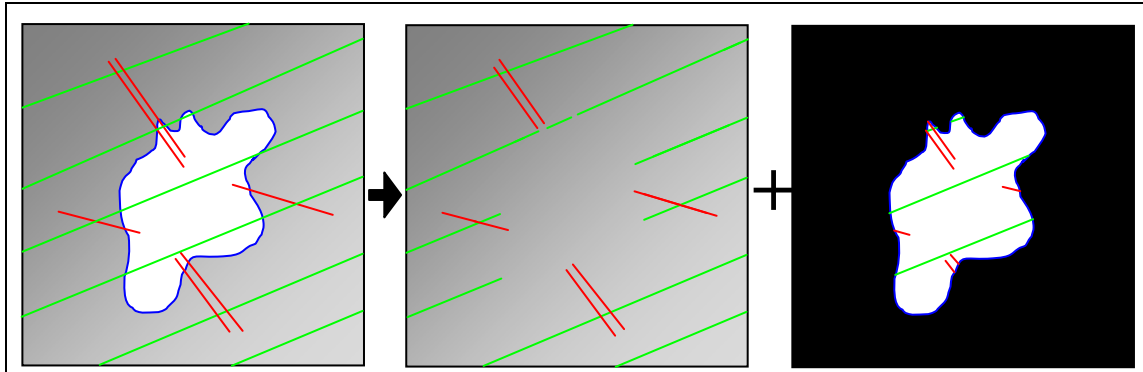
Figure 6.6: Schematic of the separation achieved by the initial thin-plate spline interpolation. Within the mass border, breast parenchyma and mass spicules have been included with the mass image

Caulkin's method for generating synthetic masses proceeded to fit appearance models to the separated masses at this stage. However we believe this separation method is unsatisfactory. There will not be consistent correspondences between the high frequency components across the set of separated masses. As we see in the following chapter, this greatly reduces the quality of fit for the texture component of the appearance model, and ultimately reduces the ability of the method to generate realistic synthetic masses. In the next section, we make a crucial modification to the separation method that resolves this problem.

6.4 Adding high-frequency components to the estimates

In the previous section we showed how the thin-plate interpolation method gives a good initial approximation to the average intensities of R_B in the region M. The aim in this section is to show how more realistic, pixel-by-pixel estimates for R_B can be computed. The key, as we show below, is to transfer the high frequency components in the region

M , from R_M to R_B .

Before introducing the new method for estimating R_B and R_M , we first summarise the

four key stages of the existing method.

1) Smooth the original region R , with a Gaussian kernel G , to obtain

$$R' = G * R$$

2) Select sample locations L , and compute the interpolating function of minimum

energy f , subject to the constraint that

$$f(x, y) = R'(x, y) \quad \forall (x, y) \in L$$

3) Use f as the estimates for R_B in the region M

$$R_B(x, y) = f(x, y) \quad \forall (x, y) \in M$$
$$R_B(x, y) = R(x, y) \quad \forall (x, y) \notin M$$

4) Compute the mass-only intensities

$$R_M(x, y) = R(x, y) - R_B(x, y)$$

Figure 6.7 shows each stage of the method for one of the regions included in Figure 6.5.

To aid visualisation, at each step a three-dimensional representation of the region is

shown alongside the two-dimension intensity image. From this representation, it is clear

the extent to which all the high-frequency components of the background tissue are

instead included with the mass.

Original region R

1) Compute smoothed region $R' = G * R$

2) Compute interpolating function f

3) Use f as the estimates for R_B in the region M

4) Compute the mass intensities $R_M = R - R_B$
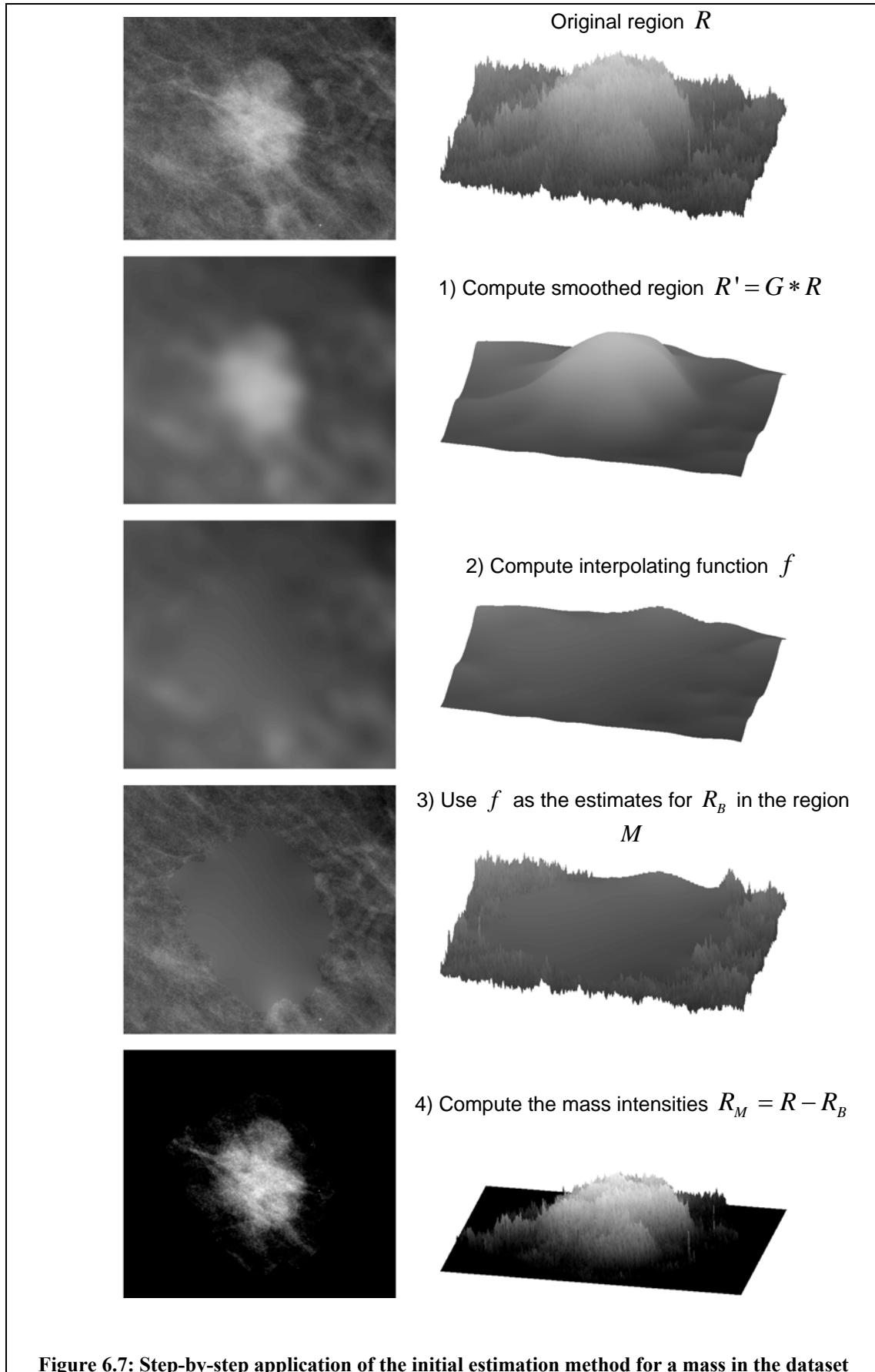
Figure 6.7: Step-by-step application of the initial estimation method for a mass in the dataset

Original region R

1) Compute smoothed region $R' = G * R$

2) Compute interpolating function f

3) Compute the mass intensities
$$R_M(x, y) = R'(x, y) - f(x, y) \quad \forall (x, y) \in M$$
$$R_M(x, y) = 0 \quad \forall (x, y) \notin M$$

4) Compute the background intensities
$$R_B = R - R_M$$
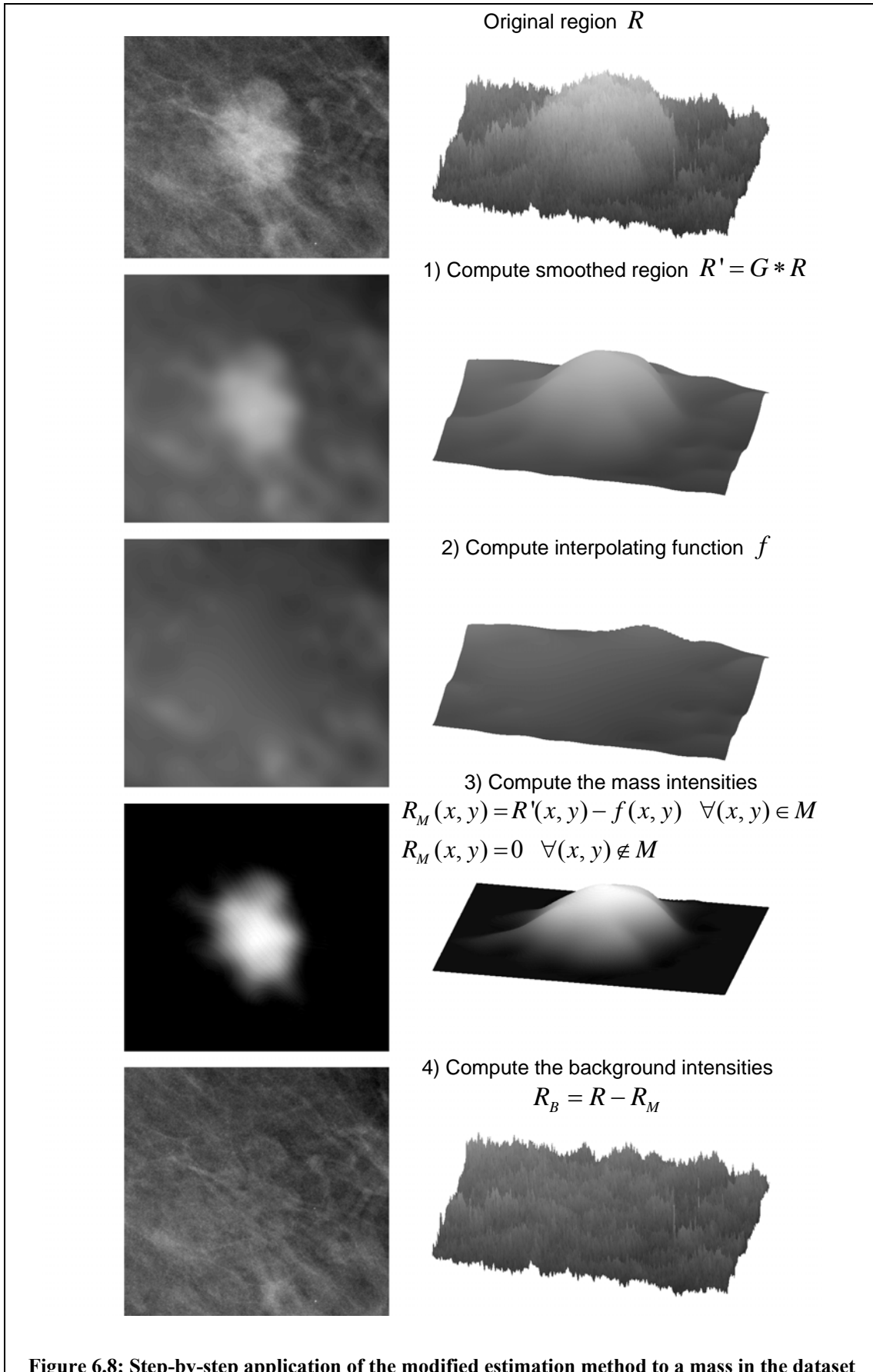
Figure 6.8: Step-by-step application of the modified estimation method to a mass in the dataset

Our proposed update to the method involves a re-ordering of the process at stages 3) and 4) as described below

3) Use the smoothed region R', and the interpolating function f, to compute estimates for R_M in the mass region[3]

$$R_M(x, y) = R'(x, y) - f(x, y) \quad \forall (x, y) \in M$$
$$R_M(x, y) = 0 \quad \forall (x, y) \notin M$$

4) Now use R_M to compute the background intensities

$$R_B(x, y) = R(x, y) - R_M(x, y)$$

All four stages of the new method are shown in Figure 6.8. Note how the high-frequency components are now included in R_B. As a result the estimated background appears as a realistic rendering of mammographic tissue, and the separated central mass is free from artefacts that would otherwise confound the appearance models presented in the next chapter.

To conclude this section, in Figure 6.9 we display the results of the new method for each of the four masses previously shown in Figure 6.5.

---

[3] To ensure we don't have a sharp cut-off at the mass boundary, we also expand M to include all pixels up to the first ring of landmark location
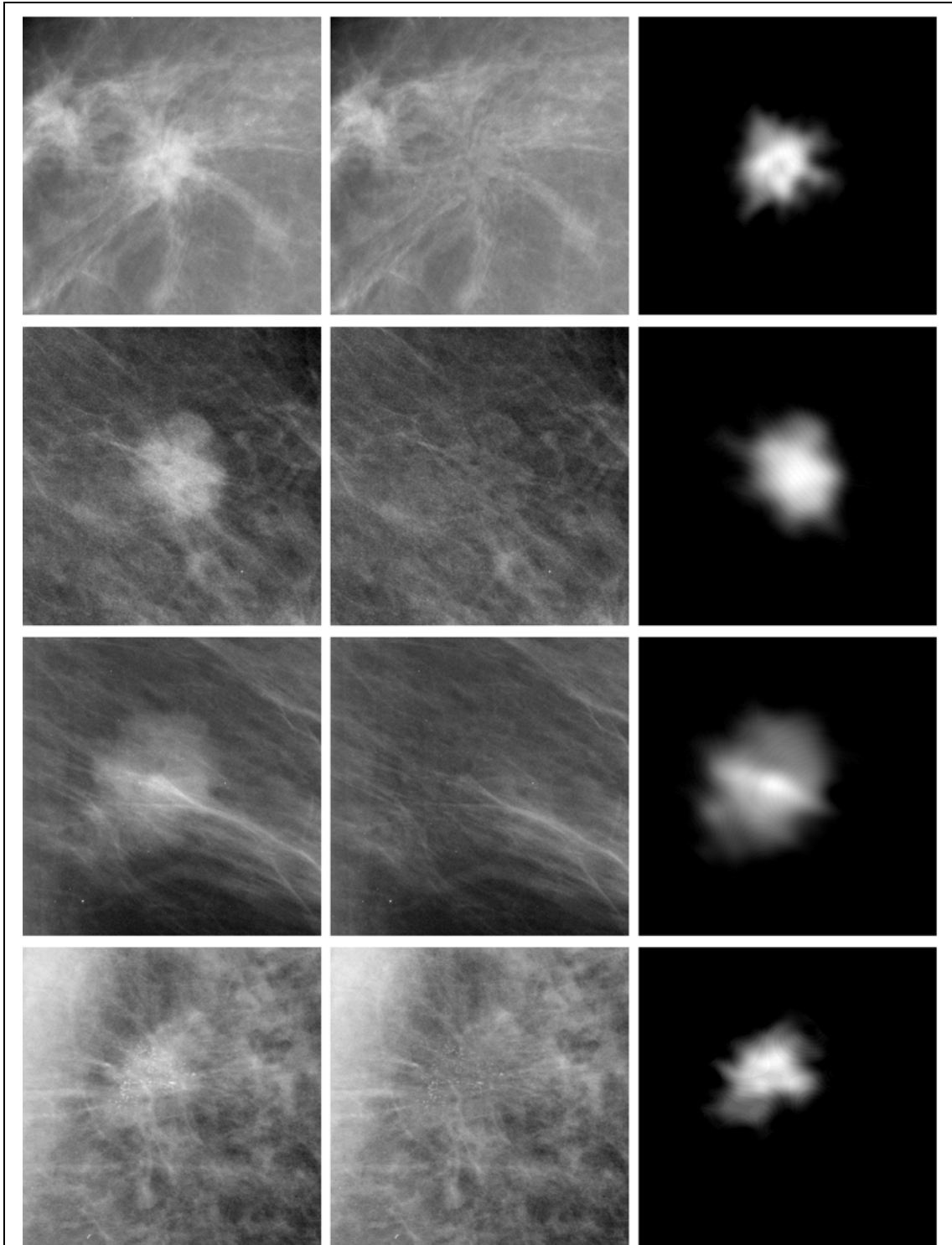
Figure 6.9: Estimating background intensities using the modified estimation method for four regions from the dataset. In each row, the left column: original region R **, centre: estimated background intensities** R_B **, right: separated mass** R_M

6.5 Iteratively updating the initial estimates

In section 6.3.1 we described the compromise between choosing the first ring of sample points close to the mass border (by varying n_1), and choosing a degree of smoothing large enough so that the thin-plate spline interpolating function isn't overly perturbed by noise (by varying the standard deviation of the Gaussian kernel, σ). Values of n_1 and σ were chosen such that the support of the smoothing filter overlapped the mass region and the first set of sample points. Across the dataset, these values were deemed optimal, however in masses that have a significant amount of intensity near the mass border, this intensity may be smoothed into the sample points. The result is an interpolation that produces higher values in f, and subsequently R_B, than we would expect.

This phenomenon is best observed by analysing intensity profiles of affected regions. For example, Figure 6.10 (a) shows one such region, from which intensity profiles have been sampled at the dashed yellow line. In Figure 6.10 (b), the intensity profile through the smoothed region R' is plotted in blue, together with the intensity profile of the interpolating function f in the region M (the dashed red line).

One solution could be to try and adapt the parameters used for individual masses, however it is not obvious how this could be achieved automatically. Instead we have developed an iterative algorithm for estimating R_B, as described below.

For each region, we proceed as before, smoothing the region before computing the thin-plate interpolating function. However, rather than continuing to compute R_M and R_B, we transfer the interpolated intensities into the original region, and compute a new smoothed region. If the new smoothed region has as a lower average intensity than the

old, we iterate again. If the average intensity doesn't decrease, we terminate the algorithm and proceed to compute R_M and R_B. The algorithm is described formally below:

Compute the smoothed region $R' = G * R$, and the interpolating function f as previously

Transfer the interpolated intensities into the original region

$$R(x, y) = f(x, y) \quad \forall (x, y) \in M$$

Compute a new smoothed region $R_{new} = G * R$ and initialise $R_{old} = R'$

while $\|R_{new}\| < \|R_{old}\|$

    Set $R_{old} = R_{new}$

    Compute new interpolating function f based on the constraint that

$$f(x, y) = R_{old}(x, y) \quad \forall (x, y) \in L$$

    Set

$$R(x, y) = f(x, y) \quad \forall (x, y) \in M$$

    Compute new smoothed region $R_{new} = G * R$

end

Compute $R_M = R' - R_{new}$

Compute $R_B = R - R_M$

Algorithm 6.1: Iteratively updating a TPS interpolation to estimate background intensities in mass regions

In cases where the intensities of sample points were not affected by the mass, the algorithm terminates after the first iteration and the estimates for R_B and R_M remain unchanged. However, consider a region where the intensities of the sample points in the smoothed region R have been increased as a result of the mass. When the interpolating

function f is computed, the values of f in the mass region M will be higher than desired. However, they will of course be lower than the intensities in the original smoothed region. Thus when we compute the new smoothed region, there is less chance of high intensities from M unduly affecting the sample points. As a result the new values of the interpolating function should, on average, decrease as desired. The process repeats until no intensities from the region M are affecting the intensity at the sample points, this being the condition we were seeking to achieve.

Figure 6.10 (c), shows an intensity profile of the smoothed region at each iteration of Algorithm 6.1, again sampled at the dashed line in Figure 6.10 (a). For this region, the algorithm terminated after the fourth iteration, with the final smoothed estimates marked by solid blue line in Figure 6.10 (c). Note that at the point of highest intensity, the final interpolated intensity is 6 grey levels lower than the initial interpolation, a change equivalent to 25% of the mass intensity at that point.

The iterative interpolation algorithm was applied to compute R_M and R_B for each region in the dataset, completing the separation process.

6.6 Discussion

In this chapter we have shown a method for separating each region in our dataset into a mass image and a background image. In each region, the mass image comprises grey levels that correspond to mass attenuation. The appearance properties of the set of mass images can be modelled and used to synthesise new examples. These synthetic mass images can then be superimposed on normal mammographic background to simulate the increased attenuation associated with a real mass.
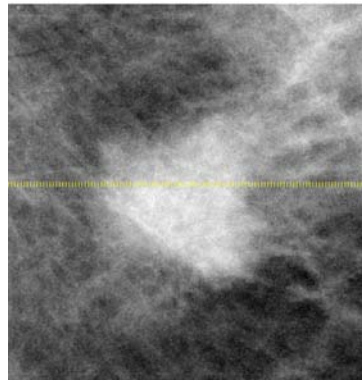
(a)
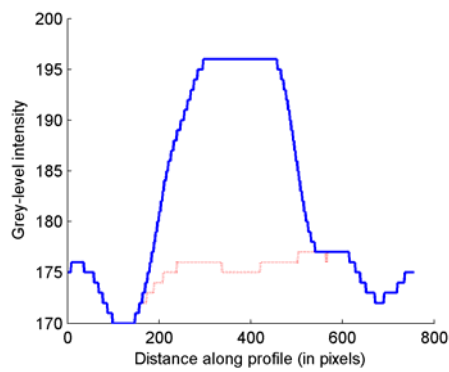
(b)
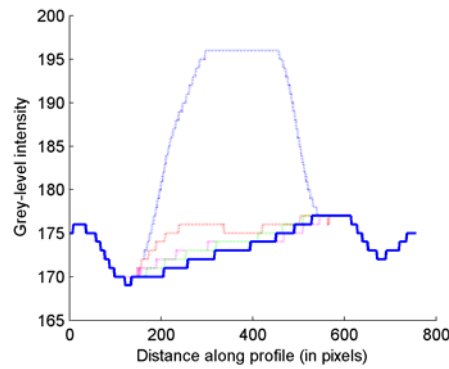
(c)

Figure 6.10:a) a region in which high intensity near the mass border results in a higher than desired initial TPS interpolant; b) intensity profiles sampled from S (blue) and f (dashed red) at the location of the yellow line marked in (a); c) intensity profiles of successive iterations of f

144

The physical model underpinning this work is the assumption that there is a consistent linear relationship between image grey level and tissue attenuation for all the digitised mammograms in our data. Thus if $g = \alpha A$ is the grey level associated with tissue attenuation A in any mammogram, then adding g to the grey level of a pixel in any other mammogram is equivalent to simulating an increase in attenuation of A at that pixel. As a result we can synthesise increased attenuation corresponding to a mass in a normal mammogram without having to explicitly calculate α.

However this only holds true for digitised mammograms with the same constant of proportionality α between image intensity and attenuation. For example, consider a normal mammogram obtained using a film-screen/digital scanner that also imparts a linear relationship between the two, but with a constant of proportionality α'. Following the equations described in section 6.2, we see that g would now correspond to an increase in attenuation of $\frac{\alpha}{\alpha'} A$.

Thus our model of mass image appearance only corresponds to a valid model of the physical properties of mass attenuation for mammograms acquired using the same film-screen combination and digital scanner.

Given the relative rates of normal/abnormal mammograms (discussed in section 2.2.7), in obtaining (and digitising) a training set of mammograms containing real masses from a screening population, it should be possible to obtain a much larger set of normal mammograms in which to synthesise masses.

However, if obtaining normal mammograms with the correct linear relationship

between grey level and attenuation is a problem, an alternative solution would be to apply the separation method shown in this chapter (and consequently the remainder of the synthesis method described in the following chapters) to mammograms that have explicitly been converted into a standard form. For example, Highnam et al. have defined a model of standard mammographic form (SMF) [56]. We recall that in section 3.2 we reviewed a synthesis method presented by Highnam and et al. [54]. In this method synthesis was performed within mammograms that had been converted into SMF. The actual synthesis method described was relatively simplistic (effectively cutting and pasting the attenuation patterns of real masses into normal mammograms). However, applying our synthesis method within such a framework might provide a successful method of simulating a mass in any normal mammogram regardless of how it was acquired. This may be an interesting area for further work.

As a final point on this topic, we note that henceforth, wherever we describe a normal mammogram in this thesis, we refer to one in our dataset, so that the consistent linear relationship between image intensity and attenuation hold.

With regard to the process of separating mass intensities, we have based our method on one earlier described by Caulkin [23]. In this method, unknown grey levels in each background image were estimated by fitting a thin-plate spline interpolant. The thin-plate spline was controlled by sampling grey levels from a set of landmark points placed in concentric rings outside the mass border. Prior to sampling the grey levels at the landmark points, the region was smoothed using a Gaussian kernel of standard deviation σ.

The positions of landmark points were controlled by three parameters n_1, n_2 and d, as

described in section 6.3. Values for σ, n_1, n_2 and d were determined experimentally (section 6.3.1). In section 6.3.1 we discussed the importance of choosing a suitable combination of n_1 and σ.

Here we note that there are many other ways in which landmark points could have been selected – for example, points could have been placed along radial profiles from the mass instead of on a rectangular grid. However, we have found that varying n_2 or d made little difference to the resulting background estimation error. This suggests that aside from choosing suitable combination of n_1 and σ, the position at which additional sampling points are located is not important.

An alternative method for selecting sample points would have been to apply a linear feature detector (for example one of the methods reviewed in section 3.3.1, or a method we develop in section 8.4.5). We could then ensure sample points are not placed on linear structures that can destabilise the thin-plate spline. As a result we would not need to apply such a high level of smoothing. This idea was not explored further, although it may be an area for further work.

Finally, in section 6.4 we described a method for transferring high-frequency components from the mass image to the background image. Our motivation for doing this was discussed in section 6.1. In the following chapter (section 7.5) we describe the effect this modification has on a model fitted to the set of mass images.

6.7 Summary

In this chapter we have described how a mammographic region containing a mass lesion can be represented as the pixel-by-pixel sum of a mass image and a background image. In this representation, the mass image contains grey levels corresponding to the pattern of increased attenuation associated with the central density of the lesion. The background image represents the appearance of all other breast tissue in the region.

We described a method for separating the mass image from the background for each region in our dataset. In performing this separation, we have effectively generated two new sets of image data, as depicted in Figure 6.11.

By separating the data in this way, we have decomposed our data into two logical groups that can be modelled in parallel.

In the mass images we have isolated the single object present in every region – the central density of each lesion. As we show in the next chapter we can encapsulate the variation in shape and texture of these objects by fitting a principal component based appearance model to the mass images.

Having removed the central density, the background images do not have a shared global appearance for which we can define a correspondence between each image. Instead they comprise a series of local interactions of texture and structure. By learning the patterns of these interactions we aim to understand how the appearance of breast tissue from which a mass has been separated differs from the appearance of breast tissue in a normal region. To emphasise this difference, in future chapter we will refer to background images as *mass backgrounds*. If the difference between mass background

and normal background can be learned, we can model the way in which a mass distorts normal structures in the breast.

Our work on constructing models of mass background appearance is described in chapter 8. However, in the following chapter we focus our attention on modelling the set of mass images.
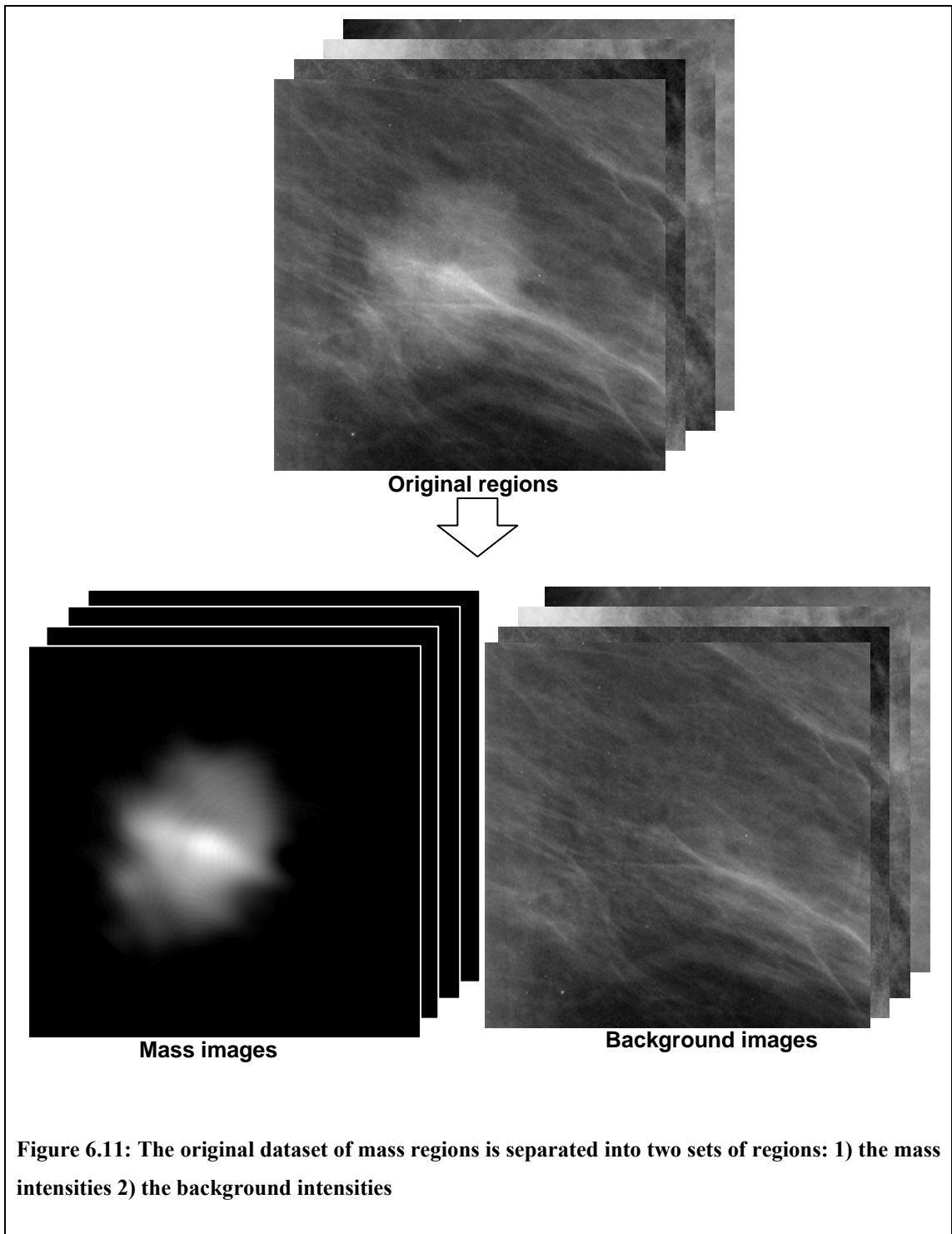
Figure 6.11: The original dataset of mass regions is separated into two sets of regions: 1) the mass intensities 2) the background intensities

Chapter 7 : Modelling the Appearance of Masses

7.1 Introduction

In chapter 6 we described a method for representing each region in our training data as the linear sum of two images. We defined the two images as the *mass image* and the *mass background* such that:

i)   The mass image contained grey levels corresponding to the pattern of increased attenuation associated with the central density of the lesion

ii)   The mass background represents the appearance of all other breast tissue in the region including spicules

In this chapter, the set of mass images becomes our sole data, as depicted in Figure 7.1. Because these are the only data we use, we no longer need to disambiguate between other uses of the term *mass*. As such, we drop the image suffix.

Our aim is to model the statistical variation in appearance of these data. The principal application of our model is to produce realistic synthetic masses, and thus in addition to describing the data, the model must be generative. This chapter describes how, using an existing model of mammographic masses as a base, we have developed a generative statistical model that optimally fits our data. A description of the remaining chapter sections is given below.

In the next section we describe in detail a model of mass appearance developed by Caulkin [23]. This was reviewed in section 3.2 and identified as a good starting point on which to base our method.
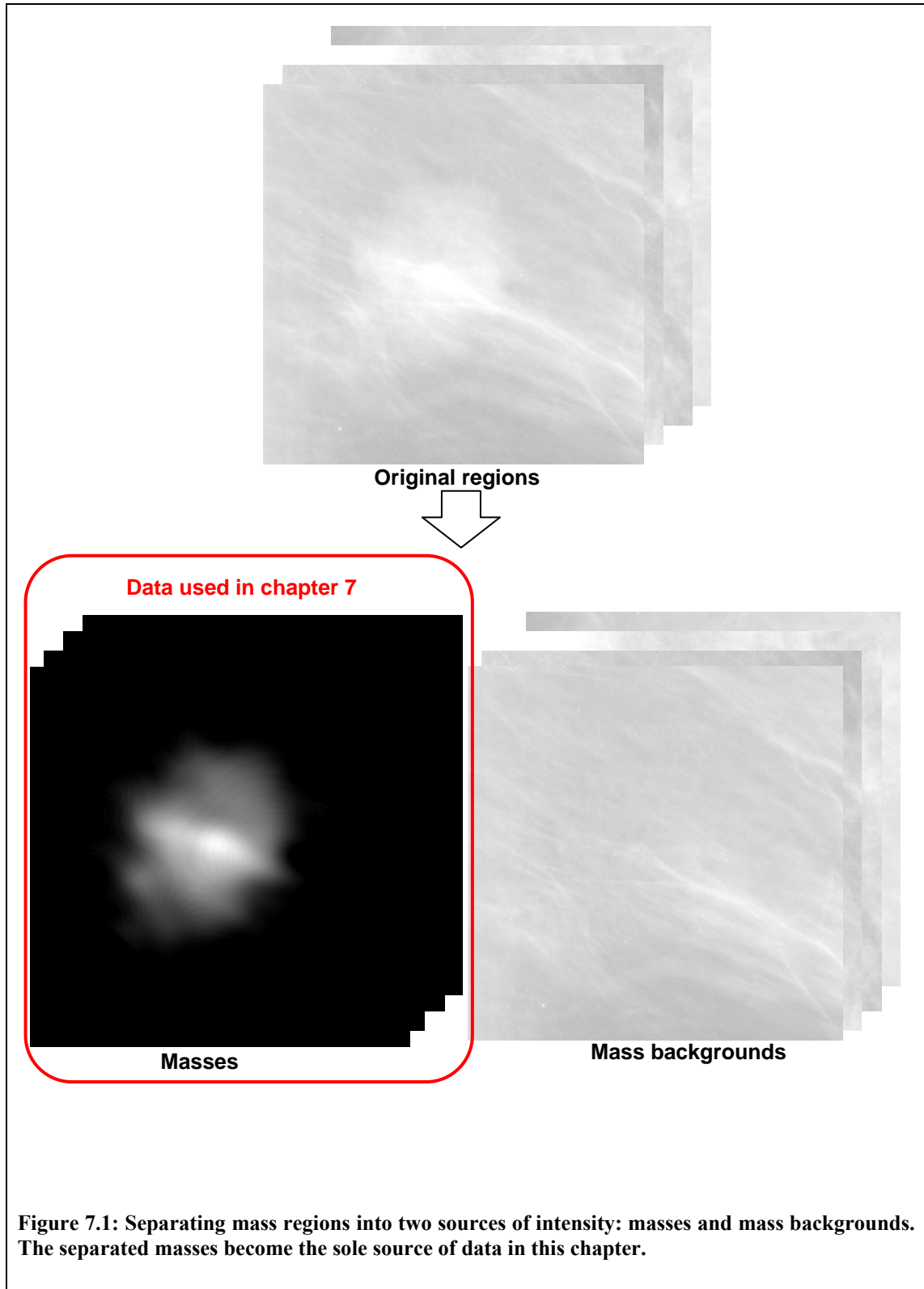
Figure 7.1: Separating mass regions into two sources of intensity: masses and mass backgrounds. The separated masses become the sole source of data in this chapter.

Having established the basic framework for a generative statistical model of our data, in section 7.3 we discuss the properties required of the model. In particular we look at how well the model fits the dataset of masses, and, importantly, show how we can measure model fit directly from the data.

In sections 7.4 to 7.6 we use the measures of model fit to show how three of the main components of Caulkin's model can be optimised resulting in a significant improvement in model fit. The resulting improvement to the full model of mass appearance is presented in section 7.7.

Finally, in sections 7.8 and 7.9, we show how the generative nature of the model can be exploited to produce realistic synthetic examples of masses. We qualitatively evaluate the appearance of such masses when superimposed on regions of normal breast tissue, and describe how this motivates the work on background modelling presented in the following chapter.

7.2 Caulkin's model of mass appearance

In section 4.2 we reviewed the class of statistical models developed as part of the active appearance model framework [27]. These models were designed to describe the variation in appearance of a training set of image objects. Here the term objects is necessarily vague – the only requirement of the object is that it has some characteristic shape and texture (where texture refers to the spatial distribution of grey level across the object) such that a correspondence can be established between each example of the object in the training data.

By design, the set of masses obtained using the separation process described in the last

chapter meet this requirement. In each mass, we can define a shape using the mass border (as annotated by a radiologist). This in turn defines mass texture as the pattern of grey levels across the shape. Approximately speaking, we expect these grey levels to be highest in the centre of the shape, tapering to a value of zero at the mass border. As such, there should be an approximate correspondence between each mass in the training data.

In our review of previous attempts to synthesise mammographic masses, we described a method developed by Caulkin [23] in which a statistical appearance model was applied to a similar set of masses that had been subtracted from mammographic backgrounds. Having fitted the model to the training data, Caulkin used the generative properties of the model to synthesise new masses that could be superimposed on real normal mammograms.

In section 3.2.1, we discussed the reasons why we think this was a good approach to modelling masses. As a result, we chose to use the model developed by Caulkin as the basis for our own synthesis method. Therefore in this section we describe the model fitted by Caulkin in more detail.

The three main components of the model were: a model of mass shape; a model of mass texture; and a combined model of shape, texture and mass size. We look at each component in turn below. In each case we describe the results of model fitting reported by Caulkin (in terms of the number of modes comprising each model) and also the results we have obtained applying the model to our own training data.

Shape

The mass border, as demarcated by a breast radiologist, defines the shape of each mass. To construct shape vectors, Caulkin first defined two reference points. These were the points of intersection of the mass border with a line connecting the nipple to the mass centroid, labelled as p_1 and $p_{n/2+1}$ in Figure 7.2. Shape vectors $\mathbf{x}_{shape} = (x_1, .., x_n, y_1, .., y_n)$ were then completed by placing n equally spaced points between the two reference points along each mass border (as depicted Figure 7.2). The number of points, $n = 150$, was chosen to be as small as could adequately describe the most complex of mass borders.

Figure 7.2: Caulkin's method for placing landmark points on a mass outline
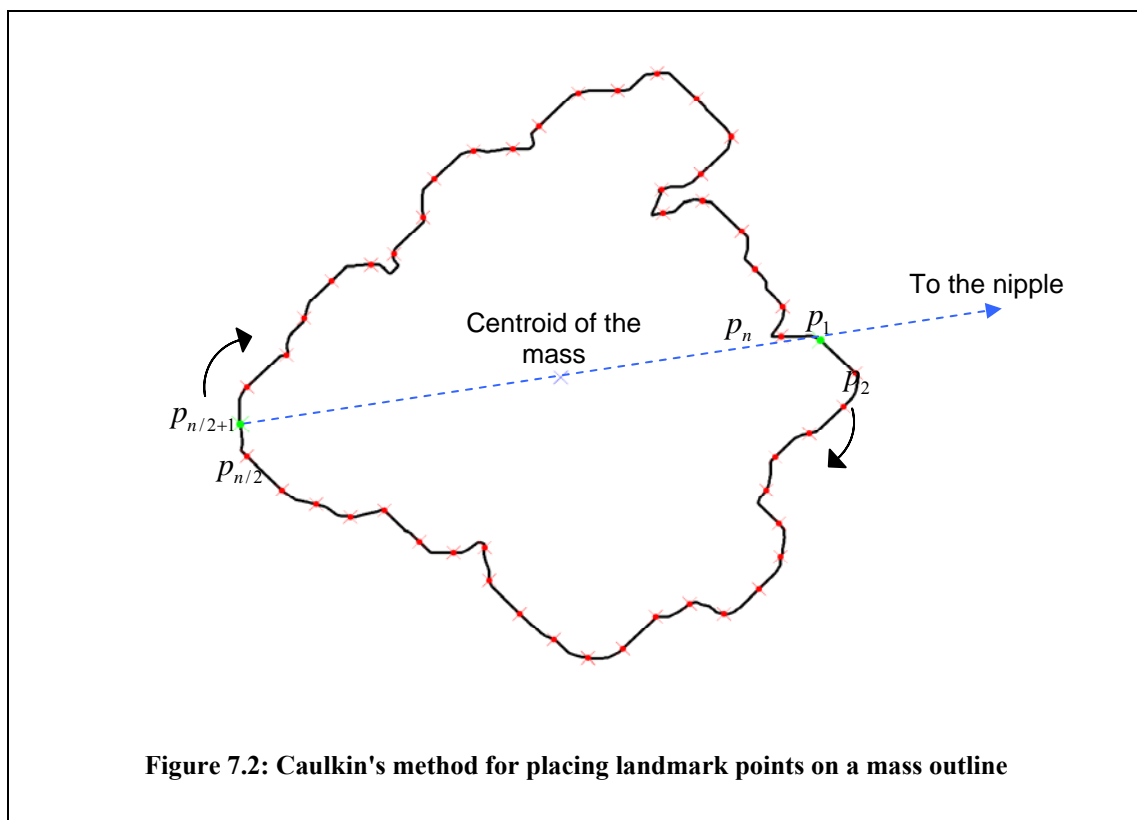
Having defined a set of landmark points on each mass border, a point distribution model (PDM) was built (see section 4.2.3) to capture shape variation. That is, the set of mass

shapes were aligned using Procrustes analysis [50], and a compact model of shape variation was built by applying principal component analysis such that for each mass $\mathbf{x}_{shape} \approx \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\mathbf{b}_{shape}$, where $\mathbf{P}_{shape}$ is a matrix of principal modes (with modes of negligible variance discarded) and $\mathbf{b}_{shape} = \mathbf{P}_{shape}^{T}(\mathbf{x}_{shape} - \overline{\mathbf{x}}_{shape})$ is the vector of model parameters stored for each shape. After discarding modes of negligible variance, Caulkin's original model comprised 23 modes describing 98% of the total shape variance. Applying this method to our data resulted in a shape model consisting of 25 modes. Given that our dataset contains nearly double the number of training examples (101 compared to 56), the fact that only 2 more modes of variance were required to describe our data suggests the shape model has captured the intrinsic shape characteristics of a mass boundary. That said, in section 7.4, we show the shape data can be further compacted to construct a model with improved generality and specificity.

Texture

Using a thin-plate spline interpolant [18] each mass border was warped to the shape mean (see appendix A.2.1), mapping the regions into a common coordinate frame. Shape free texture vectors were then extracted from each region. The size of the mean was chosen such that the texture vectors had 10,000 elements, again as a compromise between fidelity and compactness. As with shape, PCA is used to build a model of texture variation: $\mathbf{x}_{tex} \approx \overline{\mathbf{x}}_{tex} + \mathbf{P}_{tex}\mathbf{b}_{tex}$. Again keeping 98% of the total variance, Caulkin's original model yielded 37 modes. To give a direct comparison to Caulkin's method, we fitted the texture model to our data using subtracted mass images that had not had high-frequency background components removed (that is, masses obtained using the initial separation method defined in section 6.3 without further applying the methods

described in sections 6.4 and 6.5). This resulted in a model comprising 77 modes. We note that this large increase in the number of modes is most likely to be due to the increase in image resolution between our data (40μm per pixel) and that used by Caulkin (100μm per pixel). In section 7.5 we describe the effect of modelling the texture of masses obtained using the complete separation method described in chapter 6.

Combined appearance

A model of combined appearance was built by concatenating the weighted shape and texture parameters $\mathbf{b}_{shape}$ and $\mathbf{b}_{tex}$, together with the size parameter returned from Procrustes alignment[4], and applying PCA to compute $\mathbf{x}_{com} \approx \overline{\mathbf{x}}_{com} + \mathbf{P}_{com}\mathbf{b}_{com}$ as before. Note that the size parameter is not usually included in appearance models (and is usually treated separately as part of pose) but was included by Caulkin to account for any correlations between mass size and appearance. This correlation is likely to exist as, for example, larger masses would be expected to cause greater x-ray attenuation, and would thus correspond to higher grey levels in a digitised mammogram. The weighting accounts for differences in metric between the three components (shape, texture and size), and was computed as the inverse of the average variance per mode. Thus for each component, $W = k / \sum \lambda_i$ where k is the number of retained modes, and λ_i the variance along each mode. The number of significant modes retained in Caulkin's combined

---

[4] This parameter is usually described as the scale parameter in Procrustes alignment. However, we use the term size to make clear the relationship to the physical property of the mass; and also to avoid confusion when we discuss scaling the magnitudes of model parameters in section 7.6.

model was not recorded; however applying the method to our data constructed a combined model containing 65 modes (using the same texture model as before).

Having learned the distributions of mass appearance, new examples were generated by randomly sampling from the model as follows:

1) Sample a new combined appearance vector $\mathbf{b}'_{com}$, and compute

$$\mathbf{x}'_{com} = \overline{\mathbf{x}}_{com} + \mathbf{P}_{com}\mathbf{b}'_{com}$$

2) Separate $\mathbf{x}'_c$ and invert the weighting to form $\mathbf{b}'_{shape}$, $\mathbf{b}'_{tex}$ and $\mathbf{b}'_{size}$

3) Compute $\mathbf{x}'_{shape} = \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\mathbf{b}'_{shape}$ etc.

4) Map the new texture $\mathbf{x}'_{tex}$ to the new shape $\mathbf{x}'_{shape}$, scaled by $\mathbf{x}'_{size}$, to produce the new synthetic mass S_M

The process above highlights a major advantage of using the appearance model framework as a basis for mass synthesis. Having built the models, no user parameters are needed to generate new examples. The appearance of a new mass is completely determined by a vector that can be randomly (and thus automatically) sampled from the model distribution. As a result, the process can be repeated to produce limitless examples of unique masses that match the appearance properties of real lesions.

However, as we show in sections 7.4 to 7.7, we do not think the model described by Caulkin adequately fits our data. As a result the model distributions learnt from the data are not truly representative of the variation in appearance of the global set of masses. This, in turn, would mean that synthetic masses generated from the model could not be

used in place of real data. To determine how we can improve model fit, in the next section we expand on the desirable properties of this form of statistical model and describe how we can measure model performance.

7.3 Measuring model errors

In this section we define a method for quantifying how well a statistical model fits a set of training data. To do this, we first consider what properties we require a model to have. Throughout we reference the particular application of modelling mammographic masses, although we note the properties and measures we discuss are applicable to any class of image object.

7.3.1 Model properties

From the description of how a mass appearance model is constructed (section 7.2), it should be apparent that a statistical model provides a compact representation of our training data. Where previously, each mass was defined by several hundred thousand pixels, it is now fully described by its vector of combined model parameters $\mathbf{b}_{com}$. Thus the complete set of training data is defined by the set of vectors $\{\mathbf{b}_{com}\}$, and the matrix of principal modes of variation for each component, $\mathbf{P}_{shape}$, $\mathbf{P}_{tex}$ and $\mathbf{P}_{com}$.

However, in using the model to generate synthetic masses, we make two important assumptions about the model that extend its use beyond simply representing the data we already have. Firstly, we assume the model can describe any mass appearance from the global population. Secondly, we assume that anything that doesn't match the appearance of a mass cannot be represented in the model. These two model properties are usually known as generality and specificity, respectively. If we can measure the

generality and specificity of a given model, we can compare the performance of one model relative to another, enabling us to find the optimal model for our data.

7.3.2 Specificity

We assume specificity is directly linked to the compactness of the model, which itself can be measured by the total variance in the model. With the mean mass appearance at the centre of the model, the variance along each mode defines how the probability density reduces moving away from the mean. As the variance increases, the probability is spread further from the mean, increasing the likelihood that an object whose appearance does not match that of a mass has non-negligible probability in the model.

One way of reducing the amount of variance stored in a model is to reduce the number of modes retained. Taken to the extreme, a model with no modes only allows the mean mass to be generated and is therefore guaranteed to be specific. Of course such a model suffers from a total lack of generality, and is unable to explain any of the variance in the population of masses (or indeed in the training sample). Thus in constructing each of the component models, we choose to retain a fixed amount of the total variance, selecting the number modes required to match this threshold. For this work 98% of the total variance was chosen as a sufficient compromise between specificity and generality. As a result, our aim in improving the specificity of the models is to reduce the total variance in the data, prior to discarding modes of negligible variance. An alternative measure would be to try and use an objective function to measure model specificity as was described by Cootes et al. [31].

7.3.3   Generality and the residual errors for an unseen mass

To get a measure of model generality, we assume we have built shape, texture and combined appearance models from a training set of masses. In addition, we assume we have a mass not used in building the models, which we refer to as an unseen mass. The more general the full combined appearance model, the better it will be able to represent the unseen mass. To measure this, we can transform the unseen mass into a point in model space, and then reconstruct a representation of the mass from the model.

We can then compute the residual error of the unseen mass as the difference between its original form and its reconstruction in the model. Below, we give a fuller description of this process. In addition, the error calculation is depicted diagrammatically in Figure 7.3.

We call the unseen mass R_M , and assume this is in the same form as those used to train the model (i.e. with a continuous border outlined and the original position relative to the nipple known). This allows us to obtain vectors $\mathbf{x}_{shape}$ and $\mathbf{x}_{size}$, and, after warping to the shape-free space, $\mathbf{x}_{tex}$. Using the principal modes of the shape model, we can calculate the model parameters $\mathbf{b}_{shape} = \mathbf{P}_{shape}^{T} \left(\mathbf{x}_{shape} - \overline{\mathbf{x}}_{shape} \right)$, and similarly, compute $\mathbf{b}_{tex}$ and $\mathbf{b}_{size}$ (step 1 of Figure 7.3). Now we regenerate the shape, texture and size vectors as approximated by the individual models: $\mathbf{x}_{shape}^{'} = \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\mathbf{b}_{shape}$ etc. as depicted by step 2 in Figure 7.3. From here the full mass region $R_M^{'}$ can be regenerated. The root mean squared error between the mass regions R_M and $R_M^{'}$ and/or the individual vectors can be used to quantify how well the individual models represent the new mass.

Figure 7.3 Calculating the residual errors in the model representation of a mass R_M. For the error to be unbiased R_M must not have been used in the construction of the model
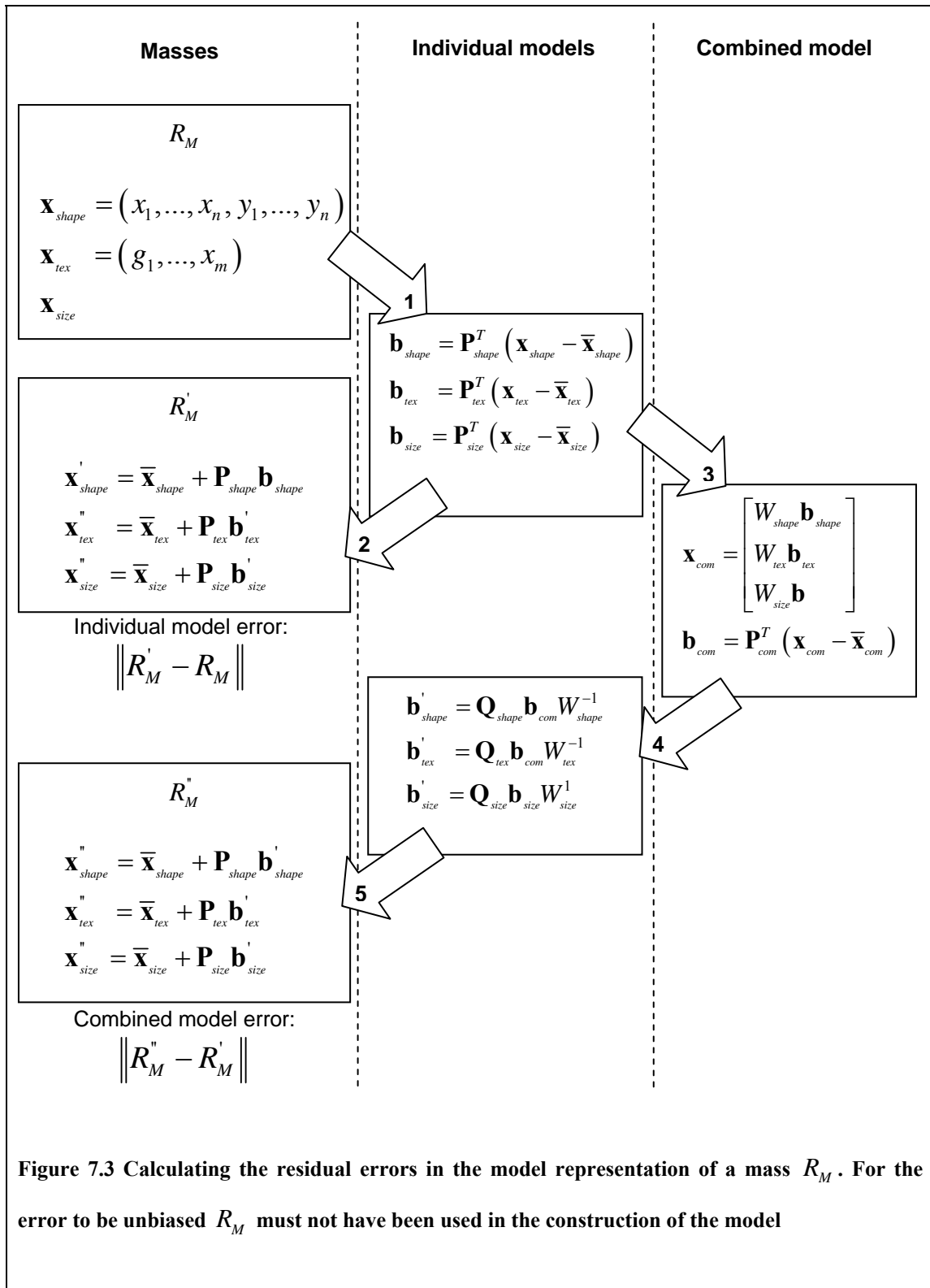
Of course, the individual models are further combined into the full appearance model, and so returning to $\mathbf{b}_{shape}$, $\mathbf{b}_{tex}$ and $\mathbf{b}_{size}$, we build the combined vector $\mathbf{x}_{com}$ and compute $\mathbf{b}_{com} = \mathbf{P}_{com}^{T} \left(\mathbf{x}_{com} - \overline{\mathbf{x}}_{com} \right)$ as depicted in step 3 of Figure 7.3. This is the representation of the original mass in the full model space and, following the reconstruction described in section 7.2, can be used to generate $R_{M}^{''}$ (steps 4 and 5 of Figure 7.3). As with the approximation from the individual models, we can use the root mean squared error between R_{M} and $R_{M}^{''}$ to quantify how well the full appearance model has described the original mass. It is also useful to consider the difference between $R_{M}^{'}$ and $R_{M}^{''}$. This shows the added error induced by the combining the individual models in the full appearance model, and as we see in section 7.6, will be useful when we optimise the weighting of the shape, texture and size parameters in the combined appearance vectors.

7.3.4   Leave-one-out testing

The process above describes how we calculate residual errors for a single mass represented by the model. For this error to be unbiased the mass should not have been used to build the model. To have a useful measure of the performance of the model, we need to calculate this error for a sample of masses large enough to be representative of the global population. However, we only have a limited number of masses available, and require all of these to construct the model. To satisfy both needs, we employ leave-one-out testing to calculate residual errors for each of the N masses in the dataset. The set of errors can then be used to quantify how well the model can describe unseen masses in the global population, and so provides our measure of the model's generality. The leave-one-out testing algorithm we apply is summarised below:

for $i = 1$ to N

       construct model from the set of masses $\left\{ R_{M,j} \mid j = 1 \ldots N, j \neq i \right\}$

       compute residual errors for mass $R_{M,j}$

end

Algorithm 7.1: Leave-one-out computation of residual errors

Such a scheme works on the assumption that the model constructed from any N-1 subset of masses is negligibly different from the model built from all N masses. This assumption holds true if we believe our set of masses sufficiently represents the global population, which itself is a condition of constructing the models. Thus using leave-one-out testing does not require any further assumptions about data than had been made already.

Having described how we can measure the desired properties of specificity and generality of a model, in the next two sections we show how we can construct an optimal model of mass appearance given a set of masses (complete with annotated borders).

7.4 Optimising the shape model

The key to building a compact representation of mass shape is to establish a set of dense correspondences between the annotated borders. Finding such a set may be problematic due to the lack of features present in each mass border. In section 7.2, we described how Caulkin defined two common feature points on each mass shape, and used these to determine the location of a further set landmark points. The reference points were constructed by extending a line from the breast nipple through the centroid of the mass, before setting the point of origin p_1 as the intersection of this line with the mass

border nearest to the nipple, and a halfway point $p_{n/2+1}$ as the point of intersection furthest from the nipple, as was shown in Figure 7.2. Fixing the origin in this manner aimed to introduce a correspondence based on the theory that masses form in ducts approximately radiating from the nipple, as in the simplified depiction of masses in Figure 7.4.

If all the masses in the dataset followed such an alignment to the nipple, then after adding in the remaining equally spaced border points and applying Procrustes analysis, we would expect to obtain a compact shape model. However, analysis of our data suggests no evidence for masses radiating from the nipple. A more realistic diagram of mass alignment is presented in Figure 7.5. In this case, the two shapes will not correspond after Procrustes analysis, so in fact fixing the origin simply introduces rotational variance in the model. That is, objects that have the same shape but differing orientations are described differently in the model. This has two negative effects: firstly, and most obviously, the model will not be optimally compact, as additional model parameters are required to describe the differing orientations of the same shape. Secondly, subtle variations may be discarded from the model because the retained modes describe a fixed amount of the total shape variation, and a significant amount of this would be due to the rotational variance. This would reduce the model's ability to represent masses in the dataset, and more generally, those in the global population.
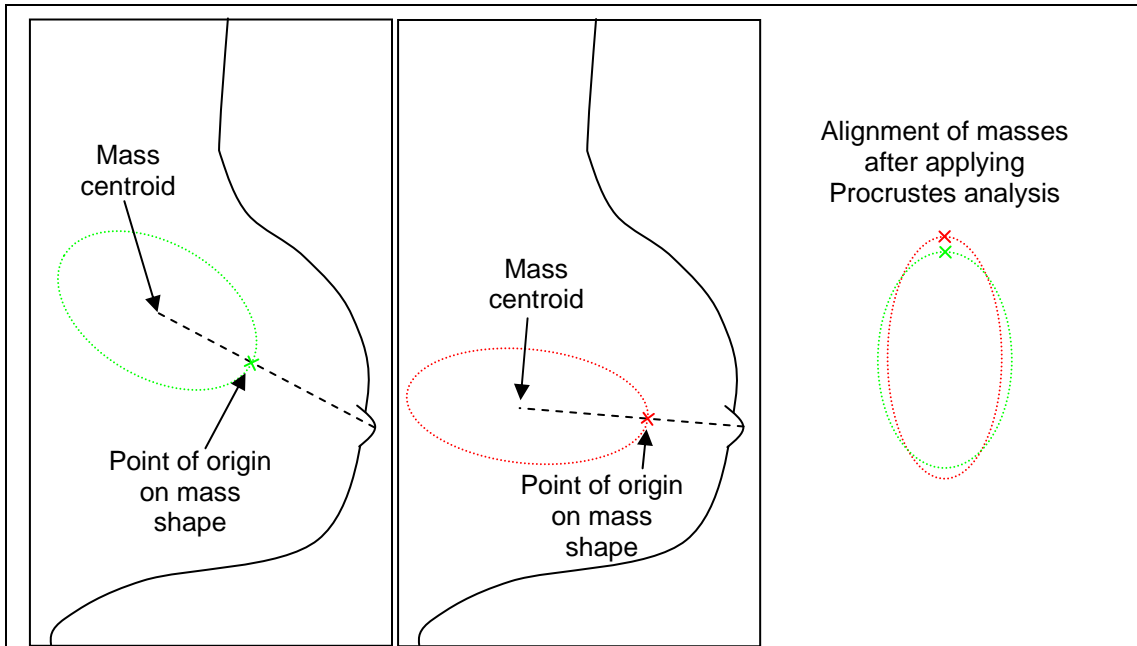
Figure 7.4: Depiction of two mammographic masses oriented towards nipple. With a fixed point of origin, the mass shapes form a compact alignment after Procrustes analysis
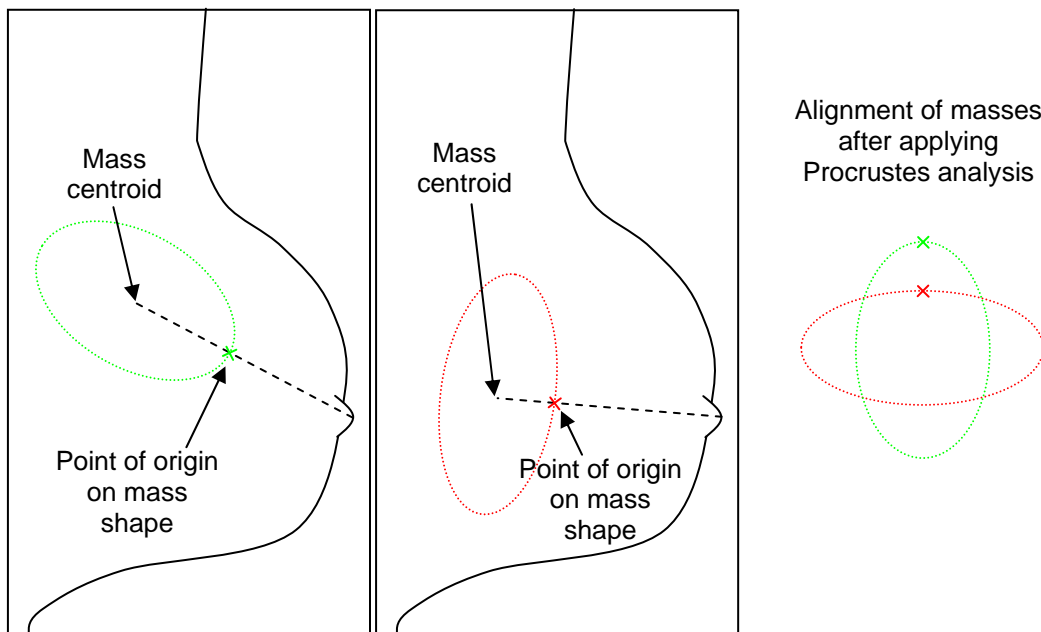
Figure 7.5: Depiction of two mammographic masses with random alignment. By fixing a point of origin, Procrustes alignment will not align the mass shapes despite their similarity

7.4.1   Optimising the point of origin on mass shapes

A solution to remove the rotational variances from the model is to allow the optimal

origin of each mass border to be selected during Procrustes alignment. This requires

modifying the iterative Procrustes algorithm. The new algorithm is given in detail below

(note as all vectors in the algorithm refer to shape, we will drop the shape suffix for

convenience, for example $\overline{\mathbf{x}}$ instead of $\overline{\mathbf{x}}_{shape}$).

Select a random seed shape as the target $\mathbf{x}_t = \mathbf{x}_i$ for some $i = 1 \ldots N$

    for each shape $i = 1$ to N

        for each point $j = 1$ to n

            permute $\mathbf{x}_i$ so that the j-th element is now the first

            align $\mathbf{x}_i$ to $\mathbf{x}_t$ to find optimal rotation, scaling and translation

            record the distance from the aligned shape to the target: $\quad d_j = \left\| \mathbf{x}_t - \mathbf{x}_i \right\|$

        end

    set $D_i = \min(d_j \mid j = 1, \ldots, n)$

    and set the origin of $\mathbf{x}_i$ as the associated j

    end

    Compute the mean of the differences across the set: $D = \dfrac{1}{N} \sum_{i=1}^{N} D_i$

    Compute the mean shape: $\overline{\mathbf{x}} = \dfrac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$

    Set $D_{old} = D$, $D_{new} = 0$ and $\mathbf{x}_t = \overline{\mathbf{x}}$

while $D_{old} > D_{new}$

    repeat

    set $D_{old} = D_{new}$, $D_{new} = D$ and $\mathbf{x}_t = \overline{\mathbf{x}}$

end

Algorithm 7.2: Procrustes alignment of shape, with optimisation of shape origin

To check the algorithm is robust to the choice of initial seed shape, the algorithm is repeated for a random selection of seed shapes. In each case the difference between the resulting mean shapes is calculated to establish whether choice of seed affects the algorithm outcome. In our experiments there was negligible difference between the

mean shapes obtained from any of the randomly selected seed shapes, satisfying the desired robustness.

7.4.2   Applying MDL optimisation to mass shapes

Whilst allowing the origin of each mass border to vary removes much of the rotational variance in the data, there may still be a better set of shape correspondences we can obtain. In section 4.2.3, we reviewed a method developed by Davies et al. [34] for computing the optimal set of correspondences for a training set of shapes. In this method, the locations of points on the continuous shape borders are allowed to vary (as opposed to sampling equally spaced points) during the alignment optimisation. The objective function used in the optimisation is a measure of compactness termed the Minimum-Description Length (MDL) that quantifies the amount of data that is required to fully describe the set of shapes given their parameterisation under the model.

In addition to allowing the origin of each shape to be optimised as described in the previous section, we applied the MDL optimisation algorithm to our set of mass shapes to see if further improvements in correspondence could be made. The results of applying each alignment method are given in the following section.

As noted in section 7.2, the spatial resolution of digitised mammograms in our data was greater than that used in Caulkin's original experiments. For this reason, we required an increase in n, the number of points sampled along the demarcated border of each mass. The mean length in pixels of the masses in our dataset was 1560 (compared to 700 in Caulkin's work) and thus a value of $n = 400$ was chosen (compared to $n = 150$ in Caulkin's work).

7.4.3   Results

After applying the standard Procrustes alignment algorithm to the set of 101 mass shapes, the total variance was 2.37×10^5. As noted in 7.3.2, the total variance of the aligned data provides a measure of how specific the resulting shape model will be, with lower variance resulting in increased specificity.

To measure model generality, the residual shape error $\left| \mathbf{x}'_{shape} - \mathbf{x}_{shape} \right|_{RMS}$ was computed for each mass using leave-one-out testing (see sections 7.3.3 and 7.3.4 for more details), and the mean of these errors taken. For the standard Procrustes algorithm, the mean generalisation error was 3.78.

Applying the origin optimisation alignment algorithm described in 7.4.1 to our dataset reduced the total shape variance by 48% to 1.24×10^5. The mean generalisation error was also reduced to 3.64.

Applying MDL optimisation in addition to allowing the origin of each shape to vary further reduced the total shape variance to 1.21×10^5. This was associated with a reduction in model generalisation error to 3.60. These data are summarised in Table 7.1.

Table 7.1: Model generalisation error and variance for methods of aligning mass shape

	Variance	Model generalisation error
Procrustes analysis	2.37×10^5	3.78
Procrustes analysis using optimised origins	1.24×10^5	3.64
MDL optimisation	1.21×10^5	3.60

As a result of optimising the origin of each shape and applying MDL optimisation, we have greatly reduced the total variance in the shape model, making the model more specific and thus less likely to generate unrealistic shapes when used to synthesise new masses. Further, because the model is not required to describe the rotational variances present in the mass shapes aligned using the standard Procrustes algorithm, the model modes are better able to describe subtle variations in the mass shape. This allows the model to be more general, as seen in the reduced mean generalisation error. To test the statistical significance of the reduction in model error, we applied the Wilcoxon signed-rank test to compare the three methods. The results of the test show that allowing the origin of each shape to optimised during Procrustes alignment produced a significant reduction in errors ($p < 0.0001$). Subsequently applying MDL optimisation produced a further statistically significant reduction model error ($p < 0.0001$).

7.5 Modelling mass texture

With regards to modelling mass texture, the only change we have made to the method described by Caulkin is to increase the length of texture vector extracted, from 10,000 to 50,000 pixels (controlled by scaling the size of the mean shape used to warp the masses into a shape-free coordinate frame). As with mass shape, this was to accommodate the increase in mass size within each image as a result of the increased spatial resolution of the mammograms in our data.

However, as a result of changing the process by which mass intensities were separated, as described in chapter 6, the data modelled is very different. Most importantly, we have transferred high-frequency components that previously were included with mass intensities into the background region. To recap, we observe that the majority of such

structures do not belong to the mass (that is, their intensity can be attributed to x-ray attenuation through normal structures in the breast superimposed in the mass region). If these structures were included in a synthetic mass, then when we added the mass to a normal mammogram we would be simulating a localised increase in attenuation above that corresponding solely to the presence of a mass.

Where high frequency components do belong to the mass (principally, spicules), these are better accommodated in the background regions because:

i) They do not form a consistent set of correspondences that can be captured using the form of appearance model described in this chapter. Analogous to the discussion on rotational variance (section 7.4), this reduces model compactness (because extra modes of variance will be needed to describe the individual structures) and may cause subtle variations in texture that are characteristic of the central density to be discarded from the model (because the high frequency components have used up the total variance retained).

ii) By including them in the background regions we have a set of data from which the interaction between mass spicules and normal tissue can be modelled. As we show in the following chapter, we design models to capture and synthesise high-frequency texture components that do not have a consistent global correspondence across an image

In this section we give evidence to support the first point. Using the optimised shape model described in the previous subsection, we built texture models of our set of masses both with and without high frequency components included. We then computed the

mean texture generalisation error $\left| \mathbf{x}'_{tex} - \mathbf{x}_{tex} \right|_{RMS}$ for each mass using leave-one-out testing. The errors for masses with high frequency components removed were dramatically reduced, the mean errors across the set being 3.80 and 1.73 respectively (Wilcoxon signed-rank $p < 0.0001$). We also note the number of modes required to maintain 98% of texture variance was reduced from 77 to 28.

7.6 Weighting shape, texture and size

The final step in building the full model of mass appearance is to combine the model parameters of shape, texture and size for each mass before further applying PCA. As noted previously, the inclusion of the size parameter is unusual in appearance models, but is important in this instance to account for the expected correlation between the size of a mass and its shape and, in particular, texture.

However, including size makes correctly weighting the individual components in the combined appearance vectors crucial. Prior to weighting, the size parameters $\{\mathbf{b}_{size}\}$ of our dataset of masses have an order approximately 10^3 smaller than the texture parameters $\{\mathbf{b}_{tex}\}$. Thus, if unweighted, real size variation would be equivalent to random texture noise and would be discarded when removing negligible modes from combined model.

We recall that Caulkin used the inverse of the mean variance per mode in each component. That is, for each of shape, texture and size, $W = k \Big/ \sum_{i=1}^{k} \lambda_i$ where k is the number of retained modes, and λ_i the variance along each mode (by construction $k = 1$ for size, and thus W_{size} is just the inverse of the size variance). In the following sections,

we show why we do not think these are a good of choice weights and present three alternative methods for computing component weights. In section 7.6.4, we quantitatively evaluate each method and thus choose the best method to use in our model.

7.6.1   Equalising mode variance

We presume that dividing each set of model parameters by the mean variance for that component was an attempt to equalise the variance of shape, texture and size in the full model. This appears to be a sensible idea, however because variance has the order of the square of the parameters, equality will not be achieved by dividing by the variance. As a simple example, assume the texture parameters $\{\mathbf{b}_{tex}\}$ have, on average, a magnitude A times larger than the size parameters $\{\mathbf{b}_{size}\}$. Then $W_{size} = A^2 W_{tex}$, and so, after weighting, we have simply reversed the problem such that $\{W_{size}\mathbf{b}_{size}\}$ now have, on average, magnitudes A times larger than $\{W_{tex}\mathbf{b}_{tex}\}$.

In the previous section we stated that in fact A is approximately 10^3 and so computing weights proportional to the inverse of each components total variance results in the size parameter of the combined appearance vectors being 1000 times larger than the texture parameters; and by a similar calculation, 200 times larger than the shape parameters. As we see in 7.6.4, this weighting does not produce a model that fits the data well.

The solution to this problem should be clear: if we are trying to equalise the mean variance of shape, texture and size in the model, we should be dividing each component by the mean standard deviation. Thus for each component $W = k \Big/ \sum_{i=1}^{k} \sqrt{\lambda_i}$. We will show

in section 7.6.4 that using mean standard deviation does indeed improve on weights calculated using the mean of the variances. However in sections 7.6.2 and 7.6.3 we describe two further methods for computing weights that produce models that describe mass appearance with significantly lower residual error.

7.6.2   Equalising texture displacement

Having shown how to equalise the total variance of each component in the combined model, we now consider whether this is in fact desirable. It may be that shape, texture and size have different magnitudes of effect on mass appearance and thus should have different magnitudes of representation in the combined model.

A better method for computing the model weights would be to quantify the magnitude of the effect of each component on mass appearance. Such a method was used in the original active appearance model work by Cootes et al. [27]. Below we show how this method is applied to our data.

Consider any mass R_M from our dataset, comprising shape, texture and size vectors $\mathbf{x}_{shape}$, $\mathbf{x}_{tex}$ and $\mathbf{x}_{size}$ respectively, with texture model parameters $\mathbf{b}_{tex} = \left[b_1^{tex}, ..., b_{k_{tex}}^{tex} \right]$. Due to the linearity of the model, if we displace any mode in $\mathbf{b}_{tex}$ by a unit then we cause a unit displacement in the model approximation to $\mathbf{x}_{tex}$. That is, if we define

$$\mathbf{b}_{tex}^{'} = \left[b_1^{tex}, ..., b_i^{tex} + a, ..., b_{k_{tex}}^{tex} \right], \quad 1 \le i \le k_{tex}$$
$$\mathbf{x}_{tex}^{'} = \overline{\mathbf{x}}_{tex} + \mathbf{P}_{tex} \mathbf{b}_{tex}$$

and

$$\mathbf{x}_{tex}^{''} = \overline{\mathbf{x}}_{tex} + \mathbf{P}_{tex}\mathbf{b}_{tex}^{'}$$

then

$$\left\| \mathbf{x}_{tex}^{''} - \mathbf{x}_{tex}^{'} \right\| = a$$

Since the shape and size vectors determine the texture vector sampled, we can estimate the displacement of the sampled texture vector caused by systematically displacing each mode of the shape and size model parameters, and subsequently calculate the weight needed to cause a unitary shift in the texture vector. We apply this process to the full set of masses to approximate the mean weight needed so that a unitary displacement along any mode of the shape, texture or size model parameters results in an equivalent displacement in the resampled texture vector. The algorithm for calculating the shape weight is given below:

for each mass $R_{M,i}$

    calculate $\mathbf{x}'_{shape} = \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\,\mathbf{b}_{shape}$

    resample $\mathbf{x}'_{tex}$ given $\mathbf{x}'_{shape}$ scaled by $\mathbf{x}_{size}$

        for each shape mode $j = 1...k_{shape}$

            displace parameter j by a constant

            set $\mathbf{b}'_{shape} = \left[b_1^{shape}, ..., b_j^{shape} + a, ..., b_{k_{shape}}^{shape} \right]$;

            calculate $\mathbf{x}''_{shape} = \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\,\mathbf{b}'_{shape}$;

            resample $\mathbf{x}''_{tex}$ given $\mathbf{x}''_{shape}$ scaled by $\mathbf{x}_{size}$;

            define $d_{j,i} = \dfrac{1}{a}\left\| \mathbf{x}''_{tex} - \mathbf{x}'_{tex} \right\|$;

        end

    end

    set $D_j = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} d_{j,i}$

    set $W_{shape} = \dfrac{1}{k_{shape}}\displaystyle\sum_{j=1}^{k_{shape}} Dj$

Algorithm 7.3: Computing the effect of shape mode variance on shape-normalised mass texture

Note that we could use the texture displacement per unit shape displacement for each

mode $D_j = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} d_{j,i}$ to define a weight for each individual shape mode as in the method

described in by Cootes. However, the per unit displacement was found to be

approximately constant across the modes, and computing the resulting model error as

described in section 7.3.4, no significant difference was found between using individual mode weights and a combined weight.

To calculate W_{size} we apply a similar algorithm to above, with the obvious change that the re-sampled textured vectors are calculated given the original shape vector multiplied by a displaced size parameter, as opposed to a displaced shape vector multiplied by the original size parameter. By construction, $W_{tex} = 1$.

7.6.3 Directly optimising weights

The methods described in sections 7.6.1 and 7.6.2 attempt to calculate weights that make the individual components theoretically commensurate in the combined model. However, an alternative approach to the problem is to see the weights as variables to be optimised given some objective function.

For a suitable error function, we refer back to the method for quantifying model generality in 7.3. Using leave-one-out testing, for each mass R_M we showed how to calculate R_M' and R_M'' as the approximations of the mass given the individual models and the full appearance model respectively. The root mean squared error between R_M'' and R_M' is the added error caused only by the approximation of the combined model. If we fix all aspects of constructing the combined model except the component weights, then any change in $\left| R_M'' - R_M' \right|_{RMS}$ is associated with the weights. Thus, in combination with leave-one-out testing, we can compute a mean generalisation error for the dataset associated with any given set of weights, as described in Algorithm 7.4.

---

for each $R_{M,i}$

    construct model from the set of masses $\left\{ R_{M,j} \mid j=1 \ldots N, j \neq i \right\}$

    compute approximations $R'_{M,i}$ and $R''_{M,i}$ for mass $R_{M,i}$

    compute $e_{Wi} = \left| R''_{M,i} - R'_{M,i} \right|_{RMS}$

end

set $E_w = \dfrac{1}{N} \sum_{i=1}^{N} e_{Wi}$

Algorithm 7.4: Computing the mean residual error associated with component weighting in the combined model for a set of masses using leave-one-out testing

Having defined an objective function, we consider the variables we seek to optimise. Given the linear nature of the models, it is only the relative magnitudes of the weights that are important, and so without loss of generality, we can fix $W_{size}=1$, and optimise W_{shape} and W_{tex}. In addition, we note that the sign of the weights is irrelevant and so fix the constraints that W_{shape}, $W_{tex} > 0$. Computing a set of weights is now formulated as a constrained non-linear optimisation over two variables, minimising the objective function E_w. All that remains is to choose an optimisation algorithm. For this we use the sequential quadratic-programming (SQP) method implemented in the Matlab® toolbox, and described by Schittkowski [104]. This has been shown to be one of the most accurate and efficient algorithms over a large range of problems [104].

To narrow the range over which we apply the optimisation, we first computed E_w for a coarse grid of values W_{shape} and W_{tex}. W_{shape} was varied from 0 to 5×10^{-3} at intervals of 1×10^{-3}, whilst W_{tex} was varied from 0 to 1×10^{-3} at intervals of 0.2×10^{-3}. The resulting error map is shown in Figure 7.6. This confirms that E_w behaves suitably as an

objective function, and provides an approximate region to start the SQP algorithm. Of the range of values initially tested, $\left(W_{shape,} W_{tex} \right) = \left(2 \times 10^{-3}, 0.2 \times 10^{-3} \right)$ corresponded to the lowest value of E_w. Four pairs of $\left(W_{shape,} W_{tex} \right)$ values were chosen symmetrically about this minima $\left(\left\{ (1.5, 0.1), (2.5, 0.1), (1.5, 0.3), (2.5, 0.3) \right\} \times 10^{-3} \right)$ and used in turn to initialise an SQP optimisation. To aid computational efficiency, only a randomly selected subset of the full training data (25 out of 101 masses) was used during each optimisation. However, when each optimisation had converged, the final weights were used to compute E_w for the full set of data. The weights corresponding to the lowest overall error were selected.
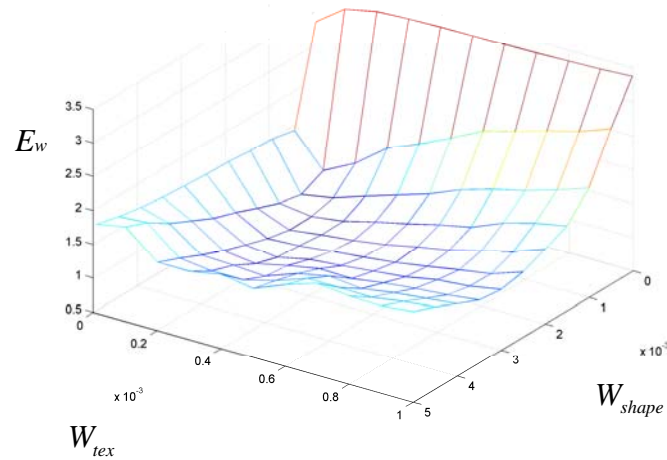
Figure 7.6: Map of mean generalisation error associated with component weights (E_w) for varying combinations of W_{shape} and W_{tex} ($W_{size} = 1$)

7.6.4   Results

In the previous section we showed how we can calculate the residual error in

approximating a mass as a result of combining the shape and texture models into a single model of appearance ($e_j = \left| R_{M,i}^{''} - R_{M,i}^{'} \right|_{RMS}$). By keeping all other aspects of model building the same, we were able to associate this error with changes to the weights applied to the shape, texture and size parameters so that they were commensurate. Again employing leave-one-out testing we used this error to define a mean generalisation error for the dataset, given a particular set of component weights (E_w).

We have discussed four methods for selecting component weights W_{shape}, W_{tex} and W_{size}. These were:

1) Computing each weight as the mean variance per mode (as used originally by Caulkin)

2) Computing each weight as the mean standard deviation per mode so that the total variance in shape, texture and size parameters were equal in the combined model

3) Applying the texture displacement method described in section 7.6.2

4) Directly optimising E_w as described in section 7.6.3

We used E_w to compare the performance of each of the four methods. The individual models used during the evaluation were the optimised shape and texture models described in 7.4 and 7.5 respectively, built from the dataset of 101 mass regions. To recap, this produced a shape model comprising 32 modes and a texture model of 28 modes. The parameters along each shape and texture mode, complete with the single size parameters, produced 61-element combined appearance vectors for each mass.

Boxplots of the set of errors ($\{e_j\}$ in the notation above) for each method are shown in

Figure 7.7, whilst the mean errors (E_w) are summarised in Table 7.2.

Table 7.2: Comparison of error E_w for different methods of selecting component weights

	E_w
1. Mean variance per mode	3.10
2. Mean standard deviation per mode	1.76
3. Texture displacement	1.00
4. Direct optimisation	0.82

Method 1 was included to enable a comparison with Caulkin's model. However there is little theoretical justification for dividing each component by the mean variance per mode and, as expected, it produced the highest errors. Dividing instead by the mean standard deviation, as in method 2, equalises the variance of the pooled set of parameters for each component. This assumes shape, texture and size affect appearance equally, and whilst producing lower errors than method 1, method 2 performed poorly relative compared to methods 3 and 4.

By construction, we would expect direct optimisation to generate the lowest errors[5], and indeed this was the case. However we were interested to see whether computing weights using the texture displacement method produced similarly low errors. In fact the

---

[5] Although we recall that each optimisation was trained only on a subset of the full data and thus wasn't guaranteed to produce a lower E_w than the other three methods

reduction in error using method 4 compared to method 3 is statistically significant ($p <$ 0.0001). As a result, the combined appearance model was constructed using the weights computed using the direct optimisation method.

Figure 7.7: Boxplots of model generalisation error for different component weighting methods
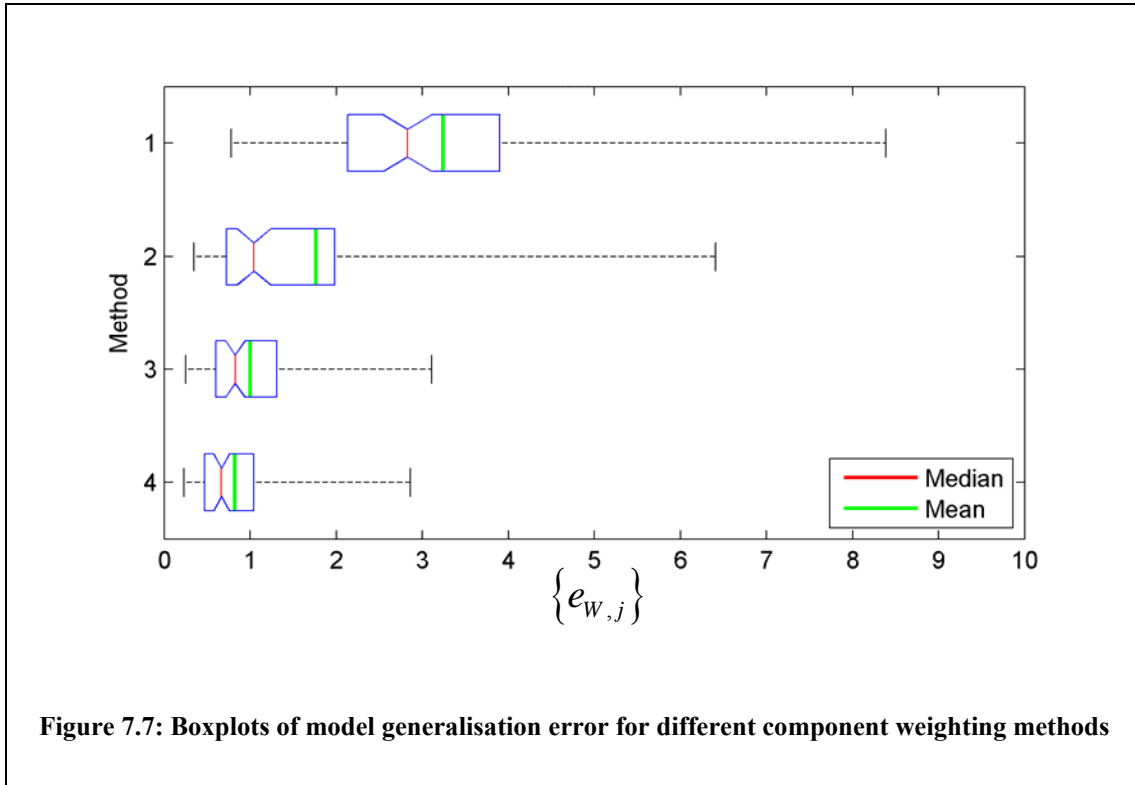
7.7 Results

Using our dataset of 101 masses, we built a combined model of appearance using the steps described in sections 7.4, 7.5, and 7.6 for improving model fit. That is, we constructed a shape model having first aligned the set of mass borders using the MDL optimisation algorithm [34]; we then constructed a texture model using masses, separated from background regions using the completed method described in chapter 6; finally we weighted the parameters of shape, texture and size to form combined appearance vectors for each mass, using weights that had been directly optimised to reduce the generalisation error associated with the combined model. This model was

compared to the model we constructed using Caulkin's method without modifications (see section 7.2).

To compare the two models, we define the total model error for a given mass R_M as the difference between itself and its regeneration from the combined appearance model (that is, $\left| R_M^{"} - R_M \right|_{RMS}$). Again, leave-one-out was employed to compute unbiased errors for the full dataset, and allowing us to compute the total mean generalisation error (MGE) for each model. These data, along with the respective number of modes contained in shape, texture and combined models are given in Table 7.3. The improvements to individual stages of the model building process have resulted in a model that produces significantly lower generalisation errors (Wilcoxon signed rank test $p < 0.0001$). Moreover the model is more compact requiring just 32 combined appearance modes compared to 65.

In the next section, we describe how we can use the model we have constructed to sample new synthetic masses.

Table 7.3: Comparison of total model generalisation error and number of model modes

		Caulkin's model	**Optimised model**
	Total MGE	3.11	1.26
	Shape	25	32
Number of modes	Texture	77	28
	Combined	65	32

7.8 Synthesising new masses

Having constructed a model of mass appearance we use the model to generate synthetic masses. The first step in this process is to sample new appearance vectors from the model distribution. To do this, we assume a Gaussian distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ over the space of combined model parameters. The covariance matrix $\boldsymbol{\Sigma}$ is diagonal such that $\boldsymbol{\Sigma}(i,i) = \lambda_i$, where λ_i the eigenvalue associated with the i-th mode (and equal to the variance of model parameters in the training data along that mode). Because $\boldsymbol{\Sigma}$ is diagonal, we can sample independently from the univariate Gaussian distribution $N(0, \lambda_i)$ along each mode.

If we label a new appearance vector sampled from the model as $\tilde{\mathbf{b}}_{com}$, then we can reconstruct a synthetic mass in the same way we reconstructed real masses to compute model generalisation errors.

In Figure 7.8 we show six synthetic masses generated using the method outlined above. Qualitatively, these masses appear to have a good likeness to real masses in the training data. Of course the ultimate test of the synthetic masses will be how realistic they appear when added to real normal mammograms. Thus we do not attempt to quantitatively evaluate their appearance in this chapter. This evaluation is performed in chapter 11 when we have completed a model of the interaction of mass appearance with the appearance of surrounding breast tissue in mammograms.

With regards to this, Figure 7.9 shows one of the six masses depicted in Figure 7.8 super imposed on two regions of mammographic background. The first background is a

region from a normal mammogram (Figure 7.9 (b)). The second is a real mass background obtained in the previous chapter (Figure 7.9 (c)). We can see that the mass background contains spicules associated with the mass that has been subtracted, and that existing breast tissue in the region appears to have been distorted (with the appearance that breast parenchyma has been pulled in toward to the centre of the region). Figure 7.9 (d) and (e) show the synthetic mass added to the two background regions. When added to the normal region, the mass does not appear to match its surroundings. Meanwhile when added to the mass background, the synthetic mass appears to be a very convincing simulation of a real malignant mass.

From this example we can see that whilst the model developed in this chapter provides a good description of the central density in mass lesions, to model the full mammographic appearance of real malignant masses we must also consider the mass background.

7.9 Discussion

In this chapter we have shown how to fit a statistical appearance model to our dataset of masses. The form of model used has several advantages with regard to our goal of synthesis. Most obviously, it is generative, thereby allowing masses to be synthesised simply by sampling from the learnt model distribution. In addition the model provides a flexible representation of appearance. For example, rather than imposing a pre-determined shape as a template for masses (such as an ellipse [101, 102]), the model constraints are learnt directly from the data.
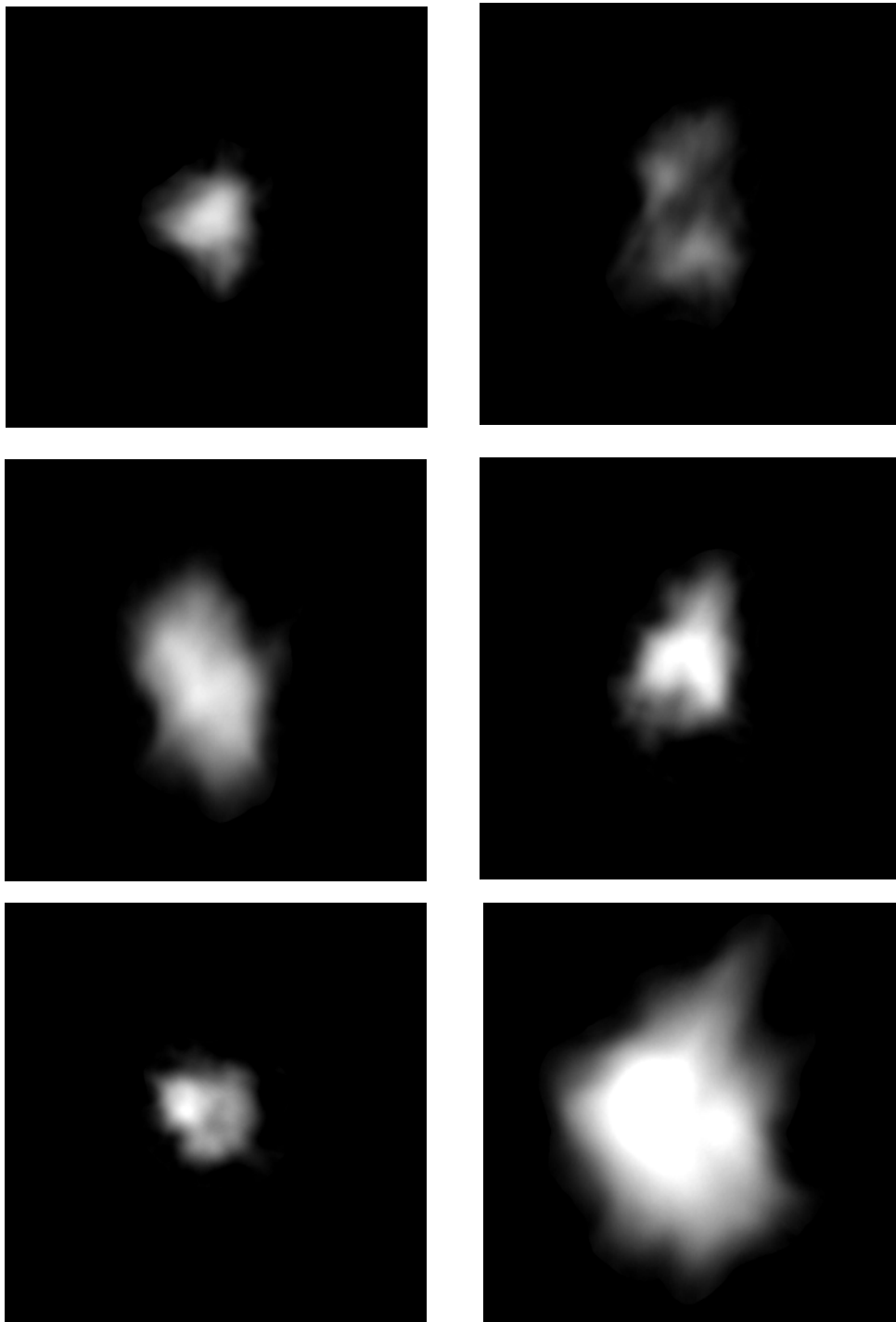
Figure 7.8: Six synthetic masses randomly sampled from the combined appearance model

Another advantage of the models used in this chapter is the fact that the combined appearance model allows for a correlation between shape, texture and size to be learnt. Our model of image appearance is designed to correspond to a physical model of x-ray attenuation through a mass. In the physical model we would expect there to be a strong correlation between the size of a mass and the amount by which it increases attenuation. Therefore it is important our model is able to represent this correlation as it appears in digitised mammograms.

The main weaknesses of the model (as applied to our data) are: the difficulty in defining correspondence between mass objects that do not contain obvious physical landmarks; the assumption of linear variability across the model space. We discuss each point below.

With regards to the problem of finding correspondence between masses, our solution has been to avoid imposing a theoretical correspondence that is not in fact evidenced by the real data. Instead, we try and select the best match in the shape of mass borders by using the MDL optimisation algorithm. We saw that this produced a statistically significant improvement in the ability of the shape model to generalise to unseen mass borders, whilst compacting the shape space.

However, as reviewed in section 4.2.3, Cootes et al. [31] have further extended the idea of using MDL to optimise shape correspondence so that optimal correspondence is searched for directly in the training images (and thus accounts for texture correspondence in addition to shape). We have not attempted to apply such a technique

to our data, although this may be an interesting area for further development.

As described in section 7.8, to sample new masses we assume a Gaussian multivariate distribution over the model space. This model may not be appropriate if there are non-linear correlations between the model parameters. PCA removes all linear correlations in the data, however any non-linear dependencies will remain in the model space. Given the heterogeneity of mass appearance, it is possible that the appearance vectors describing each mass will lie in a non-linear manifold of the space. In section 4.2.3, we discussed various attempts at modelling non-linearity in a model space (for example, fitting a mixture of Gaussians rather than a single multivariate distribution). However, a visual inspection of the model parameters (using, for example, pair wise scatter plots of the parameters along each mode) indicates no clear non-linearities. For example, the data have not fallen into two obvious clusters as did the brain stem MR data described by Cootes and Taylor [28]. To judge whether there are more subtle non-linearities in the data we would need a much larger training set of masses.

In Figure 7.9, we highlight the difference in appearance between a normal region and a mass background region. In Figure 7.9 (a) we show a mass sampled from the appearance model developed in this chapter. In Figure 7.9 (b) we show a normal region alongside a real mass background region in Figure 7.9 (c). We can see in the mass background region a pattern of spicules and a general distortion of structures in the region. In Figure 7.9 (d) and (e) we show the synthetic mass added the normal and mass background region respectively. Figure 7.9 (e) is a much more convincing simulation of a real malignant mass than Figure 7.9 (d).

Figure 7.9: a) a synthetic mass b) a region of normal mammogram c) a mass region d) adding the synthetic mass to the normal region e) adding the synthetic mass to the mass region
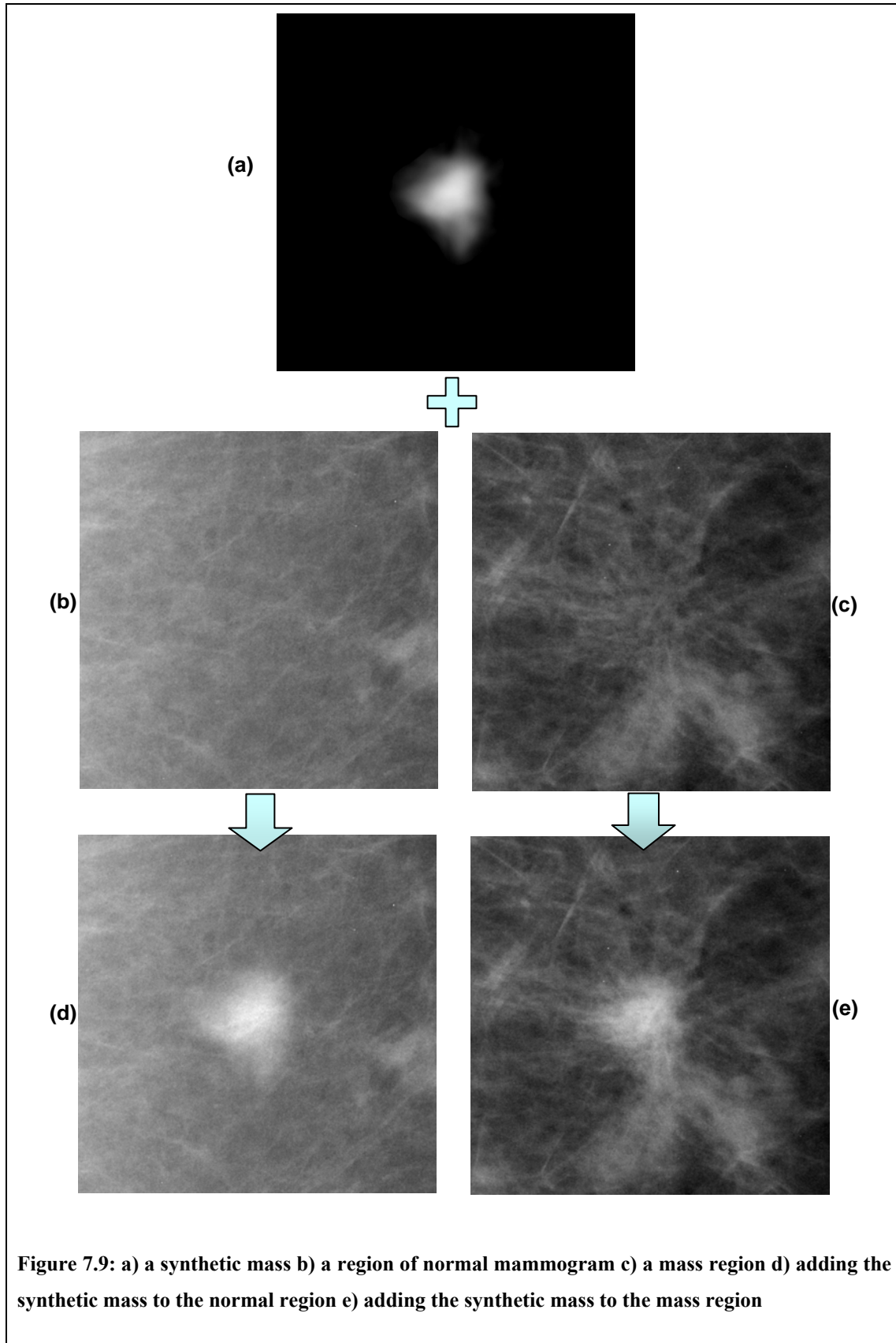
Thus our goal for the next chapter can be summarised as developing a method to modify normal regions (as depicted in Figure 7.9 (c)) so that the have the appearance of mass background regions (as depicted in Figure 7.9 (c)).

7.10 Summary

In this chapter we have presented a method for modelling and synthesising masses. In this context, the term mass has specific meaning, and refers to the images we obtained in chapter 6 by separating the pixel grey levels corresponding to attenuation through the central density of lesion from background.

We described a method developed by Caulkin for modelling this form of image object. The model learned the variation in appearance of masses across a training set of data, and could subsequently be used to sample new synthetic masses.

We observed that the basic model described by Caulkin did not adequately fit our training data. To formalise this observation, we discussed the properties we would like a model to have, and described how such properties could be measured directly from the training data. In particular, we introduced the idea of model generalisation errors that could be used to quantify how well various components of the model might describe a global set of data.

Given a method for measuring model performance, we attempted to optimise the model with respect to our data. In particular, we improved how mass shapes were aligned prior to applying PCA and improved the method for selecting weights with which to scale the parameters of shape, texture and size so that they were commensurate in the combined model. In addition, we showed how changing the way masses were separated from the

background (see sections 6.4 and 6.5) improved texture model fit.

We believe the changes were necessary improvements to the model, as evidenced by the statistically significant reduction in mean generalisation error. Having implemented these changes, we believe the model we have constructed from our data provides a good representation of mass appearance.

We showed several examples of synthetic masses that had been generated by randomly sampling from the model distribution. These masses were similar visually to the real data, although a quantitative evaluation will not be performed until such masses are added to real mammograms (see chapter 11).

Finally, it was noted that we have only described a model for the appearance of the central density in malignant mass lesions. Our goal is to synthesise the complete appearance of malignant mass lesions – including spicules associated with the central density and the distortion of normal breast tissue in the mass region. This additional appearance information is stored in the background images from which masses were subtracted (see section 6.6). The difference in appearance between a normal mammogram region and mass background, and the difference this makes when a synthetic mass is superimposed, was highlighted in Figure 7.9.

Thus in the next chapter, we focus on these background images and show how their appearance can be synthesised in normal mammograms.

Chapter 8 : Modelling the Appearance of Mass Backgrounds

8.1 Introduction

To understand the motivation for this chapter, we first recall the steps we have taken towards modelling and synthesising mammographic masses in the preceding two chapters.

In chapter 6, we described how we could represent each mass region in our dataset as the linear sum of two images:

Original mass background → mass + mass background

In chapter 7, we showed how we could model, and subsequently synthesise, masses. However, as we showed in Figure 7.9, simply adding these to a region of a normal mammogram does not replicate the realistic appearance of a malignant mass. A normal mammogram region does not have the same appearance properties as a mass background.

Thus the aim in this chapter is to be able to synthesise the appearance of a mass background in a previously normal region of mammogram. To achieve this we will attempt to learn the appearance properties of the dataset of mass backgrounds obtained by the separation process described in chapter 6. As such, these become the sole data for consideration in this chapter (we will examine the correlation between masses and mass backgrounds in the following chapter), as highlighted in Figure 8.1.
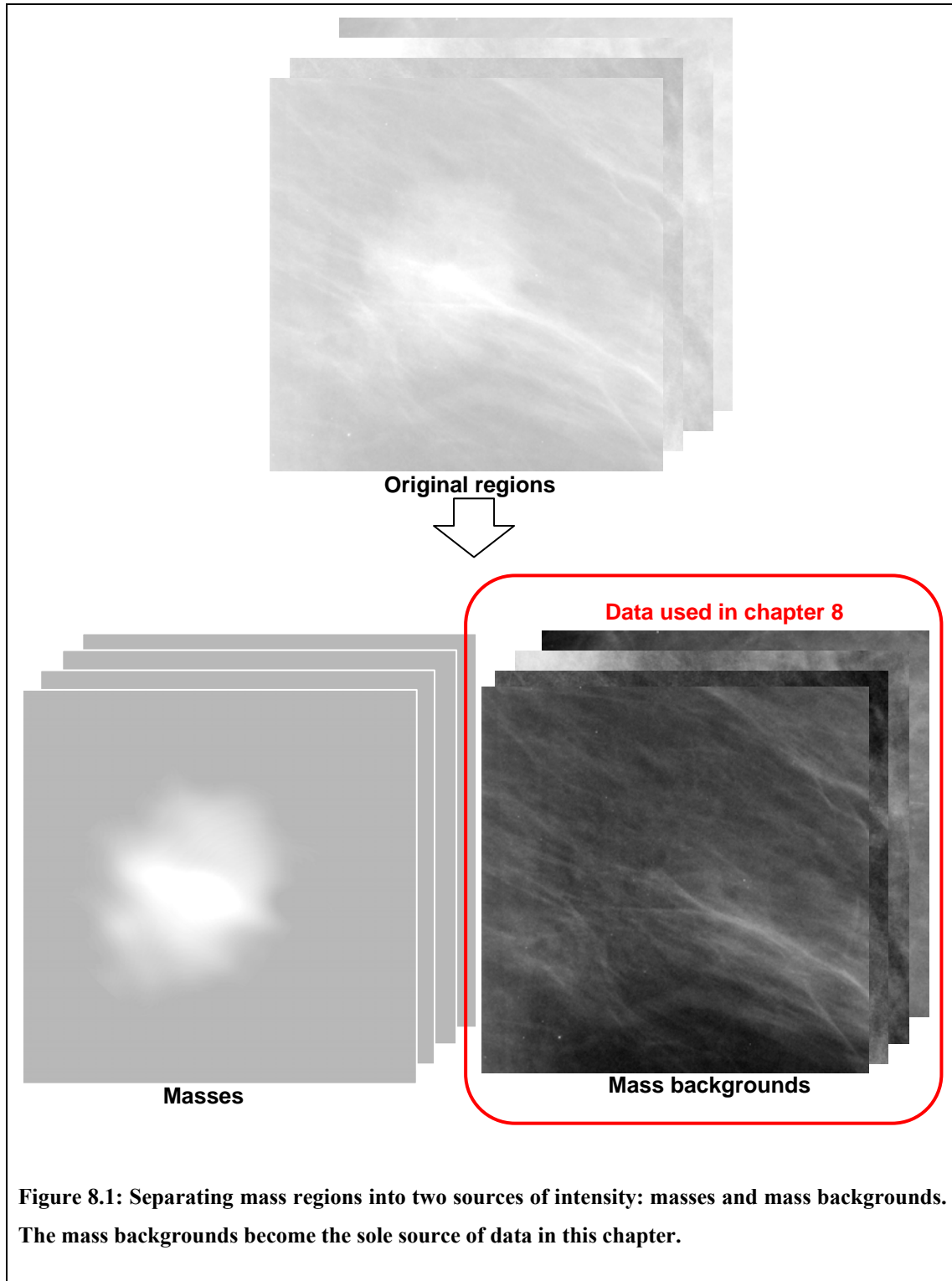
Figure 8.1: Separating mass regions into two sources of intensity: masses and mass backgrounds. The mass backgrounds become the sole source of data in this chapter.

A description of the remaining sections in this chapter is given below.

Section 8.2 gives a brief description of previous attempts to synthesise mass spicules, in each case as part of a larger method to generate synthetic malignant masses. We describe how focusing on spicules alone cannot reproduce the complete appearance of a mass background and use this to motivate the work for the remainder of the chapter.

In section 8.3 we return to the topic of multiscale image transforms and their use in analysing mammographic structures. We reintroduce the dual-tree complex wavelet transform (DT-CWT), reviewed earlier in section 4.3.3, and describe how its properties may be beneficial when analysing and synthesising mammograms.

Section 8.4 then presents two further transformations of the DT-CWT, the inter-level product (ILP) and inter-coefficient product (ICP). We show these transformations in their original form as published by Anderson et al. [7, 8] and in addition describe our own modifications to the transforms. We show how the modified transforms can be used to gain greater information about linear structures in the mammogram regions.

Having selected the DT-CWT as a useful tool for decomposing mammogram regions, sections 8.5, 8.6 and 8.7 present three methods for synthesising mass background appearance in normal regions. In each case, we take the DT-CWT of a normal region and modify the coefficients, before inverting the decomposition to reconstruct synthesised mass background appearance.

8.2 Previous models of spicule appearance

When a malignant mass is spiculated (as are the majority of the masses in our dataset), the most obvious difference between a normal region and the region surrounding the mass is the appearance of the spicules. These represent additional structures to those

already present in normal tissue. Of the literature on synthesising mammographic abnormalities reviewed in section 3.2, three papers explicitly describe simulating spiculated masses [23, 99, 102]. We provide a brief reminder of the method for generating spicules in each case below.

Ruschin et al. [99] simulated mammographic abnormalities, including spiculated masses, using various pre-defined parameterised functions. As such, an individual spicule was generated by first generating a 'super'-ellipse (that is, a shape described by the equation $(x/A)^n + (y/B)^n = 1$, where $0 < n < 1$) of constant intensity, before bisecting the ellipse along the shorter axis and discarding one half, and finally blurring to achieve a smooth drop-off at the border. A set of between 10 and 20 individual spicules were sampled, randomly oriented and centred on the centre of a simulated mass. Judging how well the model described spicule appearance is difficult because no quantitative evaluation of synthetic mass realism was reported. However, it appears to be an overly simplistic representation of spicule appearance, based neither on a physical model of spicule composition nor an image model of how spicule appearance varies in real data.

Saunders and Samei [102] also used a predetermined shape model to simulate individual spicules, however an attempt was made to learn grey level appearance from real data. Triangles were used as a template of spicule shape. Spicule intensity was defined in profiles perpendicular to the longest axis of the triangle, based on a set of intensity profiles measured from real spiculated masses. Details of how many spicules per mass were generated, and their locations and orientations, were not given. In an evaluation of how realistic simulated masses appeared to a radiologist, spiculated

masses were rated less real than non-spiculated masses[6]. This indicates that the appearance of synthetic spicules may not have been realistic.

Finally Caulkin [23] fitted a statistical shape model to learn the shape of individual spicules from a dataset of 602 real annotated spicules (drawn from a set of 57 masses). New spicule shapes were randomly sampled from this model. As in Saunders and Samei's work, grey level texture was described with respect to grey level profiles sampled perpendicularly to the spicule shape. In this case, elliptical profiles were fitted to the set of grey level profiles extracted from the training data, and the resulting distribution used to specify the texture of a spicule sampled from the shape model. The position of each spicule in the training data was measured with respect to its location on the mass border, the distance to the centre of the mass, and the orientation relative to the radius at the point. The joint distribution of these three parameters was learnt and used to generate a position for each new spicule. Finally the distribution of the number of spicules associated with a mass was learnt from the data and used to specify how many spicules should be generated for each synthesised mass. Of the three methods reviewed this included the most thorough description of how spicules could be modelled and synthesised. However, in a discussion of his work, Caulkin reported that many of the synthetic spicules appeared unrealistic. As a consequence, synthetic spiculated masses were distinguishable from real masses by mammogram readers in an observer study.

---

[6] An exact breakdown of numbers is not reported, however this fact is visually apparent from a histogram of the radiologists ratings included in the paper

From this review we see that attempting to synthesise spiculated masses has proved a difficult problem, and one that has yet to be fully addressed. Although the methods described above used different models of spicule appearance, all three are based on the same idea: that spicules can be modelled as individual objects, a number of which can be generated independently and added to a mass. Whilst this approach may be capable of generating some examples of malignant masses, we believe it will never be able to generate the complete appearance of breast tissue in a mass background. In particular, we think the approach suffers from the following three drawbacks:

i)   Modelling and generating spicules individually does not allow for correlation or interaction between spicules. For example, adjacent spicules of a particular mass may share appearance characteristics (especially those relating to position and orientation).

ii)   Some spiculated masses contain areas of spicules that are very hard to describe as groups of individual spicules. Recall from Tabar and Dean [115] in section 2.2.4 the 'sheaf of wheat' appearance of some spiculated regions. Figure 8.2 depicts one such example. In the enlarged region, the spiculated area appears as a striped linear texture. Assigning individual spicules within this area would be a near impossible task, and would not sufficiently describe the appearance of the region.

iii)   Modelling spicules independently from the background region into which they are inserted fails to account for any interaction between local background structures and the spicules. An often described feature of malignant masses is how the spicules anchor the mass within the surrounding breast tissue, distorting

local structures. The appearance could be described as a tension between the breast parenchyma and the mass spicules. This effect suggests that in addition to adding spicules to a normal region, the region itself must be modified.

Whilst the first drawback could be addressed by learning the joint appearance distributions of spicules associated with a mass (indeed this is discussed by Caulkin in his thesis [23]), it is hard to see how an individual model of spicule appearance can overcome the limitations described in points ii) and iii). Instead, we propose a more complete solution to the problem. Using techniques influenced by the texture synthesis literature, we propose jointly synthesising any structures associated with a mass in addition to any distortions to the surrounding tissue in regions of normal mammograms.

To the best of our knowledge, no previous methods for synthesising signs of disease in mammograms have attempted such a complete reconstruction of the appearance of mass backgrounds.
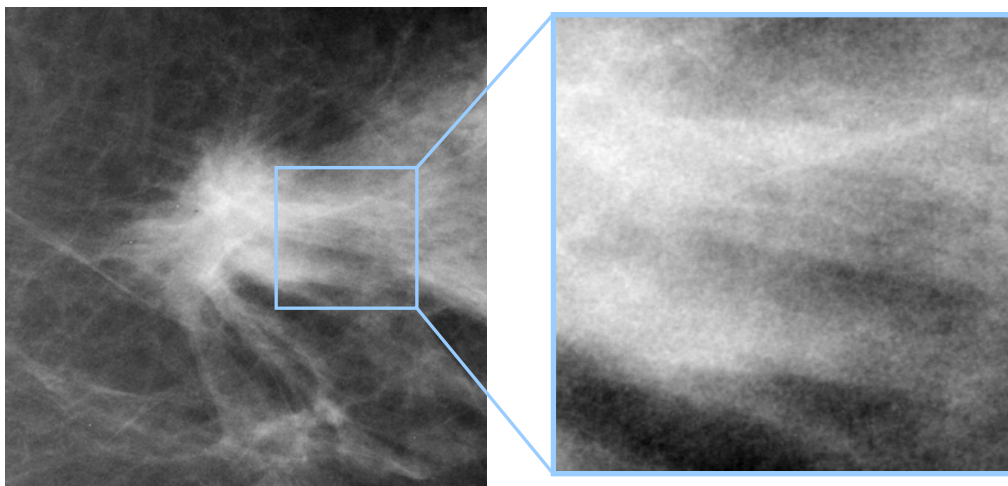
Figure 8.2: Left image: A malignant mass with associated "sheaf-of-wheat" spicules. The image on the right shows an enlarged area where such spicules are evident

To achieve this requires a richer description of the appearance of tissue structures present in mammograms than that provided by the grey level intensities alone. In the following section we describe how mammographic texture can be decomposed across multiple scales and orientations, producing a rich description of local structure.

8.3 Multiscale decompositions of mammograms

In this section we return to the topic of multiscale image transforms and their use in decomposing mammographic texture. A review of previous work in this area was given in section 3.3.1, whilst a technical review of three particular transforms (the monogenic signal, steerable pyramids and the dual-tree complex wavelet transform) was given in section 4.3.

8.3.1   Introduction

Mammograms are rich in features across a range of scales and orientations. When analysing a mammogram region, a useful first step would be to obtain information on a structure or feature at any location in the region. For example, at any pixel we might ask:

- Does this pixel belong to part of a larger structure?

- If so, what type of structure (for example, linear, blob-like)?

- What size (or scale) is the structure?

- How is the structure oriented?

Clearly this information cannot be inferred from the pixel grey level alone. Instead, a commonly used solution is to decompose a region in terms of orientation and scale

using a bank of filters. The filter responses at each orientation and scale provide a rich set of data from which the information above can be inferred (either explicitly or implicitly) and subsequently used to solve a variety image analysis tasks.

Table 8.1: Summary of literature in which decomposition across scale and orientation have been used in particular mammographic applications

Author	Filter bank/decomposition	Mammographic application
Wai et al. [121]	Monogenic signal	Detecting curvi-linear structures (CLS)
Pan et al. [86]		Breast segmentation
McLoughlin et al. [78]	Complex log-Gabor filters	Detecting CLS
Schenk and Brady [103]	(at multiple orientations and scales)	
Petroudi et al. [92]	MR8 filter bank	Texture analysis and classification
Rose [98]	Steerable pyramid	Texture analysis and synthesis

In the literature review in section 3.3 we discussed several examples of this approach from the mammography literature, describing in each case the choice of filters used and the task they were employed to solve. Table 8.1 lists notable examples of this work.

Of the work reviewed, Rose's [98] attempt at synthesising mammogram texture is most closely related to our goal for this chapter. In this case a real-valued steerable pyramid

was used to decompose mammograms across scale and orientation. The idea behind this approach is that modelling and synthesising in the domain of the pyramid coefficients (that are separated in subbands localised in scale and orientation) is easier than trying to model and synthesise original image intensities directly. In keeping with this approach, we used the steerable pyramid in our initial attempts to synthesise the appearance of breast tissue in mammograms. However, as previously discussed in section 4.3.2 of chapter 4, the coefficients of a real-valued steerable pyramid are not shift invariant.

The models for synthesising mammographic texture we propose in sections 8.6 and 8.7 rely on learning how patterns of decomposed coefficients within each subband correspond to particular patches of mammograms. For this to work successfully the relative position of a patch within an image should not affect the pattern of decomposed coefficients it produces at a particular frequency. If the coefficients of a decomposition are shift dependent, this property won't hold. Thus we require a shift invariant image decomposition.

Returning to the multiscale transforms reviewed in section 4.3, we note that shift-invariant pyramid features could be obtained by using the complex steerable pyramid. However, this doubles the redundancy of coefficients in the pyramid. With this in mind, we choose instead to use the dual-tree complex wavelet transform (DT-CWT). The DT-CWT has a lower redundancy and is more computationally efficient than the complex steerable pyramid.

In the following section, we give a brief reminder of how the DT-CWT is constructed and describe the set of filters we have used in our work.

8.3.2   Dual-tree discrete wavelet transform

The dual-tree complex wavelet transform (DT-CWT) was developed by Kingsbury [65] to overcome the lack of shift invariance and poor directional selectivity inherent in standard real discrete wavelet transforms.

Shift-invariance and directional selectivity are achieved by computing two parallel real discrete wavelet transforms with a 90º phase difference. The responses of the two transforms are combined to form 6 complex valued subbands at each frequency level of the transform. By construction the subbands are strongly oriented at angles of approximately ±15º, ±45º and ±75º. The magnitudes of the coefficients are invariant to shifts in the input image, whilst the phase of coefficients can be used to infer local structural information in an image.

The DT-CWT matches the properties that originally led us to choose the steerable pyramid as useful tool for modelling and synthesising mammographic appearance. That is, it is able to separate an image into subbands localised in scale and orientation, whilst the invertibility of the filters provide a suitable framework for the synthesis methods we propose in sections 8.6 and 8.7. Given the additional property of shift invariant coefficient magnitudes and the possibility of capturing structural information through local changes in coefficient phase, we use the DT-CWT throughout the remainder of the chapter.

In all our work we used an implementation of the DT-CWT kindly supplied by Nick Kingsbury. The implementation used orthogonal near-symmetric filters comprising 13 and 19 elements for the first level, and 14 element Q-shift filters from the second level onwards. Further details on the design of these filters are described by Kingsbury [67];

here we note that the filters:

    i)   allow perfect reconstruction

    ii)   are linear phase

    iii)   are composed of rational coefficients

The impulse responses for each subband obtained using these filters are shown in Figure 8.3, highlighting the directionality of the subbands.

Figure 8.3: Impulse responses for the real and imaginary parts of the 6 DT-CWT subbands
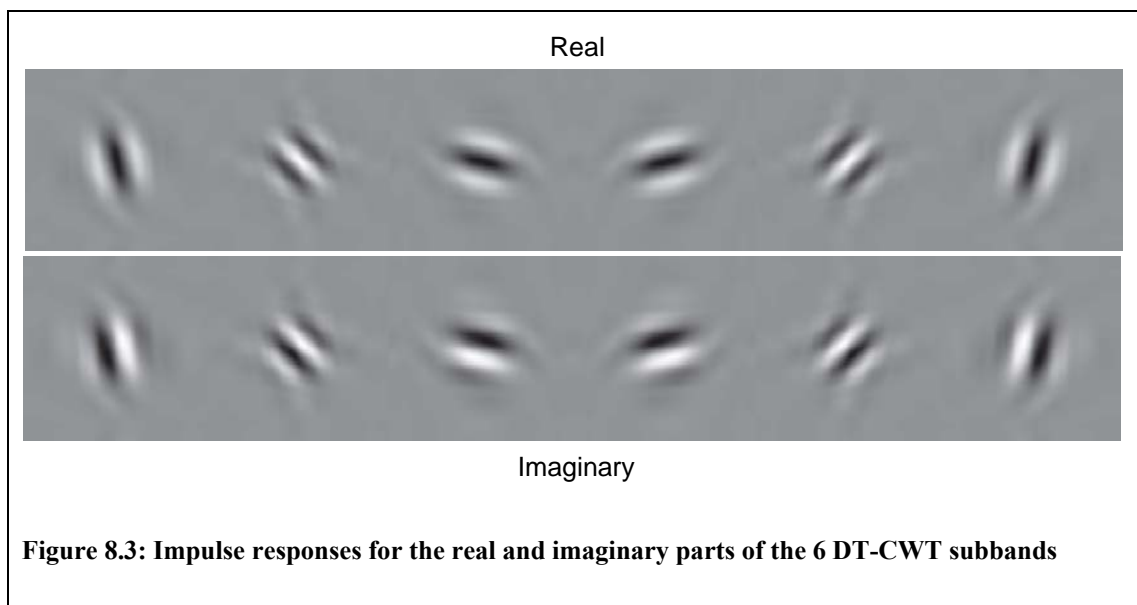
Having selected the DT-CWT as a suitable tool with which to continue our work, in the next section we describe how the phase of DT-CWT can be used to detect linear structures in mammogram regions.

8.4 Detecting linear structures in mammograms

Much of the appearance of mammograms is formed by the superimposition of a large number of linear structures at different orientations and scales. In attempting to model mammogram appearance a natural requirement is to be able to detect such structures

and describe their properties. In this section we explore how this can be achieved using the DT-CWT. In particular, we look at how the phase of DT-CWT coefficients can be used to infer local structural information. Returning to the work summarised in Table 8.1, we observe that using phase information formed an important part of the feature measures used in the three CLS detection algorithms [78, 103, 121].

8.4.1   Introduction

Because the DT-CWT decomposes an image across scale and orientation, it is a natural tool for analysing linear structures and of course this was a key motivation in choosing to use it.

At a simple level, information on linear structures can be obtained using only DT-CWT coefficient magnitudes. The presence of a structure at a particular location can be determined from the maximum magnitude across all subbands. Meanwhile information on the scale and orientation of the structure can be inferred from the specific subband that generated the maximum response.

However the magnitudes alone do not give information about what kind of structure is present at the location - for example, the magnitudes alone cannot differentiate between a line and edge. Moreover, taking structure orientation as equal to the orientation of the subband with maximal response, limits angular resolution to six directions across 180 degrees.

In the next two sections we review work presented by Anderson et al. [7, 8] that show how the phase of DT-CWT coefficients can be used to infer both structure type and a more precise computation of orientation.

8.4.2   Computing orientation from phase: the inter-coefficient product

The inter-coefficient product (ICP) was introduced by Anderson et al. [8]. It is based on the observation that if a strong linear structure (such as an edge or line) lies within the support of a DT-CWT coefficient, the phase of the coefficient varies linearly with respect to the distance of the structure to the location at which the coefficient is sampled. The rate of change of phase (that is, the magnitude of the phase gradient) is constant for all structures whilst the direction of the phase gradient is perpendicular to the structure. Therefore by computing the phase gradient with respect to a fixed direction in the image, the orientation of the structure can be inferred. This idea is shown in Figure 8.4. The figure shows the magnitude and phase coefficients from a single oriented subband in the DT-CWT decompositions of three lines. The lines have orientations of 0º, 12º and 24º to the vertical respectively. The 4th oriented subband is shown as this produces maximal responses for each line. In the images of Figure 8.4, intensity varies from black to full colour with coefficient magnitude, whilst the hue of the colour represents phase. For each of the lines in Figure 8.4, the phase sampled at the white arrow is shown in the coloured bar overlapping the subband image. The change in phase of this sample is shown in the axes under each image. For the vertical line, there is no change in the phase as the sample is taken in parallel to the line. In the 12º and 24º lines, note how the phase changes twice as quickly in the 24º line than the 12º – for example the coloured bar for the 12º line shows just over one complete cycle of hues (from light blue to light blue) whilst the colour bar for the 24º line includes two complete cycles (from red to red). Thus sampled at a fixed direction, phase change is proportional to structure orientation.

Figure 8.4: Magnitude/phase plots of DT-CWT coefficients for three lines oriented at 0º, 12º and 24º degrees to the vertical. The colour bars depict the phase as sampled at each of the white arrows. The schematic axes show the relative change in phase for each line
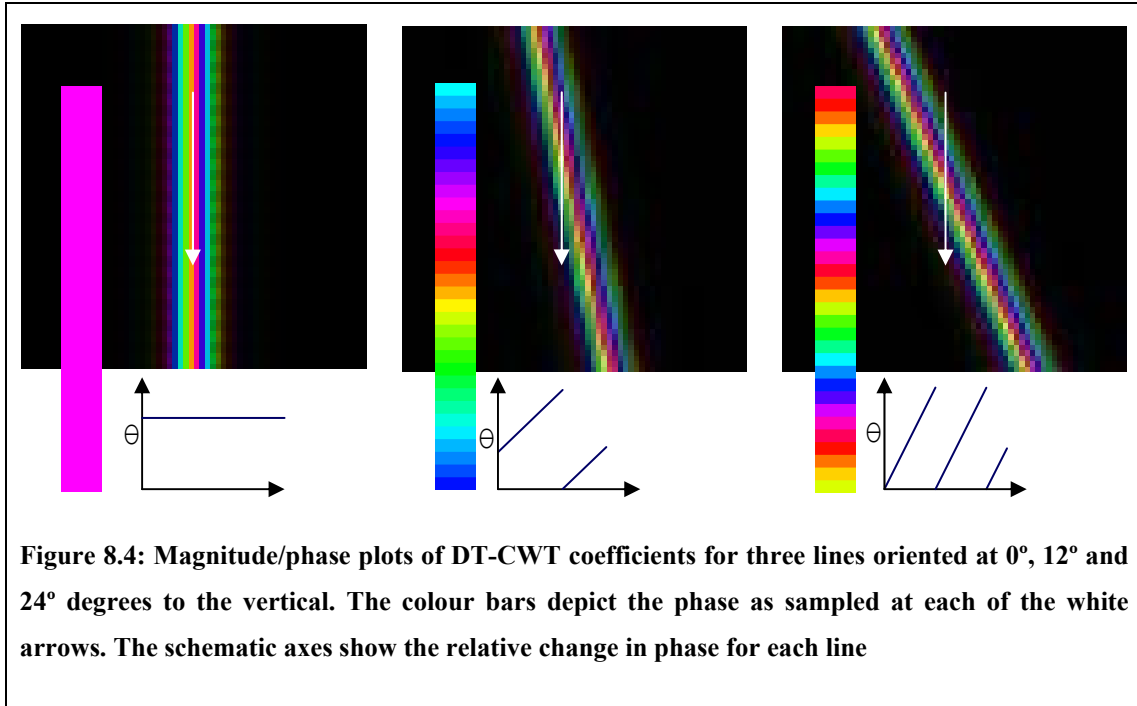
To obtain the phase gradient, each wavelet coefficient is multiplied by the complex conjugate of its neighbour producing a complex number that has magnitude equal to the product of the wavelet magnitudes, and phase equal to the difference of the wavelet phases – hence the name 'inter-coefficient product'. The neighbouring coefficients are chosen in the direction closest to the orientation of the subband. Thus horizontal coefficients are used for subbands 1 and 6, vertical coefficients for band 3 and 4, and opposing diagonals for band 2 and 5. We further discuss the exact choice of coefficients used in section 8.4.4.

To move from phase gradients to orientations it remains to compute the constant of proportionality between them. This was done empirically by Anderson, computing values of 4.49 for bands 1, 3, 4 and 6 and 8.98 for bands 2 and 5 (note this ratio is relative to the distance between coefficient locations and therefore remains constant across all scales).

From its construction, an ICP transform has the same number and size of subbands as the DT-CWT from which it was computed.

8.4.3   Computing structure type from phase: the inter-level product

We look now at a second transform of DT-CWT coefficients introduced by Anderson, named the inter-level product (ILP) [7].

Like the inter-coefficient product described in the previous section, the transform is based on the observation that the phase of a DT-CWT coefficient changes linearly in the support of an image feature. However, as the name implies, rather than computing phase differences between neighbouring coefficients in the same subband, the inter-level product calculates the phase gradient between neighbouring levels of equivalently oriented subbands in the dual-tree.
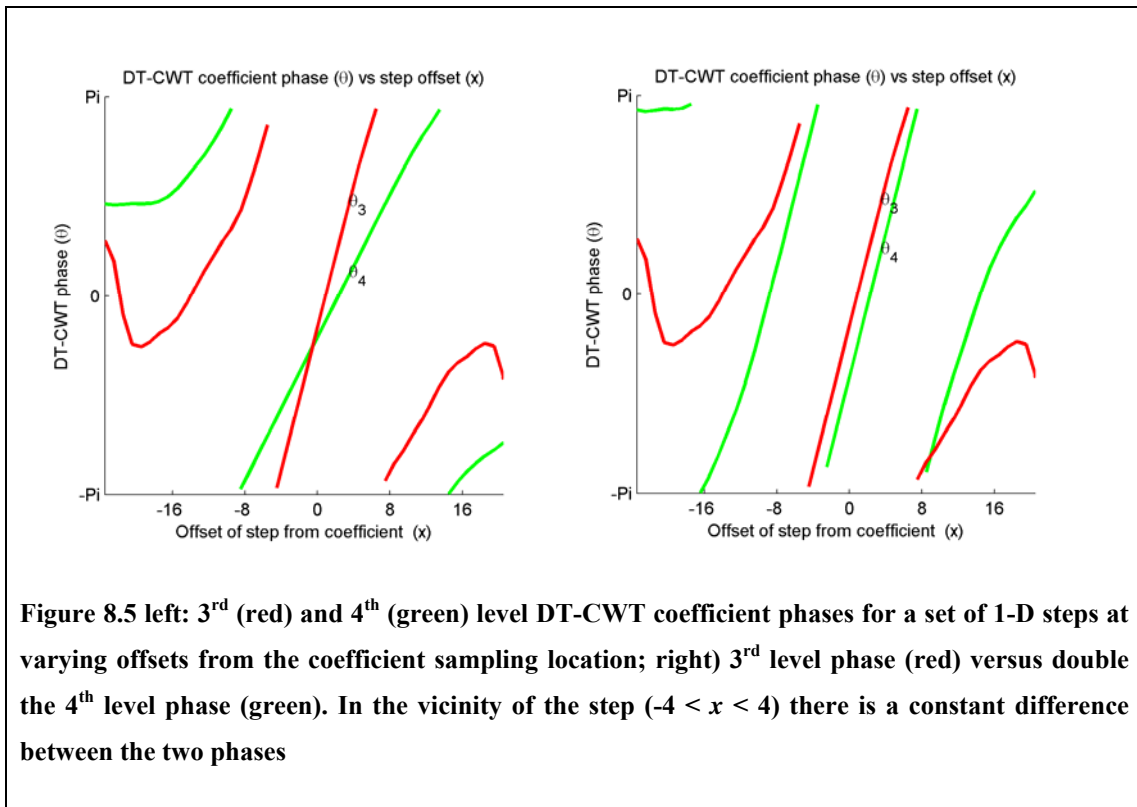
Figure 8.5 left: 3rd (red) and 4th (green) level DT-CWT coefficient phases for a set of 1-D steps at varying offsets from the coefficient sampling location; right) 3rd level phase (red) versus double the 4th level phase (green). In the vicinity of the step (-4 < x < 4) there is a constant difference between the two phases

In the vicinity of a feature, the phase of a dual-tree coefficient at level l varies twice as fast as the phase of the coefficient at the corresponding location in level $l+1$. Therefore the phase difference between level l and double the phase of $l+1$ is constant regardless of spatial location of the feature. This is depicted in Figure 8.5 where the red line depicts the phase of a coefficient in the third level of a 1-D DT-CWT for a set of steps at varying offsets from the coefficient. Note how this is linear in the region between step offsets of approximately ±3. In the left-hand plot, the green line shows the phase of the corresponding location in the fourth DT-CWT level. Because the support width of the fourth level is double that of the third, the range of offsets for which phase is linear is double that of the third level. As a consequence phase varies at half the rate. In the right hand plot, the green line now depicts double the fourth level phase. In the region in which the third level coefficient is linear, the gradients of the two lines are equal, thus the difference between the phases is constant regardless of the offset. Moreover, the value of the inter-level phase gives information about the type of structure present.

For example, in the first oriented subband of a 2-D DT-CWT, an ILP phase of 0 corresponds to an ideal positive edge, $\pi/2$ a positive line, π a negative edge and $-\pi/2$ a negative line. This phase-to-feature relationship is shown on a set of real and imaginary axes in Figure 8.6. Thus the phase of an ILP coefficient is analogous to the local phase computed in the monogenic signal and used in phase-congruency measures (see monogenic signal, section 4.3.1). The magnitude of an ILP coefficient is a measure of the consistency of feature strength between adjacent levels.

By using both the ICP and ILP we can infer structure type and gain more precise measure of orientation than using DT-CWT magnitudes alone. However, before

showing examples of applying the ILP and ICP to mammograms we suggest some modifications to the transforms.

Figure 8.6: Relationship of feature type to ILP phase. Positive edges correspond to a phase of 0; positive lines $\pi/2$; negative edges π; and negative lines $-\pi/2$
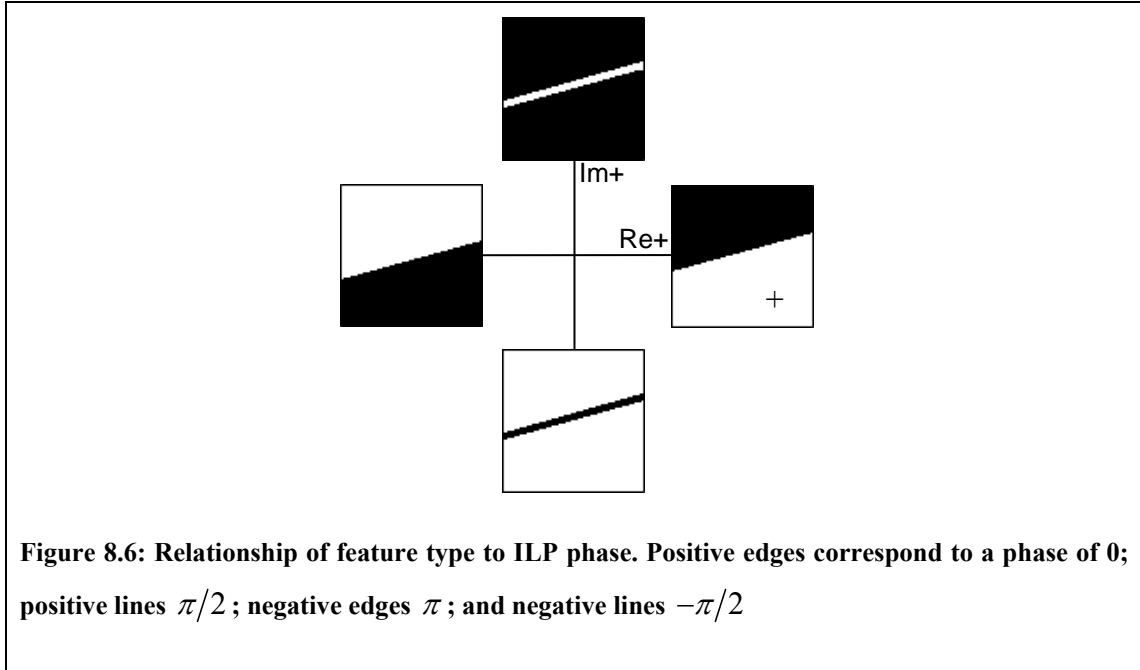
8.4.4 Modifying the ILP and ICP transforms

After initial experiments using ILP and ICP transforms as published in the papers by Anderson, we noticed there were some aspects of the coefficients produced by the transforms that did not behave as we would like with respect to rotations of image features. In this section we present the modifications we have made to the transforms in light of this.[7]

The basic premise of our modifications is that if we denote the phases of an ICP coefficient and an ILP coefficient in the vicinity of a structure as θ and λ respectively,

---

[7] From personal communication, we understand changes to how ICP and ILP orientation varied with rotation were presented in Anderson's thesis [7]. However this was published after the time the work in this section was implemented

then the phases of corresponding ICP and ILP coefficients to a copy of the structure rotated through ϕ radians should be $\theta + \phi$ and λ respectively[8].

That is, as a measure of orientation, ICP phase should vary consistently with rotation, whilst as a description of structure type ILP phase should be rotation invariant.

In light of this, the changes we have made to the transforms are:

i)   Computing the ICP symmetrically about a coefficient

ii)   Discarding the concept of edge polarity

iii)   Defining ICP phase consistently over a 2π range

ICP symmetry

In Figure 8.7 we show how the coefficients used to compute phase differences in the original ICP transform are not consistent with rotation. On the left-hand-side of the figure is an edge oriented at approximately 15 degrees, thus producing maximum wavelet and ICP coefficients in the 1st subband. The central ICP coefficient (marked with a blue cross) is computed using the product of the blue and green DT-CWT coefficients. The edge-feature is then rotated through 90 degrees as shown in the right-hand of Figure 8.7. It now produces maximal coefficients in the 4th subband. The coefficients in the central column have an identical relationship to the rotated edge as the central row of coefficients in the original feature. However, the central ICP

---

[8] By corresponding ICP/ILP coefficient we mean the coefficient from the subband most closely aligned with the structure, in an equivalent location relative to the orientation of the feature

coefficient is now computed from the product of the blue and red coefficients. The ICP coefficient assigned to the central location in the original image will instead be equivalent to the green ICP coefficient in the rotated image.
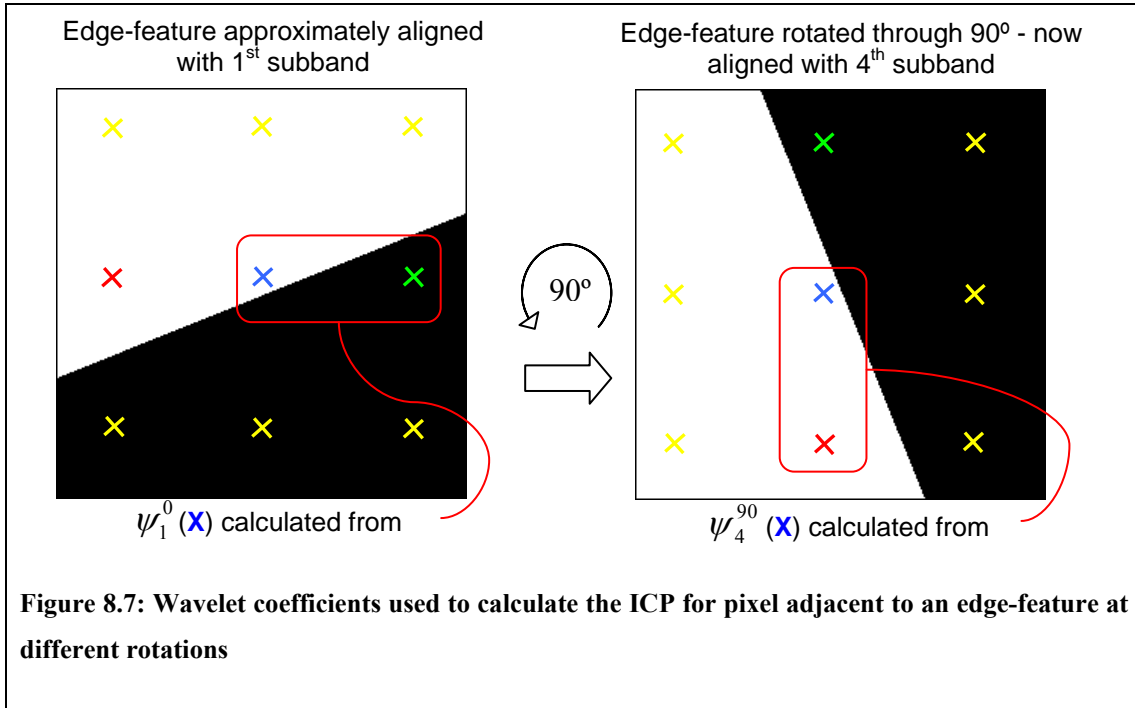
Figure 8.7: Wavelet coefficients used to calculate the ICP for pixel adjacent to an edge-feature at different rotations

The difference occurs because the calculation of the ICP does not act symmetrically about the coefficient it's assigned to. In the left column of Figure 8.8, we show the coefficients used to compute the ICP for each of the six oriented DT-CWT subbands, clearly highlighting the asymmetry about the pixel to which the product will be assigned.

Instead, we propose using the neighbouring coefficients on both sides of the centre location when computing the ICP. The coefficients used in this case are shown in the right column of Figure 8.8, and are symmetrical about the centre location.

There are several ways to combine both neighbouring coefficients to compute a single ICP coefficient. Options include computing the average of the two products between the

centre location and its two neighbours; taking the maximum of the two products or interpolating the neighbouring coefficients to find new coefficients a single sample width apart and symmetric about the centre, before computing the phase difference between these. In practice we have found taking the maximum of the two neighbouring products works well. Thus in the example in Figure 8.7, because the edge lies between the blue and green DT-CWT coefficients, the blue/green product will have a much larger magnitude than the red/blue product. Therefore, in both the original and rotated images, the blue/green product will be assigned to the blue ICP coefficient.

Note by taking the maximum in this way the new ICP by itself is no longer invertible. However as we show below, this can be remedied by using the new ICP in conjunction with changes we propose to the ILP.

Edge polarity

For a 1-D signal, specifying four pure feature types (positive step, negative step, positive impulse, negative impulse) and matching each feature to an ILP phase makes sense. However, we believe the equivalent in two dimensions – positive and negative lines, positive and negative edges – makes a false distinction between positive and negative edges. The key difference is that where the time axis of a 1-D signal has an implicit direction, there is no such direction in the 2-D plane of images. Therefore positive and negative edges are not different types of structures, but the same structure at different orientations.
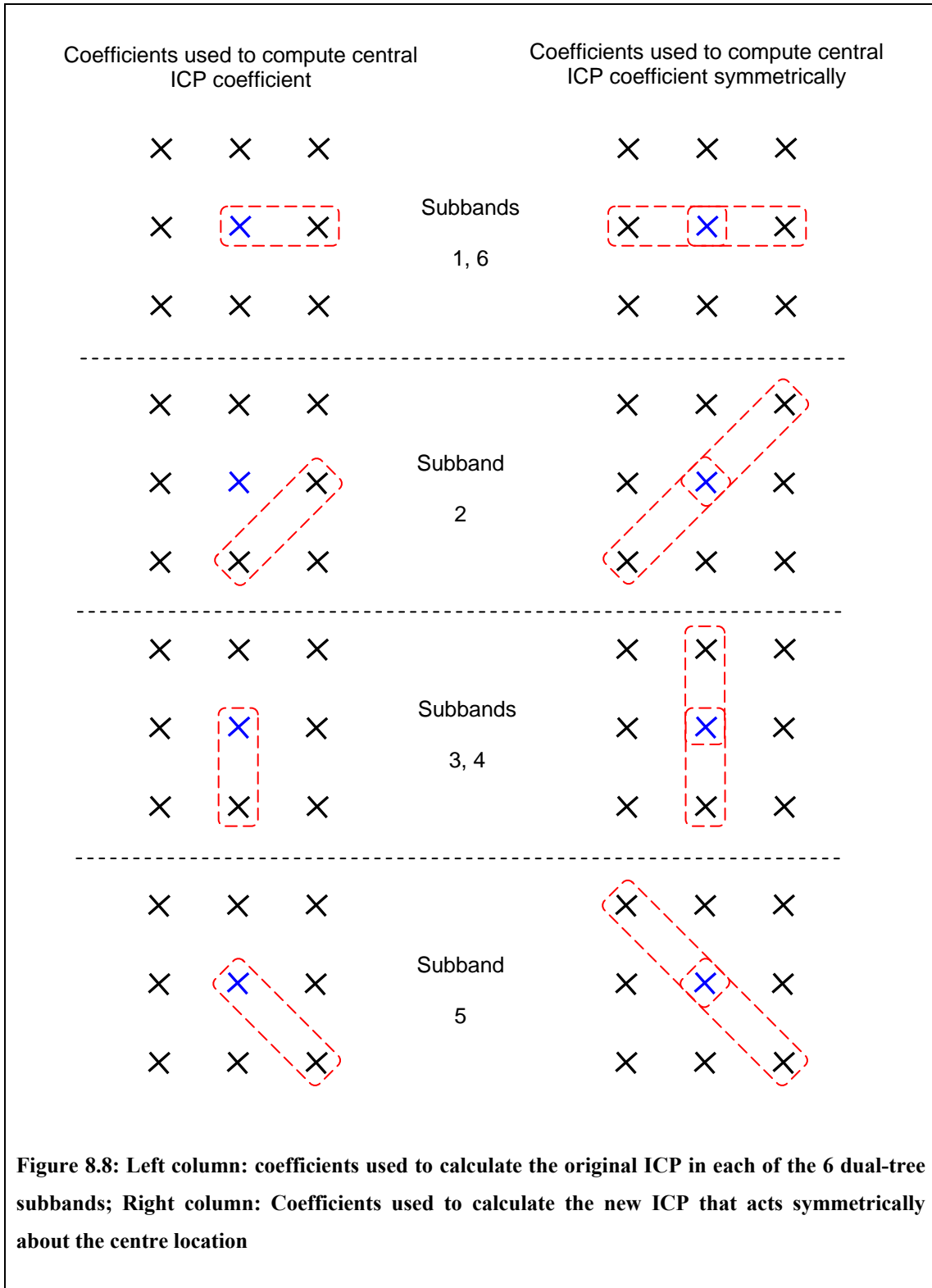
Figure 8.8: Left column: coefficients used to calculate the original ICP in each of the 6 dual-tree subbands; Right column: Coefficients used to calculate the new ICP that acts symmetrically about the centre location

To see the impact of this on ILP coefficients, consider a pure positive edge initially oriented to be maximal in the 1st oriented subband. Using the phase-to-feature relationship depicted in Figure 8.6, an ILP coefficient in the vicinity of the edge will have a phase of 0. If we rotate the edge through 90 degrees, the corresponding ILP coefficients (now the 4th oriented subband) will still be 0 as desired. However if we now rotate the edge by another 90 degrees, the edge is now a negative edge aligned in the 1st subband, and the corresponding ILP coefficient will have a phase of π.

Further, consider the ILP phases of features that are not ideal and so lie between the four feature types. These are depicted by the red (positive line to negative edge), yellow (positive line to positive edge), blue (negative line to negative edge) and green (negative line to positive edge) arrows in the left-hand axes in Figure 8.9. Again we consider the effect of rotating through 90 degrees.

Feature phases in the first three bands of the original image map to equivalent phases in the latter three bands of the rotated image. However for features in the latter three bands of the original image, phases varying from a positive line to a negative edge (depicted by the red arrow) translate into phases varying from a positive line to a positive edge (depicted by the yellow arrow) in the rotated image. Likewise a phase varying from a negative line to a negative edge (depicted by the blue arrow) translates into phase varying from a negative line to a positive edge (depicted by the green arrow) in the rotated image

More formally, we can define the set of phases depicted by the red arrow as:

$$\lambda_{red} = \pi/2 + \alpha, \quad \alpha = 0 \to \pi/2$$

And the phases depicted by the yellow arrow as:

$$\lambda_{yellow} = \pi/2 - \alpha, \quad \alpha = 0 \to \pi/2$$

And therefore

$$\begin{aligned} \lambda_{yellow} &= \pi/2 - (\lambda_{red} - \pi/2) \\ &= \pi - \lambda_{red} \end{aligned}$$

Similarly we can show that:

$$\lambda_{green} = \pi - \lambda_{blue}$$

Thus to establish ILP phases that are rotationally invariant, we map any coefficient with phase $\pi/2 < \theta < 3\pi/2$ to a new coefficient with phase $\theta' = \pi - \theta$. This transform is trivial to implement, we simply force the real part of each ILP to be positive (that is, $a + bi \mapsto |a| + bi$). This can be visualised as folding the left-half of the ILP phase-to-feature axes onto the right, as depicted in right-hand axes of Figure 8.9.

Figure 8.9: Transformation of ILP phase: mapping coefficients with phase $\pi/2 < \theta < 3\pi/2$ to $\theta' = \pi - \theta$
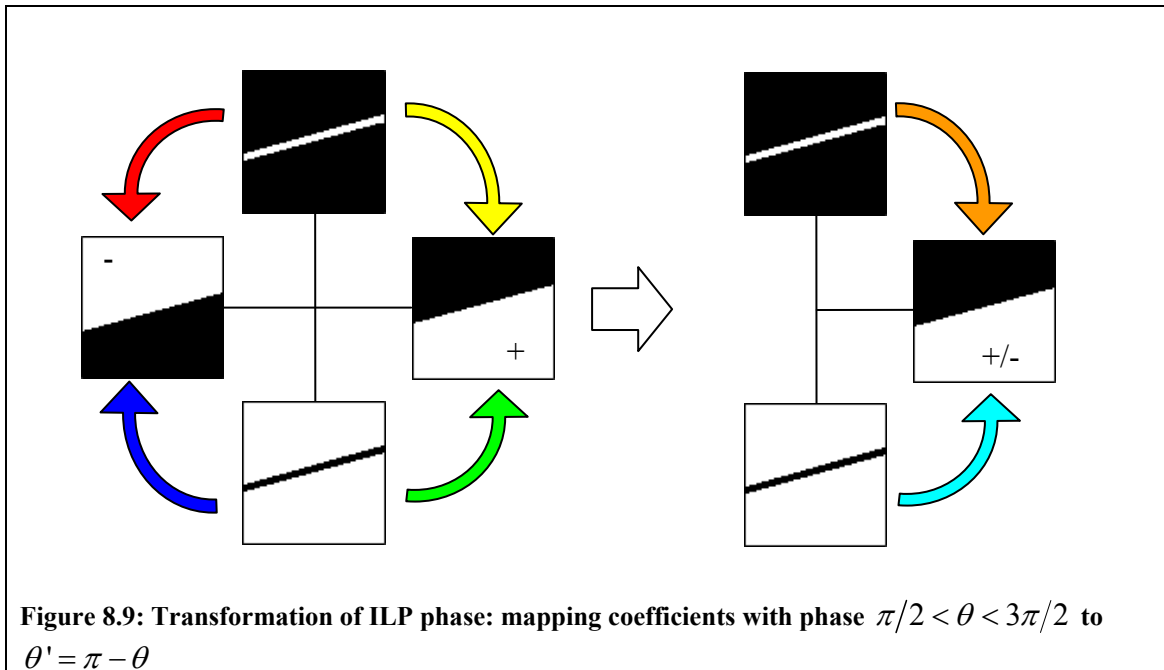
We have effectively discarded the idea of edge polarity, and in doing so have made the transformed ILP coefficients rotationally invariant. However as a result, we have lost information – one consequence of which is that the transform is no longer invertible. Below we show how this information is instead encoded with the ICP transform.

ICP phase range

The range of orientations that can be obtained from ICP phase spans approximately π radians (from -0.70 radians for the 6th subband, through bands 1 to 4, finishing at 2.71 radians for the 5th subband). Because of this a structure oriented between the 5th and 6th subbands may produce some maximal ICP coefficients with a phase of say θ and others of $\theta + \pi$. Moreover, if in any subband the phase of an ICP coefficient in the vicinity of a structure is θ, then the phase of the corresponding ICP coefficient to a copy of the structure rotated through ϕ radians may either be $\theta + \phi$ or $\theta + \phi + \pi$ depending on the angle of rotation ϕ.

By expanding the range of obtainable ICP phase to span 2π we can remove this inconsistency and in doing so, encode the information lost in modifying the ILP transform above.

This is achieved by fixing all coefficients with an untransformed ILP phase of $-\pi/2 < \lambda < \pi/2$ to have an ICP phase in the range $0 \rightarrow \pi$, and all coefficients with untransformed ILP phase $\pi/2 < \lambda < 3\pi/2$ to have an ICP phase in the range $\pi \rightarrow 2\pi$ by adding or subtracting π from the ICP phase as necessary.

Thus rather than having a concept of edge polarity, the change in structure is encoded as a change in orientation though 180 degrees. Conceptually we believe this is a more

accurate way of describing the rotations of structures.

The changes to the transforms are trivial to implement – for a given location, if we let the ILP coefficient be $a + bi$ and the ICP coefficient be $c + di$ then we simply apply the following operations:

$$
\begin{aligned}
&\textit{if} \quad d < 0 \\
&\textit{then}: \quad c = -c \\
&\qquad\qquad d = -d \\
\\
&\textit{if} \quad a < 0 \\
&\textit{then}: \quad a + bi \mapsto -a + bi \\
&\qquad\qquad c + di \mapsto -c - di
\end{aligned}
$$

The first operation fixes all initial ICP phase to lie in the range $0 \to \pi$, whilst the second operation adds π to the ICP phase of features that previously had ILP phase in the range $\pi/2 \to 3\pi/2$.

When we come to invert the transformed ICP and ILP coefficients we simply apply:

$$
\begin{aligned}
&\textit{if} \quad d < 0 \\
&\textit{then}: \quad a + bi \mapsto -a + bi
\end{aligned}
$$

and can subsequently reconstruct the original DT-CWT coefficients exactly.

Applying the transforms as described above, if we denote the phases of an ICP coefficient and an ILP coefficient in the vicinity of a structure as θ and λ respectively, the phases of corresponding ICP and ILP coefficients to a copy of the structure rotated through ϕ radians will be $\theta + \phi$ and λ respectively, as desired.

In the following section we show the results of applying the new ICP and ILP

transforms to mammogram regions and present a method for detecting linear structures based on the transforms.

8.4.5   Applying the ICP and ILP to mammograms

In this section we show images of ICP and ILP coefficients computed for mammogram regions. In addition, we show a scheme for detecting linear structures in mammograms.

Figure 8.10 depicts a mass background from our training data alongside images formed from the ILP coefficients of the region. The ILP images have been formed by taking the maximal coefficients across the six oriented subbands within the second, third and fourth frequency levels of the ILP transform. As with Figure 8.4, a phase-magnitude colour scheme in which intensity corresponds to magnitude and hue to phase is used to depict complex coefficients. For each frequency level the images have been enlarged to the size of the original region. In this colour scheme blue to purple hues correspond to phases close to $\pi/2$. In terms of ILP coefficients, this corresponds to positive lines (that is, a light line against a dark background). As we would expect, these are the dominant structure types in the images.

Figure 8.12 shows the equivalent ICP coefficients for the mass background depicted in Figure 8.10. Meanwhile Figure 8.11 and Figure 8.13 show the equivalent figures for a normal region of mammogram. The normal region has been specifically selected because of its striated texture. Note that when comparing the ILP coefficients in Figure 8.10 and Figure 8.11, the distribution of phase colours appears similar suggesting the type of structures present in both regions are similar.
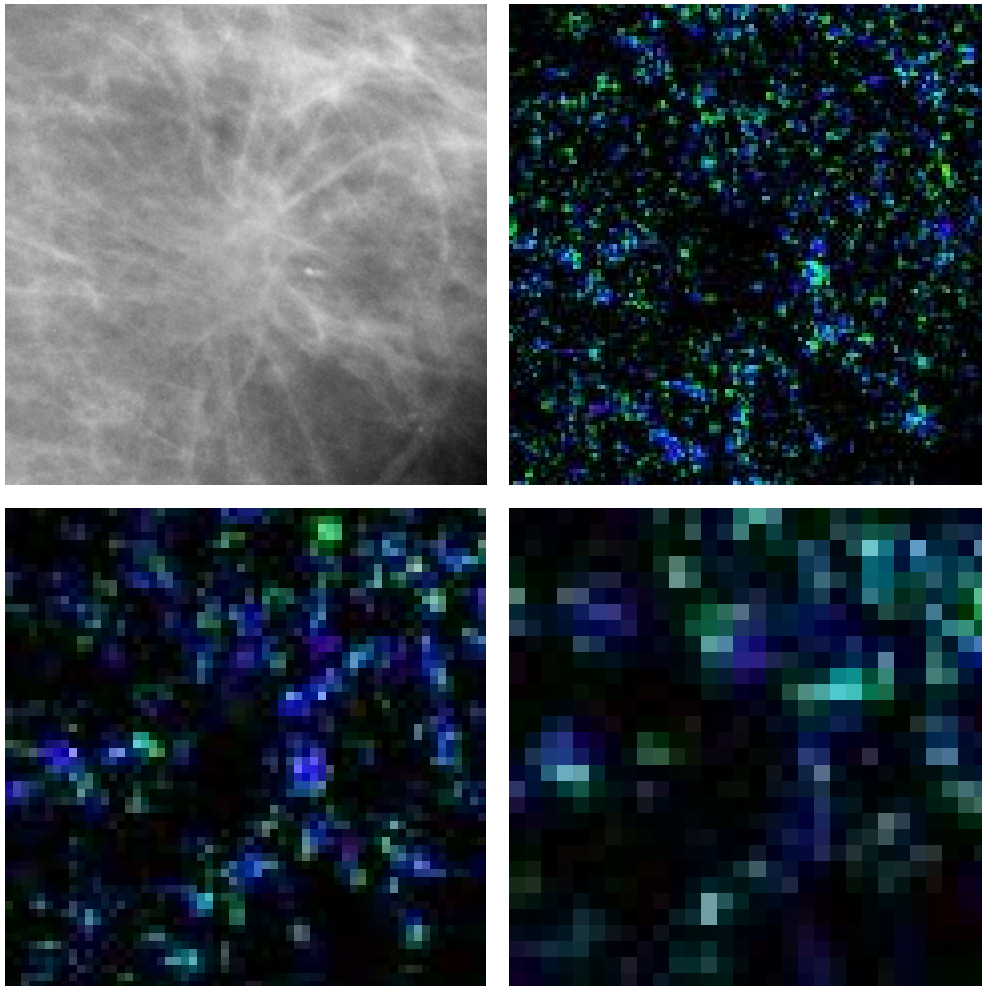
Figure 8.10: ILP coefficents in the 2nd, 3rd and 4th levels for a mass background from our training data

Figure 8.11: ILP coefficients in the 2nd, 3rd and 4th levels for a region of a normal mammogram with striated tissue appearance
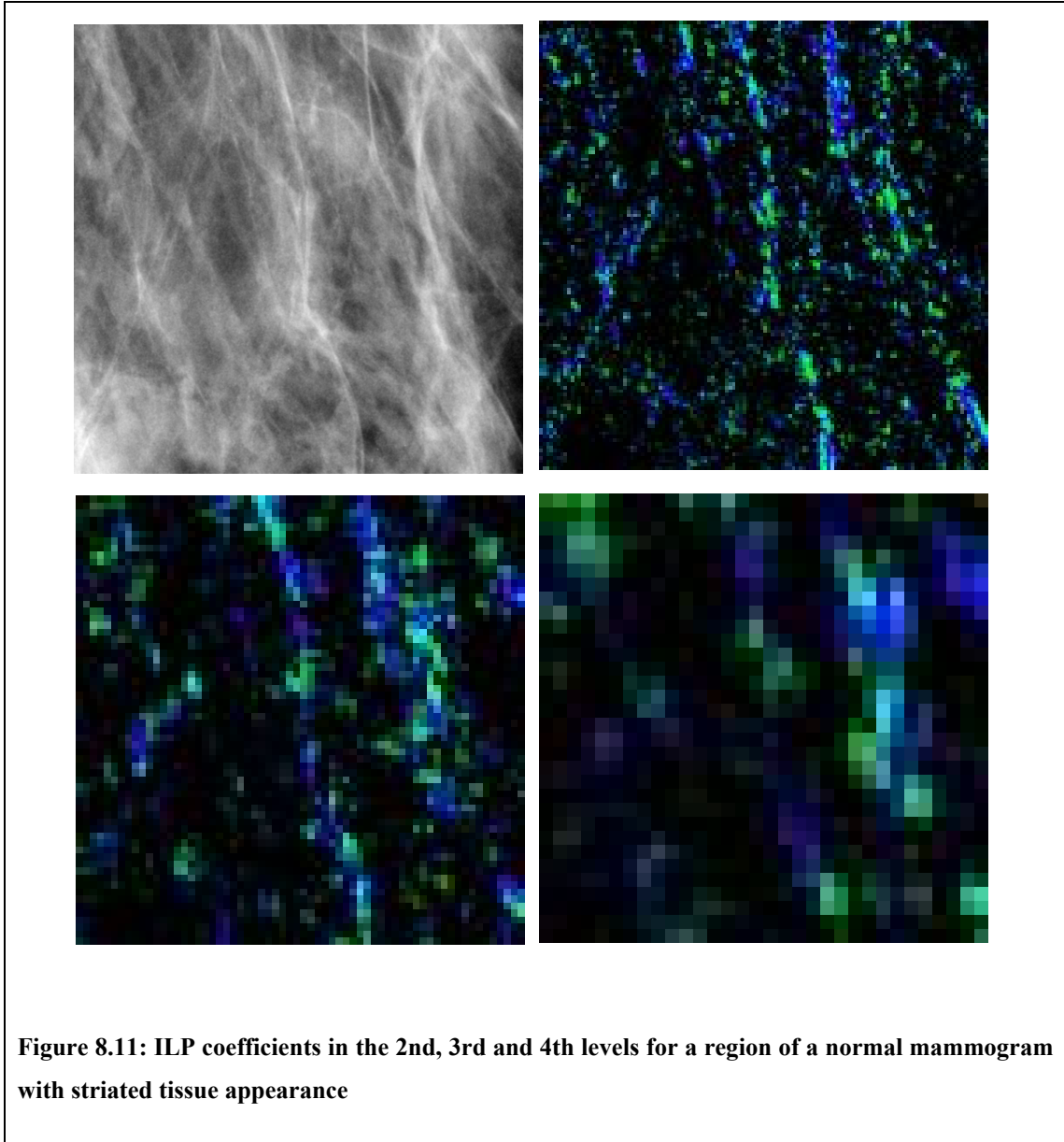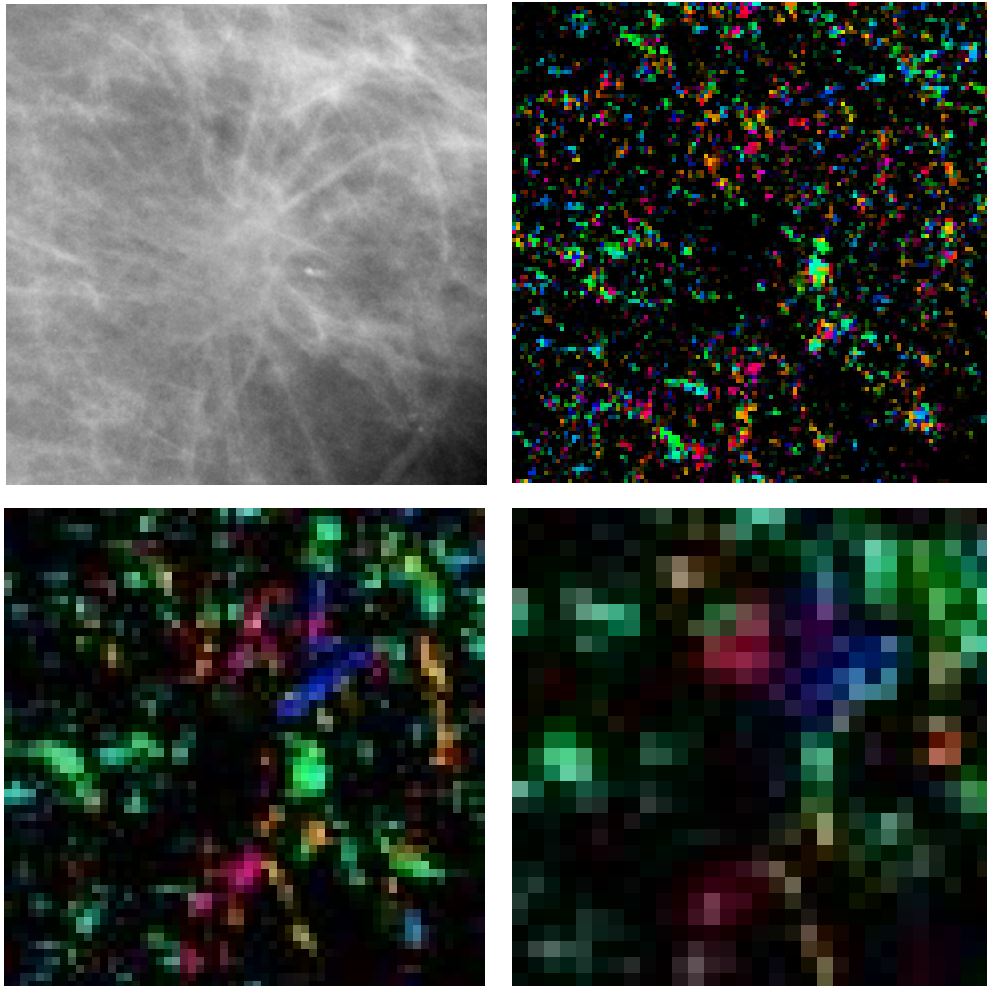
Figure 8.12: ICP coefficients in the 2nd, 3rd and 4th levels for the same mass background depicted in Figure 8.10
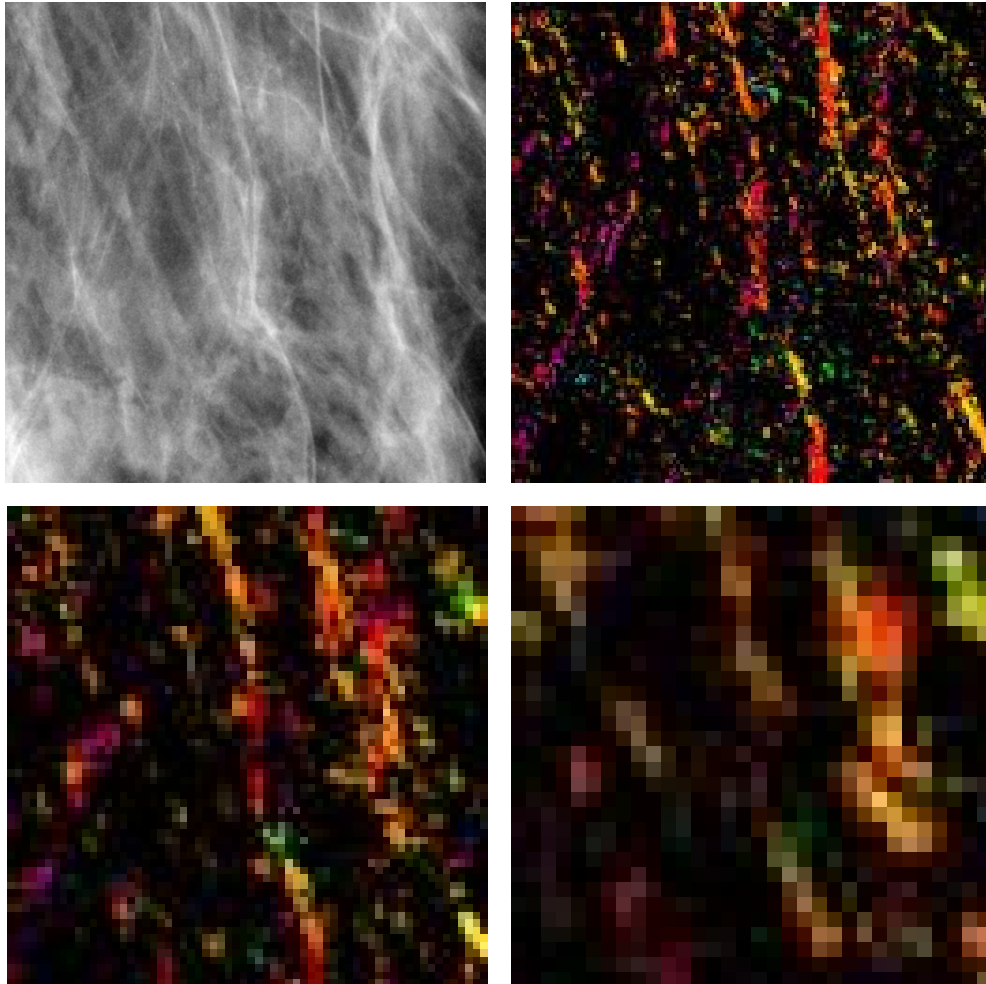
Figure 8.13: ICP coefficients in the 2nd, 3rd and 4th levels for the normal region depicted in Figure 8.11

However, comparing the ICP coefficients in Figure 8.12 and Figure 8.13, we can clearly see how the spread of orientations differs between the two regions. In Figure 8.12, the radial pattern of structures in the mass background produces ICP images with colours throughout the phase spectrum. In Figure 8.13, the structures in the normal region are generally aligned vertically, resulting in a dominance of red and yellow hues in the ICP images.

For some tasks it may be desirable to make a hard classification of linear structures in a region, for example, as used in section 8.7.2. We present a scheme for doing so below, although other ways this may be achieved are discussed in section 8.4.6.

The line detection method incorporates the outputs of ICP and ILP coefficients into a basic framework such as that used in common edge or line detection algorithms (for example, the Canny edge detector [22]) and proceeds as follows:

i)   For each frequency level, compute the maximal ICP and ILP coefficients across the six oriented subbands at each location

ii)   Using the magnitudes of the ILP coefficients as a measure of line strength, and the phase of ICP coefficients as an orientation map, perform non-maximal suppression

iii)   Given the remaining coefficients, compute a map of weak lines based on both a threshold of ILP magnitudes and a threshold of ILP phases. The ILP phase acts to select lines and not edges

iv)   Similarly create a map of strong lines using stricter ILP magnitude and phase

thresholds

v)   Perform hysteresis to compute the final line map

The result of applying this map to the striated normal region shown earlier is depicted in Figure 8.14. To compute weak and strong magnitude thresholds, a histogram of DT-CWT magnitudes at each level was computed for all regions in our training data. At each level the weak threshold was set empirically to include 75% of all locations and the strong threshold 95%. For the ILP phase, a weak threshold was set at 0 (thus favouring any structures between a positive line and an edge) whilst the strong threshold was set at $\pi/4$.

Note that in our implementation, we explicitly detect lines at each frequency scale, as this is the information we use in the synthesis algorithm described in section 8.7. If instead we wanted to compute a combined line map across all frequency scales, then we could simply take the maximal coefficients across all orientations *and* scales (up-sampling lower frequency levels as necessary) in the initial step of the detection scheme. Alternatively, a phase-congruency measure could be computed using the magnitudes and phase of ILP coefficients. This is discussed further in the next section when we consider other possible ways of combining the information provided by DT-CWT, ILP and ICP coefficients.

8.4.6   Discussion

In the sections above we have described two phase-based transforms of the DT-CWT that provide added information to using DT-CWT magnitudes alone. We described the ILP and ICP transforms as they were originally presented by Anderson et al, and

showed modifications to the transforms so that their coefficients varied more consistently with structure rotation.

Figure 8.14: Lines detected at the 2nd, 3rd and 4th levels for a region of a normal mammogram with striated tissue appearance
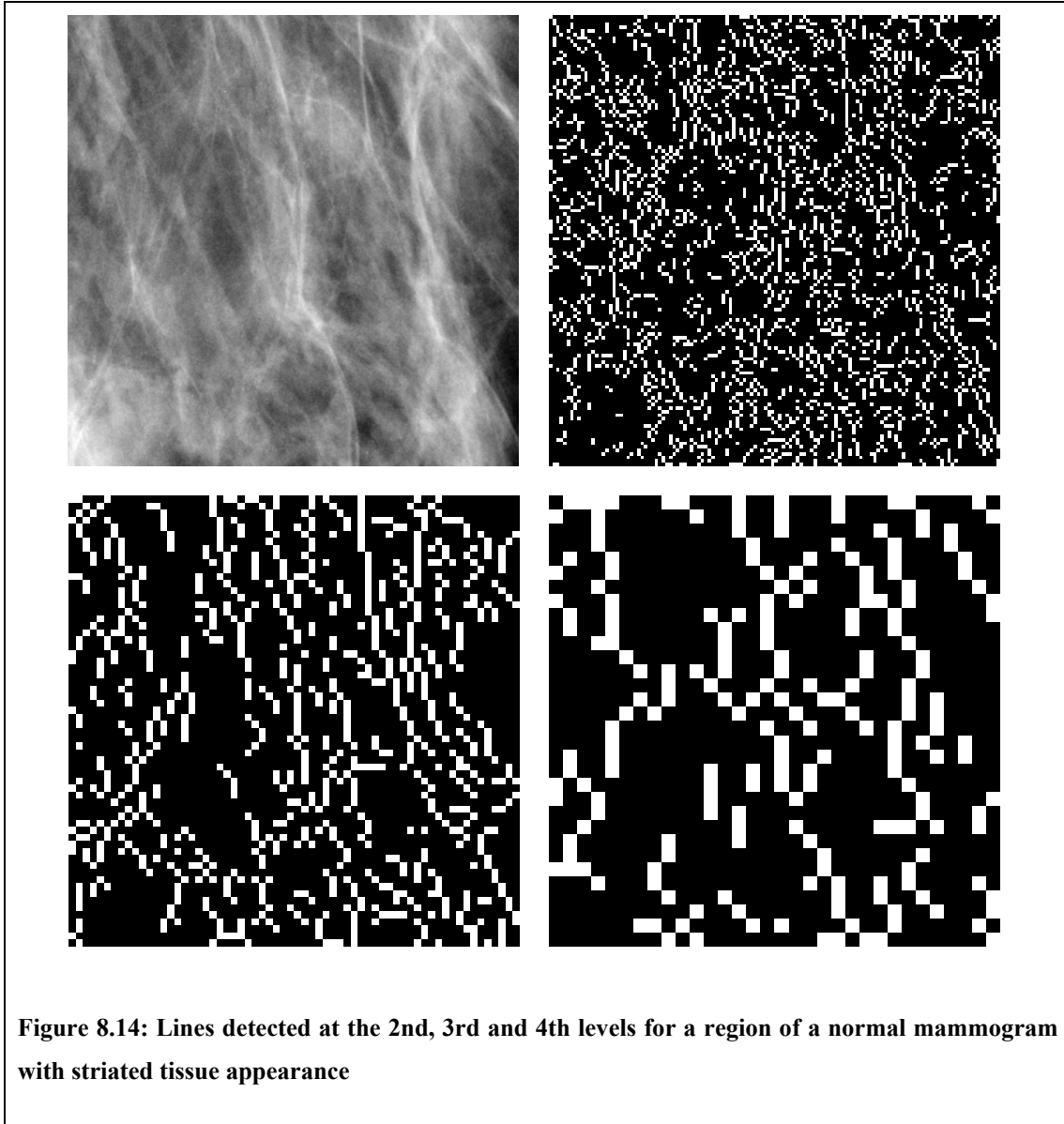
The CLS detection scheme we proposed in section 8.4.5 uses the information provided by the ILP and ICP transforms. Conceptually, our algorithm is similar to those presented by Wai et al. [121], Schenk and Brady [103] and McLoughlin et al. [78]. However, in each of these cases the algorithm was used to produce a map of CLS at the

226

pixel level of a mammogram. As such information was combined across all scales in a measure of phase-congruency (see section 4.3.1). However rather than combining the information from all levels into a single measure, it may also be possible to use the distribution of energy and phase across scales to determine the complete cross-sectional shape of a linear structure.

We also observe that in McLouglin et al.'s algorithm, a measure of feature consistency is computed in all orientated subbands. This allows linear features (that produce a peak magnitude in one direction) to be distinguished from round features (that have approximately equal magnitudes in all orientations). We recall that one of reasons for using the DT-CWT instead of a transform such as the monogenic signal was to obtain feature measures at more than a single orientation. However in the line detection algorithm we presented in section 8.4.5, only the coefficients from the subband of maximal response were used.

For these reasons, we do not think that we have fully exploited the information provided by DT-CWT (and ILP/ICP) coefficients. Given this, we are working on developing a more complete model of the local relationships between DT-CWT coefficients in mammograms. Our aim is to use such a model not only to detect linear structures mammograms, but to be able to differentiate between different classes of structure, such as spicules, ducts, blood vessels or ligaments. However, work in this area is beyond the scope of this thesis and is only discussed further as a topic of future work (see section 12.3).

Having selected the DT-CWT as a suitable decomposition with which to work, and developed further tools to analyse linear structures within our mammogram regions, we

proceed to synthesising the appearance of mass backgrounds in normal regions.

8.5 Directly transferring mass background appearance

Our approach to simulating the appearance of a mass background can be summarised as follows:

1. Take a region from a normal mammogram, and compute its DT-CWT

2. Modify some or all of the DT-CWT coefficients to match the properties of a mass background

3. Invert the DT-CWT to reconstruct a region now with mass background appearance

The remainder of this chapter describes how step 2 might be achieved.

In this section we describe our first attempt, in which DT-CWT coefficients from a target normal region are replaced with coefficients taken from the DT-CWT of a real mass background.

For this method, we assume that all the abnormal structures in a mass background are captured in subbands above some frequency in the dual-tree pyramid, and contained within some contiguous region common to each subband. Moreover, we assume that the coefficients transferred within this region do not encode significant other structures or texture that will produce an unrealistic appearance in the reconstructed target region. We will discuss the validity of these assumptions later; however we first give a more detailed description of the method.

To begin we define more carefully what we mean by a normal region, a mass

background and the set of coefficients to be transferred.

We have already defined mass backgrounds in chapter 6. Recall that each mass background is an $m \times n$ area of mammogram from which a single malignant mass has been subtracted. The centre of the region corresponds to the centroid of the subtracted mass, and the region extends to include all abnormal structures associated with the mass.

For a given mass background, we define the set of coefficients to be transferred with respect to a subset of pixels within the main region. This subset of pixels must be contiguous and contain all abnormal structures associated with the mass. However it does not (and indeed will not) be rectangular. To form this inner region we take the set of pixels enclosed by the annotated mass border (labelled as M in chapter 6) and dilate it until it includes all spicules labelled by an expert radiologist. We label the dilated set of pixels M_d. Note that due to the decimation in the dual-tree transform, the elements of M_d do not directly specify locations in any DT-CWT subbands. Instead for each subband in the L-th level of a transform, the set of coefficients to transfer can be computed from M_d as

$$M_{d,L} = \left\{ \left(\left\lceil \frac{x}{2^L} \right\rceil, \left\lceil \frac{y}{2^L} \right\rceil \right) \mid (x,y) \in M_d \right\}.$$

A target normal region can be any rectangular region of breast tissue extracted from a mammogram (digitised at the same spatial resolution) containing no signs of disease. Obviously the dimensions of this region must be larger than the maximum width and height of M_d. Within this region we specify a pixel C_N to be the centre of the

transferred region.

Having made these definitions, we can now proceed with the algorithm for transferring the dual-tree coefficients, as detailed in Algorithm 8.1.

Randomly select a mass background from the training set

Decompose the target normal region and mass background into a dual-tree complex wavelets consisting of subbands $\{N_{L,\theta}\}$ and $\{B_{L,\theta}\}$ respectively, where $L = 1,\ldots,5$ represents the subband level and θ the subband orientation

Compute the centroid of M_d, $c_M = \dfrac{1}{n} \displaystyle\sum_{(x,y)\in M_d} (x,y)$ where n is the number of pixels in M_d

for level $L = 1$ to 5

    Compute the co-ordinates of the dual-tree coefficients to be transferred based on the region M_d and the factor of two down-sampling between DT-CWT levels

$$M_{d,L} = \left\{ \left(\left\lceil \frac{x}{2^L} \right\rceil, \left\lceil \frac{y}{2^L} \right\rceil \right) \mid (x,y) \in M_d \right\}$$

    Compute the down-sampled offset between the target and mass centres

$$c = \left\lceil \frac{c_N - c_M}{2} \right\rceil$$

    for each orientation θ

        Transfer the dual-tree coefficients from the mass subband into the target subband

$$N_{L,\theta}(x+c, y+c) = B_{L,\theta}(x,y) \qquad \forall (x,y) \in M_{d,L}$$

    end

end

Invert the modified tree $\{N_{L,\theta}\}$ to reconstruct the modified target region

---

Algorithm 8.1: Transferring DT-CWT coefficients from a mass background into a normal region

Clearly this is a somewhat simplistic method for simulating the appearance of a mass background region. However as a first attempt it does produce regions with mass background properties, albeit with the addition of several artefacts, as seen in Figure 8.15. In the next section we consider a modification to the algorithm to overcome the mismatch of tissue orientations.

Normal region   Mass background   Modified region

Figure 8.15: Synthesised region obtained by directly transferring DT-CWT coefficients from a mass backgorund into a normal region. Orientations are not yet aligned betweened the regions
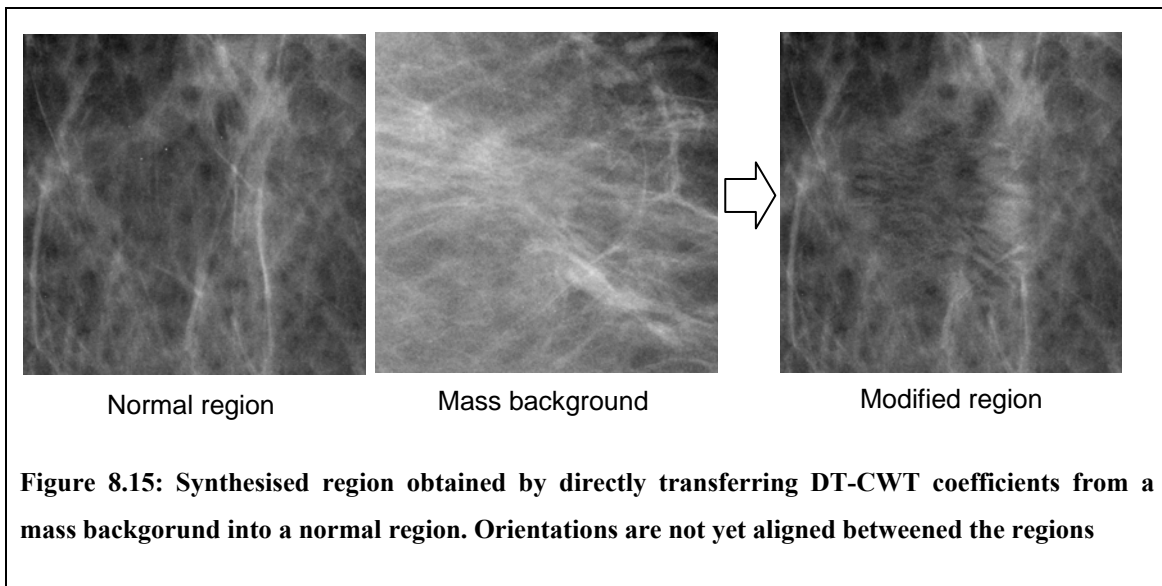
8.5.1   Aligning structure orientations

As previously discussed, mammograms gain their appearance from the complex superimposition of ducts, blood vessels, fat and other tissues that comprise the breast. This can give rise to the appearance of a dominant orientation in regions where several clearly defined structures are aligned in a similar direction.

In the method for modifying normal appearance described in the previous section, if in addition to abnormal structures, the mass background region has an underlying dominant orientation, the structures that give rise to the orientation will be encoded in the transferred DT-CWT coefficients. As a result the appearance of a dominant

orientation will be transferred to the target region. If this orientation differs from the existing dominant orientation of the target region, there will be a mismatch at the borders of the transferred region causing unrealistic appearance. This is apparent in Figure 8.15. The normal region contains strong linear structures with a primarily vertical orientation. In contrast, the dominant linear structures in the mass background are approximately horizontal. When DT-CWT coefficients are transferred from the mass background into the normal region, the mismatch in orientation causes a notably unrealistic texture in the centre of the modified region.

This is depicted schematically in the top row of Figure 8.16. To prevent this mismatch of orientations, we rotate the mass background prior to computing its DT-CWT so that its dominant orientation matches that of the target region, as depicted in the bottom row of Figure 8.16.

As an initial test, the angle of rotation for a mass background to match a target region was selected manually. Results of this process suggested such a scheme was feasible, however, in keeping with our aims for the overall mass synthesis method we must be able to compute this angle automatically.

This was achieved by using the ICP and ILP transformations of the DT-CWT described in section 8.4. Recall that the phase of ICP coefficients provide a measure of local orientation, whilst ILP coefficients provide a measure of structure type and strength. Given a region, we compute its DT-CWT and subsequently its ICP and ILP transforms. Then the dominant coefficient across all scales and orientations is computed at every location for both transforms (up-scaling coefficients from lower levels as necessary). Coefficients with ILP phase of less than zero are discarded to remove edges from

further consideration (if present, the pectoral muscle in particular would confound the next step). Finally a histogram of ICP orientations, weighted by ILP magnitudes, is computed for the retained coefficients. Where a region has a dominant orientation, the histogram will be peaked.

Figure 8.16 Top row: Misaligned orientation in the mass and target regions; bottom row: Rotating the mass background to match up orientations (schematic regions)
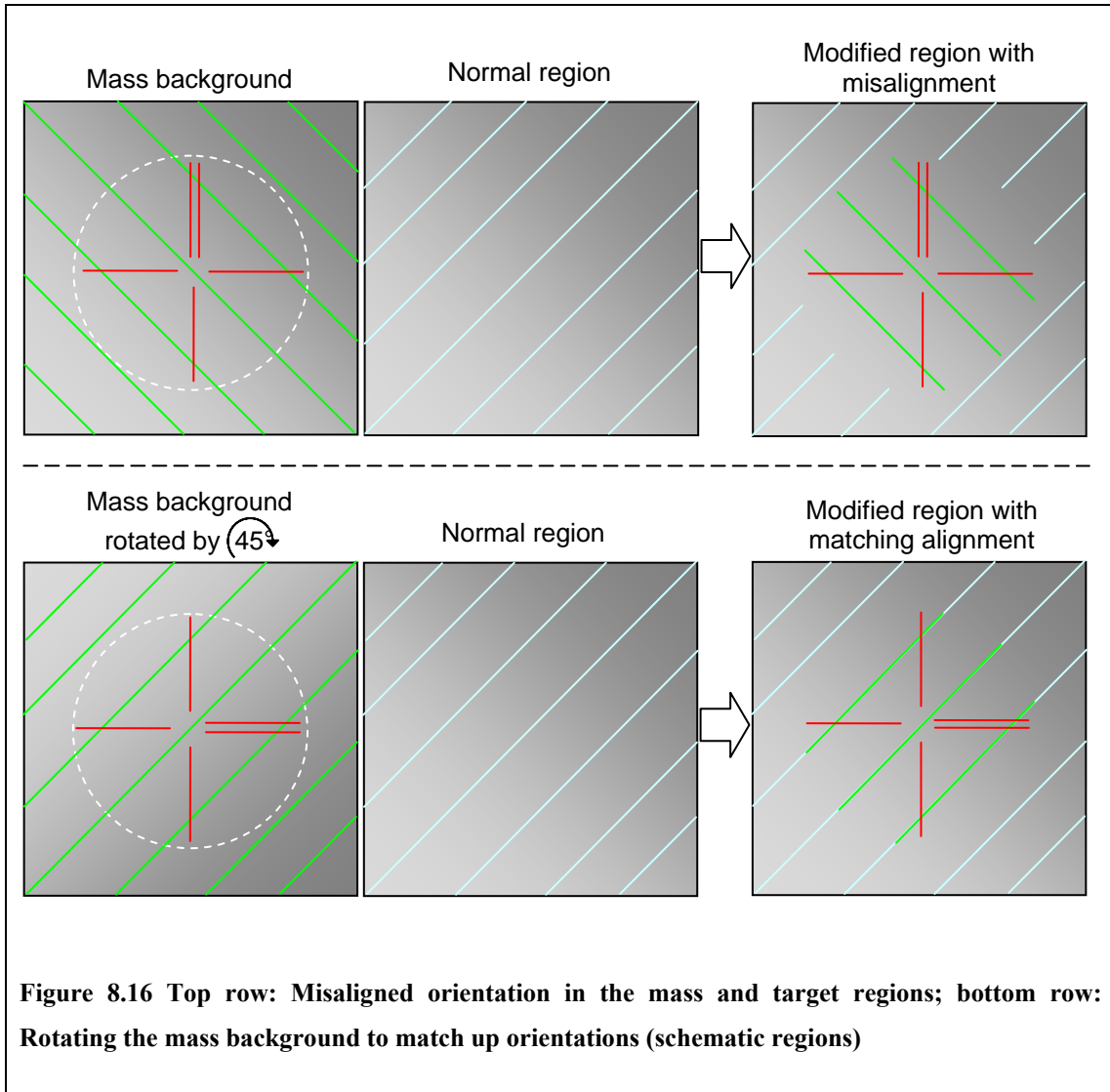
To match the orientation between a mass background and a target region we compute orientation histograms for both regions, each containing 360 bins. Having computed both histograms, we circularly permuted the mass background histogram, one bin at a

time, at each step calculating the cross-correlation with the target region histogram. The shift that produced the highest correlation across all 360 bins was taken as the angle by which to rotate the mass background.

Figure 8.17: Orientation histograms for the normal region and mass region depicted in Figure 8.15. The histograms are matched to compute the most suitable rotation for the mass background. The right-hand image shows the resulting modified region following this alignment regions
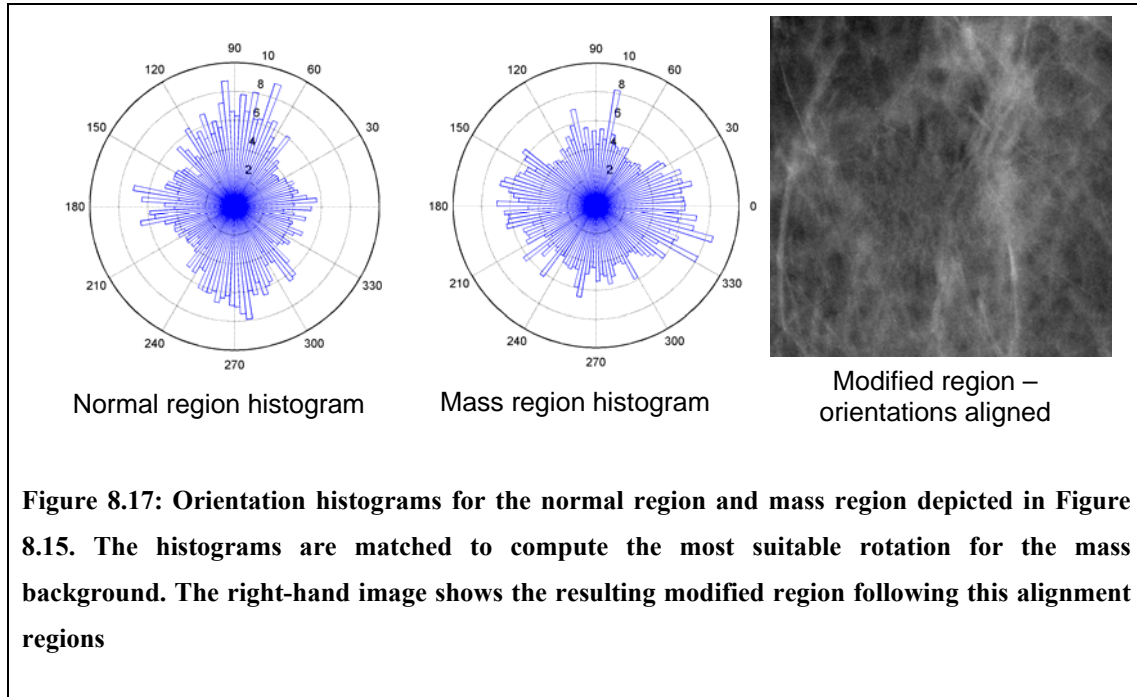
The orientation histograms of the normal region and mass background depicted in Figure 8.15 are shown in Figure 8.17. These are displayed as rose plots to make clear the circular nature of the bin centres. For these two regions, an angle of 102º was calculated as the best angle by which to rotate the mass background. The result of applying this rotation prior to the transfer process is shown to the right of the two histograms in Figure 8.17. The modified region in this case appears much more realistic than the case prior to the rotation, depicted in Figure 8.15.

8.5.2   Results

In this section we show several examples of applying the transfer process described

above. The normal regions have been randomly selected from disease-free mammograms (in the following chapter we look in more detail at how target regions should be selected).
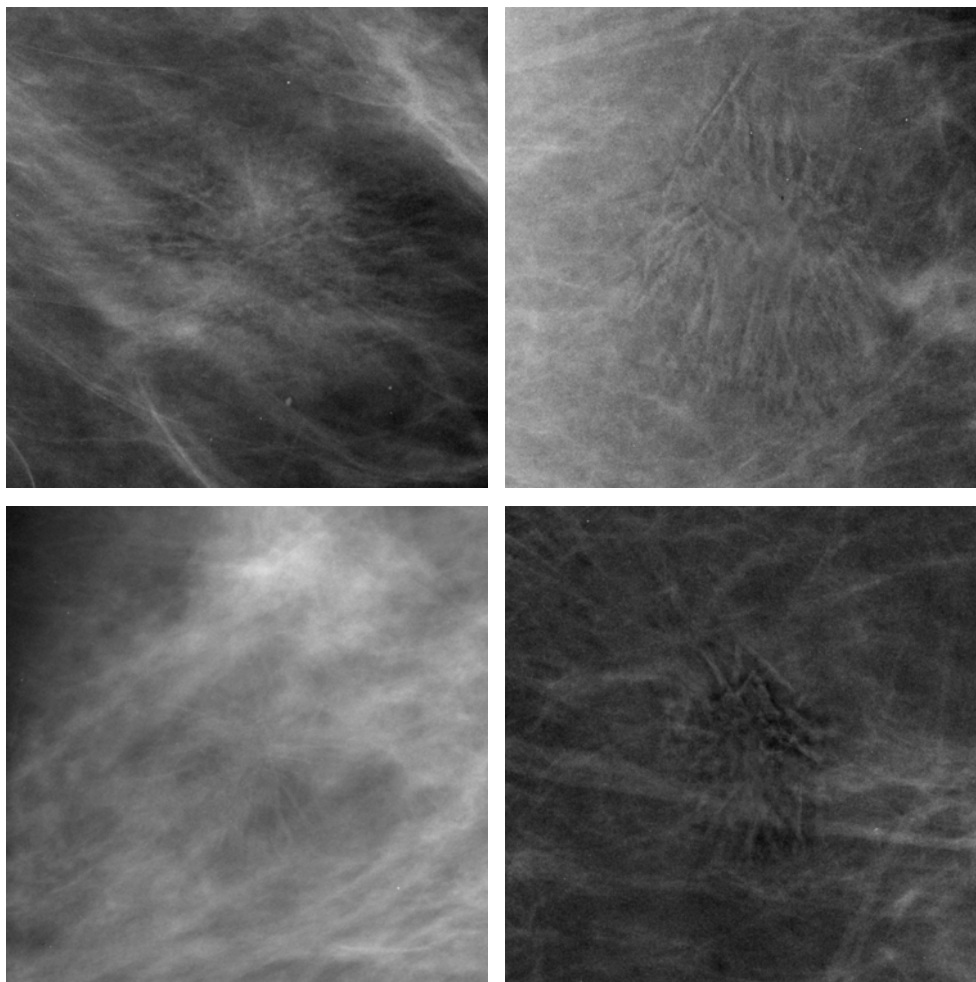
Figure 8.18: 4 normal regions with mass appearance synthesised by directly transferring DT-CWT coefficients. Subjectively, the regions on the left contain the desired appearance of a mass background; the regions on the right have unrealistic artefacts

Note we shall only qualitatively evaluate the modified regions at this stage. A quantitative evaluation of regions produced by the method (with the addition of a

synthesised mass) will be reported in chapter 11. For now we judge how well a region has been modified by assessing whether any artefacts have been introduced into the region. We know the method will transfer the appearance of abnormal structures from the mass background; however it is apparent that additional unwanted modifications can be made to the target region. For example, there may be an overall blurring to the region, or there may be areas of texture generated that clearly do not match the appearance of mammogram tissue in the remainder of the region.

Figure 8.18 shows four previously normal regions that have been modified to appear like mass backgrounds using the direct transfer method. The regions have been modified with varying degrees of success, as assessed subjectively. In the left column, we show two regions exhibiting mass background appearance with no noticeable areas of unrealistic texture. In contrast, the two regions on the right contain obviously identifiable areas of unrealistic texture. We will discuss how and why these artefacts occur in the next section

8.5.3   Discussion

We have described a method for modifying normal mammograms so that they match the appearance of mass backgrounds. This comprises directly transferring DT-CWT coefficients encoding abnormal structures from a real mass background to the DT-CWT of the normal region.

This process was a useful proof-of-concept: it showed that our basic three stage approach (1. Compute DT-CWT of target region 2. Modify DT-CWT coefficients 3. Reconstruct region) was capable of producing mass background-like appearance in previous normal regions.

However the direct transfer of DT-CWT coefficients from an existing mass background has several shortcomings we would like to improve upon.

Firstly, the method is capable of generating regions containing artefacts that would instantly identify the final mass background as synthesised. There are two main reasons such artefacts occur. Most obviously, the second part of the assumption stated at the start of 8.5, (that coefficients in the transferred region do not encode significant other structures or texture that will produce unrealistic appearance in the reconstructed target region) does not hold. For example, if the chosen mass background is from a particularly striated mammogram, then the DT-CWT coefficients transferred encode all of the high-frequency texture, reproducing the texture in the normal region. Of course, if the rest of the normal region is fatty, then having a large patch of high-frequency, oriented texture (in addition to any abnormal structures we wanted to transfer) will not look real. Conversely, transferring from a fatty mass background into a fatty-glandular normal region will produce a conspicuously smooth area in the centre of the modified region.

To understand a second, more subtle, cause of artefacts in the modified regions, we recall the theory on the inter-level and inter-coefficient products of the DT-CWT discussed in section 8.4. Remember that structure type and orientation could be inferred by analysing changes in local phase (either spatially or across frequency levels). In directly transferring DT-CWT coefficients from one region to another, we maintain the local phase relationships between all internal locations in the transferred region (both spatially and between levels). However there may be mismatches in phase between transferred and existing coefficients, either spatially at the borders of the transferred

region within each subband, or between the transferred coefficients in the 5th level and the existing coefficients in the 6th level. These discontinuities in phase may generate spurious structures in the reconstructed region.

Aside from artefacts introduced in modified regions, there is another major shortcoming in directly transferring appearance from existing mass backgrounds: we are limited by the set of mass backgrounds we have. Compare the situation to generating synthetic masses in the previous chapter. There we used the finite set of masses to learn a model of the appearance of the general population of masses. As a result we were able to generate a potentially infinite set of unique synthetic masses that match the appearance properties of the real set. Here we can only ever generate mass background appearance derived directly from a real example. Of course even using the same mass background the final appearance of two modified normal regions will differ because of the differences in the normal tissue, however we cannot claim to be able to generate the full spectrum of abnormal mass background appearance.

This problem is arguably more serious than that of generating artefacts in the modified regions. It may be possible to design a scheme to remove artefacts – for example by analysing and matching the texture of normal/mass background pairs prior to transferring the DT-CWT coefficients and then applying a regularisation step to remove discontinuities in local phase after the transfer. However the direct transfer method will always be limited by the finite set of training masses from which appearance can be transferred. For this reason, we propose learning a model of mass background appearance, which can then be used to generate a unique modified appearance in target normal regions. Such a model is described in the following section.

8.6 Synthesising mass background appearance from texture models

We saw in the previous section that by modifying the coefficients of the DT-CWT of a normal region we could replicate the appearance of a mass background. Coefficients were modified from the 5th level of the transform upwards, so that high-frequency structures associated with a mass were generated, whilst maintaining the underlying appearance of the normal region. However replacing the coefficients in the normal region with coefficients directly copied from an existing mass background did not always produce realistic results. Moreover, by directly transferring appearance from a real mass background the set of possible regions we can generate is limited by the finite set of mass backgrounds in the training data.

Thus instead of directly transferring DT-CWT coefficients, we require a method for generating new DT-CWT coefficients that match the properties of those from real mass backgrounds, without being exact copies of them. In addition, the coefficients generated must match the coefficients that aren't replaced in the normal region – both at the borders of the modified region and from 6th level of the transform downwards. Without this criterion, artefacts of unrealistic texture will appear in the reconstructed region, as we saw in Figure 8.18.

Our solution is to learn a statistical model of the DT-CWT coefficients from real mass backgrounds. Like the model for masses described in the previous chapter, this will govern the way in which the coefficients can vary, so that samples randomly drawn from the model are unique whilst sharing the properties of the real data used to train the model.

However, applying the same form of model as used with masses will not work with mass backgrounds. In the model of mass appearance each mass was ultimately described by a single appearance vector and thus a new unique mass could be generated by randomly sampling a single vector from the model. We seek an analogous method for modelling mass backgrounds, but unlike a mass, a mass background cannot be thought of as a single entity with a shared set of appearance correspondences across the dataset. Therefore it is not possible to fit a global appearance model to the set of mass backgrounds. Instead, we turn to the texture synthesis literature, and look at models that can describe appearance locally on a patch-by-patch basis.

8.6.1   Local texture modelling

In section 4.4 we reviewed previous work in texture synthesis, notably the work by Efros and Leung [40] and Rose [98]. Given the importance of these algorithms in influencing the synthesis methods developed in the remainder of the chapter, we briefly review both algorithms again below.

Efros and Leung's algorithm can be seen as the basic template for local patch-based texture synthesis. Given a sample texture, all possible square symmetric patches are extracted and stored. Given an initial seed (either a patch sampled randomly from those stored or a partially filled image), new texture is generated by adding pixels successively to the image until all pixels are filled. For each unfilled pixel, the surrounding partially filled patch is extracted and the distance between this patch and all stored exemplar patches calculated. The central pixel of the best matching exemplar is then added to the new image. Despite its simplicity, the algorithm has been shown to be one of the most effective and flexible texture synthesis algorithms.

Rose [98] produced an algorithm that adapted the main structure of the Efros and Leung algorithm to a generative probabilistic framework. Instead of being stored as exemplars, the patches in the sample texture were treated as feature vectors in some multi-dimensional space. A Gaussian Mixture Model was fitted to model the distribution of all such vectors. During synthesis, partially filled patches were extracted about a central unfilled pixel, but rather than searching for the closest matching patch from the training set as in Efros and Leung's algorithm, the filled pixels were used to condition the learnt distribution, allowing values for the unfilled pixels to be randomly sampled from the model.

One advantage of the probabilistic approach for synthesising mammograms is that adding more feature vectors does not increase the computational expense of synthesis. In the exemplar based algorithms such as Efros and Leung's, if the set of patches increases (either by using a larger training image or selecting multiple training images), then the additional patches must be checked for a match at synthesis time. This makes generating texture from a large training set of images infeasible. In contrast, whilst adding more feature vectors increases the cost of learning the model distribution, the cost is not passed on to synthesis. Because the modelling stage need only be performed once, this makes using a large training set of examples possible. This is important for mammograms where a single region is unable to capture the full spectrum of mammographic texture.

The algorithm developed by Rose suggests a local patch-based probabilistic model could produce realistic regions of mammogram texture. However, by operating directly in the image domain, the algorithm had limitations when filling in regions of an existing

mammogram. For the final region to appear realistic, the overall (low-frequency components) appearance should be maintained, with new detailed structure and texture (high frequency components) added. For a particularly clear example, consider generating texture in a region overlying the edge of the pectoral muscle, as shown in the left of Figure 8.19. If the synthesis is successful, in addition to generating realistic mammographic texture, the near straight line of the pectoral muscle edge should be maintained in the region. In right of Figure 8.19, we show the result of applying our implementation of Rose's algorithm. Here we trained the model on a set of 10 regions each of which overlapped the pectoral muscle sampled from normal mammograms. Note how despite successfully generating realistic local texture, the underlying structure of the pectoral muscle is lost, even though pectoral regions were used as training data.

Figure 8.19: Synthesising mammogram texture using Rose's algorithm for local image patches. Realistic texture is created but the structure of the pectoral muscle is lost.
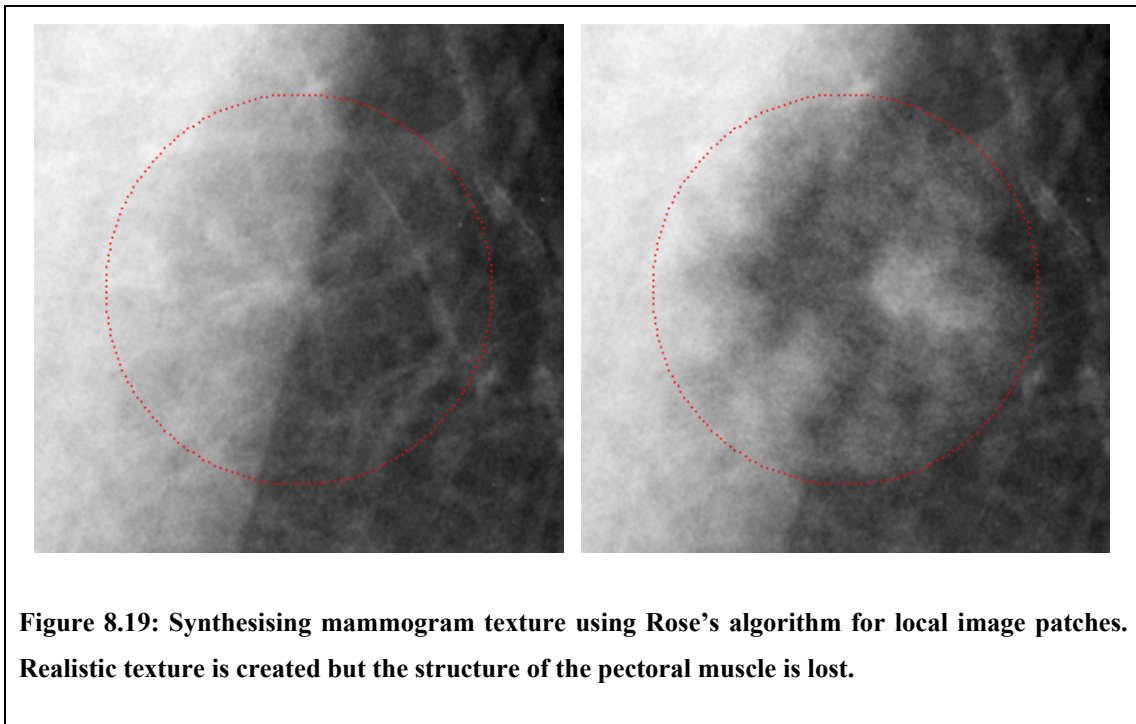
This example provides the motivation for computing a multiscale decomposition of the

training images before applying a probabilistic texture synthesis algorithm. Firstly, if we are using the synthesis algorithm to modify regions as we describe in this chapter, then the low frequency coefficients of the region need not be discarded. By choosing to only generate new coefficients above some frequency level in the decomposition, we guarantee the overall appearance (such as the structure of the pectoral muscle) will be maintained in the reconstructed region.

Moreover, rather than attempting to model all image frequencies at once (as we do when working directly in the image domain), we can apply separate models to each level of the decomposition. Effectively each model has a simpler task, because it needs only to learn the textural variations of a particular scale. Thus the set of individual models combined should provide a better description of the data than a single model encompassing all scales. Subsequently, the combined models should produce more realistic synthesis results.

A similar argument applies for applying a texture model in conjunction with an image decomposition across multiple orientations. Of course there are no orientations for which we can simply use the existing coefficients, but explicitly separating responses to structures into oriented subbands gives the texture models a better chance of learning the important variations in the data.

For these reasons we propose a texture synthesis algorithm that combines the generative probabilistic model used in Rose and Taylor's algorithm with the DT-CWT. Instead of training models on a set of images, we first compute the DT-CWT of each image. Separate Gaussian Mixture Models are then fitted for each frequency level of the transforms. Subsequently, rather than building a single local texture model describing

the probability distribution of image patches, we build multiple local models of patches of DT-CWT coefficients[9]. The composition of the DT-CWT patches reconstructs image texture. Thus by modelling the DT-CWT coefficients, we are ultimately modelling the texture of the mass backgrounds in our training set.

8.6.2   Choosing data to model

Before we can build local texture models of DT-CWT coefficients, we must first decide what coefficients should be included in the patch for a given location and scale. Clearly this is more complicated when using a multiscale multi-orientation decomposition than it is when working in the image (where a patch is simply defined as square set of pixels about a central location).

Firstly we look at the direct analogue to an image patch and consider a local patch of coefficients within a single subband. As with the algorithms in the image domain, we take this local patch to be a square set of coefficients about a central location. To align the central location with a coefficient location the square must have odd-dimensions. The size of the patch must be large enough to capture the structure of the texture being modelled. However, unlike working in the image domain, we do not have to capture the structure of all frequencies in a single patch. This allows us to select smaller local patches within each subband. For example, a 5×5 patch of coefficients in the 4th level of a DT-CWT has an area of 80×80 pixels relative to the image. Moving up through finer scales, by the time we reach the top-level of the transform, the lower levels should have

---

[9] Note that a local patch of DT-CWT coefficients may in fact comprise multiple patches sampled at the same location across different subbands. For convenience we refer to this as a single patch.

captured long-range interactions in the image. Therefore an equivalent 5×5 patch of coefficients in the top-level should suffice to capture the local high-frequency structure. Note that capturing the long-range interactions between structures in the mass backgrounds is vital for the synthesis model to succeed. This is because we must learn not just how mammogram texture appears, but the unique appearance of texture and structure in mass backgrounds that distinguish them from normal mammogram regions.

To test the theory that 5×5 patches are sufficiently large, we look at the (linear) correlation between DT-CWT coefficients at each location in a square patch and the central coefficient for a large set of patches drawn randomly from the training data. In Figure 8.20, we show plots of these correlations for a single orientation band within the top four transform levels. The colour of each square in a plot depicts the magnitude of the correlation coefficient (the covariance divided by the product of the standard deviations) between that location and the centre. The colour scheme ranges from blue (zero) to a maximum at red; the value of this maximum is depicted next to each plot (note we have set the central pixel to zero to improve contrast, although obviously the correlation coefficient here is one).

As we see below, because the complete feature is a concatenation of the elements from local patches across multiple subbands, we want to choose as small a local patch as possible. We can see that sampling a 5×5 patch about the centre coefficients should be sufficient to capture local correlations.

So far we have talked about the DT-CWT decomposing mammographic texture into subbands localised in scale and orientation. However this does not mean that coefficients within each subband can be treated independently. Thus having defined a

local patch within a single DT-CWT subband we now consider what combination of subbands should be included to form the complete feature vector at any location in a given scale. We also note that because the DT-CWT coefficients are complex, the real and imaginary parts effectively form two separate subbands at each scale and orientation.

Maximum
correlation
= 0.37

Level 1

Maximum
correlation
= 0.36

Level 2

Maximum
correlation
= 0.41

Level 3

Maximum
correlation
= 0.42

Level 4
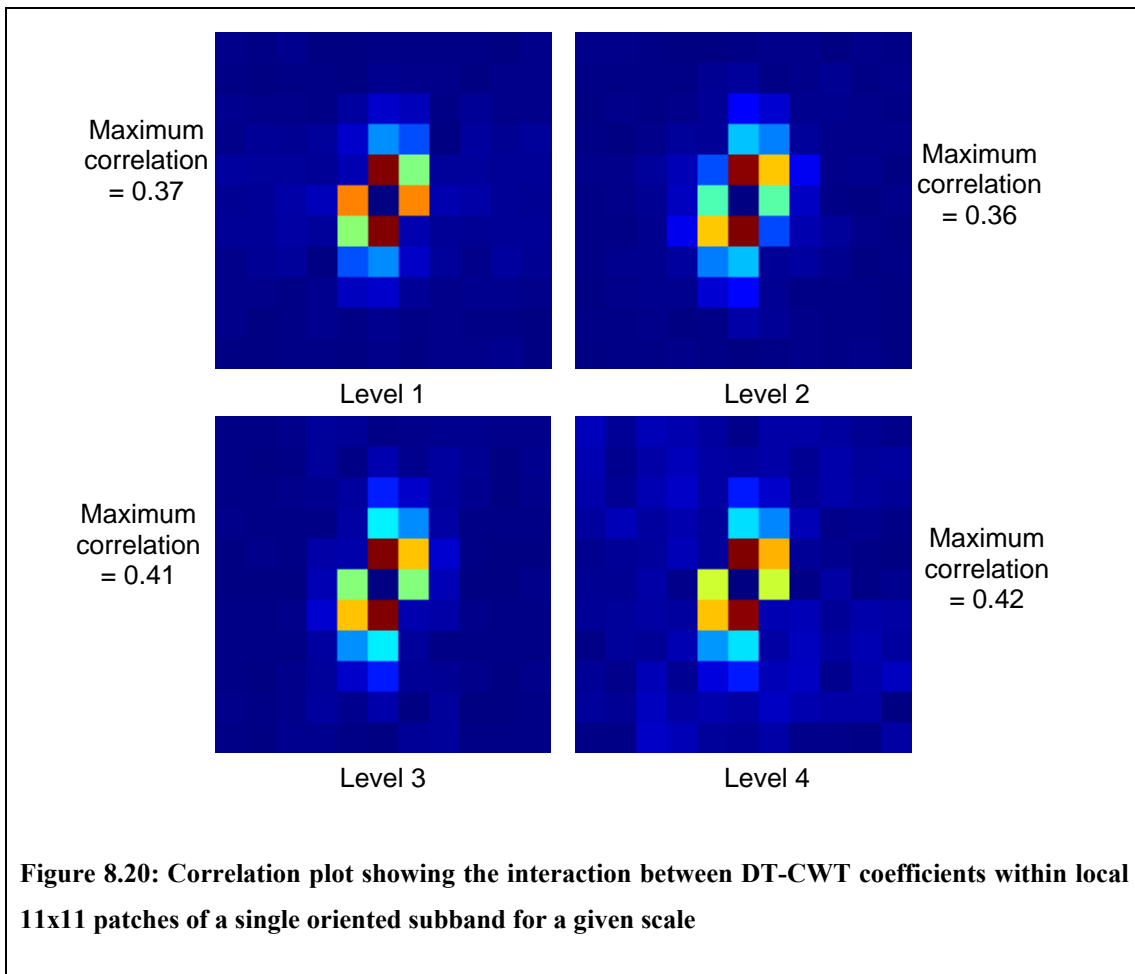
Figure 8.20: Correlation plot showing the interaction between DT-CWT coefficients within local 11x11 patches of a single oriented subband for a given scale

In Figure 8.21 we show similar plots to those in Figure 8.20. However this time the plots show the correlations between the real and imaginary DT-CWT coefficients at equivalent locations in each oriented subband for a given scale. So for example, the first row (or column, by construction the plots are symmetric) shows the extent to which a

coefficient in the first oriented subband is correlated to the real part of the other five oriented subbands, its own imaginary part and finally the imaginary part of the other five subbands (here the leading diagonal of correlation coefficients that should be one have been set to zero to improve contrast).

As we would expect there are strong correlations between adjacent orientations. For example the real part of coefficients in band 1 has strong correlations to the real part of coefficients in band 2 and the imaginary part of coefficients in band 6. Clearly our model(s) must account for these correlations, and we can do this in one of two ways:

i)   Group local patches of DT-CWT coefficients for adjacent orientation bands, (for example $\{1,2,3\}$, $\{3,4,5\}$, $\{4,5,6\}$) and fit separate models to each. At synthesis time, we would then conditionally sample groups of patches sequentially to generate coefficients across all six subbands

ii)   Group local patches of DT-CWT coefficients for all six oriented subbands and fit a single model at each frequency level

The advantage of the first approach is the greatly reduced dimensionality of the feature space produced by the combination of patches. For example, if as suggested above, we choose 5×5 local patches, then the feature vector comprising patches from all six subbands has $25 \times 6 \times 2 = 300$ elements. Adequately populating such a feature space requires a very large amount of data and in turn makes learning the distribution of the feature space difficult. Thus splitting the data into separate models may be beneficial.

However building separate models has several disadvantages. The models have a greater redundancy. For example, using the suggested groupings above, we would need

to fit to three 150-dimensional spaces as opposed to a single 300-dimensional space. The cost of this redundancy is also passed on at synthesis time when we have to condition each group of patches in turn.

Figure 8.21: Correlation plots showing the interaction between the real and imaginary parts of DT-CWT coefficients across the 6 oriented subbands for a given scale (scale levels 1 to 4 shown).
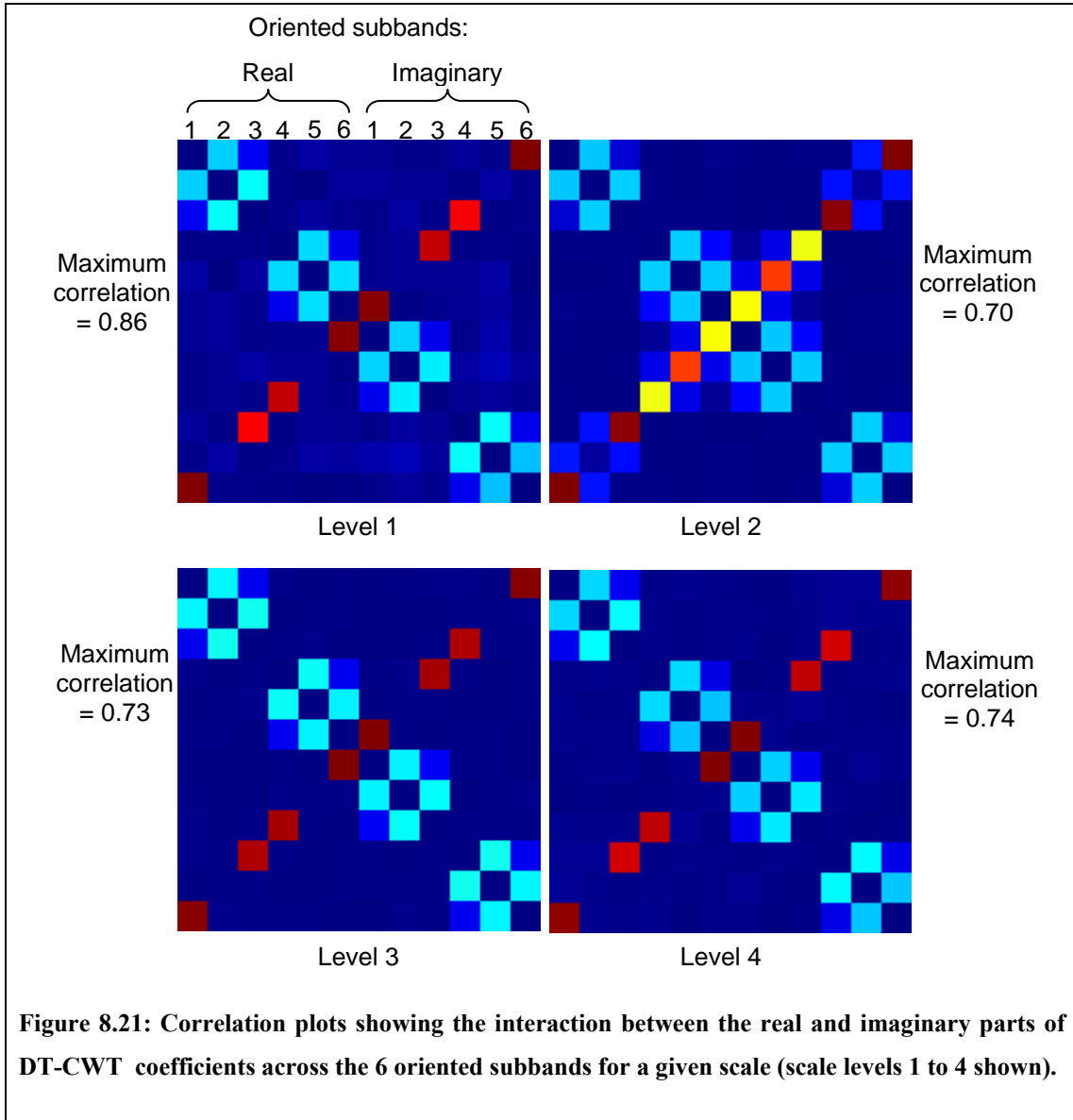
Given that we have no shortage of data and the fact that model fitting, unlike synthesis, needs only to be performed once, we choose to use the second approach. That is, we combine local patches of both the real and imaginary parts of DT-CWT coefficients

from all six oriented subbands for the model at each frequency level. Other possible ways of generating feature vectors from the DT-CWT coefficients of a particular level are discussed in section 8.6.6.

The final step in choosing what data we model is to consider correlations across scale. Figure 8.22 shows the correlation between DT-CWT coefficients at different scales given a fixed location. The correlations have been computed for both real and imaginary parts of DT-CWT coefficients sampled from all six oriented subbands. We see there is evidence that coefficients between frequency levels are correlated, and as we would expect, the strongest correlations are between adjacent scales. However the magnitudes of the correlations are approximately a factor of 10 smaller than both the correlations between spatially adjacent DT-CWT coefficients and correlations across orientation.

Figure 8.22: Correlation plot showing the interaction between DT-CWT coefficients across scales 1 to 5
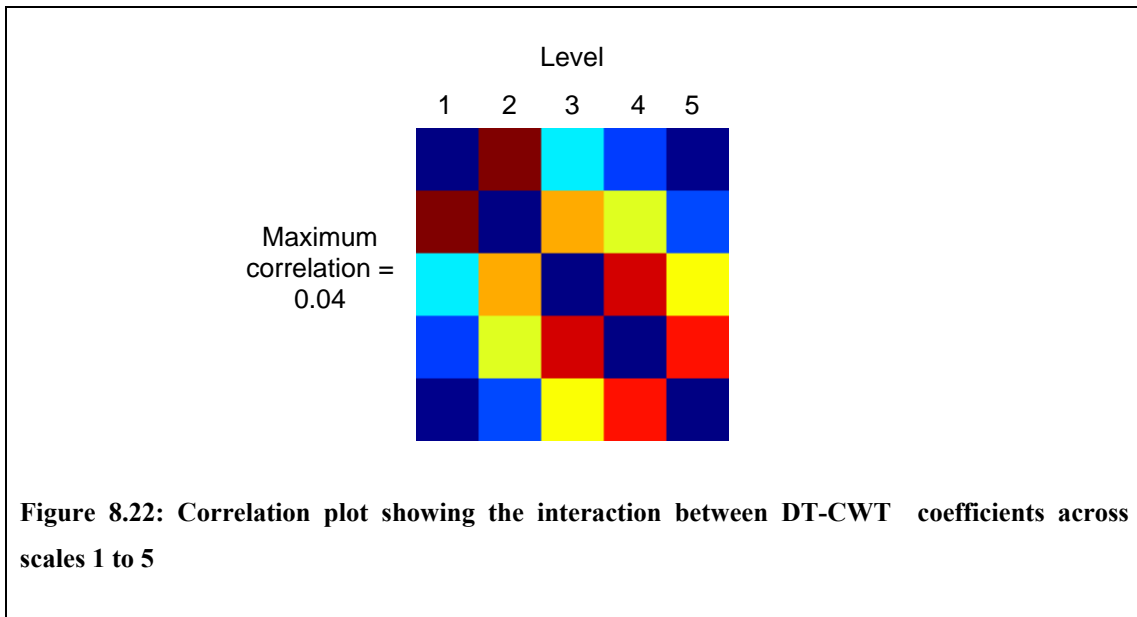
We have already stated our intention to build separate texture models for each frequency level, but we do not want to ignore the interaction of coefficients between scales. In particular, we have already seen that linear structures persist across scale

in a DT-CWT, and therefore the coefficients that generate such features must be correlated across scale.

Thus, to each local patch from an oriented subband we append the coefficient sampled at the centre location of the patch from the next coarsest DT-CWT level. Note that as we saw in the construction of the ILP in section 8.4.3, the coefficients in the coarser level must first be interpolated to lie on the same grid of locations as the level above.

We summarise this section by giving a formal labelling to the composition of the feature vector as defined above.

For a given location (x, y) in the L-th level of a DT-CWT comprising subbands $B_{L,\theta}$, we extract the feature vector $\mathbf{x}$ as:

$$\mathbf{x} = \left\{ Q_\theta, Q_\theta^{'} \mid \theta = \pm 15°, \pm 45°, \pm 45° \right\}$$

$$Q_\theta = \left\{ B_{L,\theta}\left(x \pm u, y \pm v\right) \mid u,v = 0,1,2 \right\}$$
$$Q_\theta^{'} = B_{L+1,\theta}^{'}\left(x, y\right)$$

Where Q_θ are the local patches of coefficients sampled from oriented subbands within level L and $Q_\theta^{'}$ are the oriented coefficients sampled from the next level ($B_{L+1,\theta}^{'}$ being the interpolation of $B_{L+1,\theta}$). As a result of this construction, each feature is a 312 element vector.

 In the next section we describe how we can model the distribution of all such vectors in each level.

8.6.3   Fitting a Gaussian Mixture Model

In the previous section we defined how we form feature vectors to describe scale-specific patches of DT-CWT coefficients. Having extracted all such vectors from the training set transforms, we have populated a large multi-dimensional feature space for each of the five highest-frequency scales of the transforms. We now attempt to model the distribution of these feature spaces.

To each feature space we fit a Gaussian Mixture Model using the k-means clustering algorithm [76]. This was chosen because of it had been successfully used in similar texture modelling applications [92, 98]. As a result of applying the algorithm to a feature space, the probability distribution at each point in the space is defined as the weighted sum of k multidimensional Gaussian kernels. Thus for any feature vector $\mathbf{x}$ in this space, the probability of $\mathbf{x}$ can be written as

$$P(\mathbf{x}) = \sum_{i}^{k} P(\mathbf{x} \mid i) P(i)$$

where i indexes the i-th Gaussian component and $P(\mathbf{x} \mid i) \sim N\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$. The mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$ of each Gaussian component are computed as the sample mean and covariance of each cluster returned from the k-means algorithm, whilst the prior probabilities for each component $P(i)$ are computed as the proportion of the total points assigned to each cluster.

In selecting k, we must make sure we choose a value high enough to allow the model of combined components to be rich enough to describe the complexities of the distribution. In addition, the k-means algorithm can be run so that unnecessary clusters (that is, one to which no – or very few – points are assigned) are dropped, so a conservatively large value of k can be chosen. To assess this, models were built for each level using

increasing values of k (starting at $k = 10$ and increasing in increments of 10) until at least one cluster was dropped during the algorithm. Because the k-means algorithm is susceptible to getting stuck in local minima, having selected a value of k, we repeated the modelling 100 times at each level, choosing a different randomly selected initialisation of clusters for each repeat. Within each level, the model that returned the most compact clusters was selected.

We label the model fitted to data extracted from the L-th level of the transforms G_L ($L = 1,...,5$). In the next section we describe how the models can be used to synthesise appearance in target regions of normal mammograms.

8.6.4   Sampling new texture from the model

In this section we show how the Gaussian Mixture Models of DT-CWT feature vectors can be used to synthesise new coefficients and subsequently modify the appearance of normal regions.

The first step is to define a region in a dual-tree transform within which to generate new coefficients. In the direct transfer method described in section 8.5 the region was based on the extent of abnormal structures in the real mass background transferred. Of course we no longer have a template real mass to base this choice on. Instead we simply define the region with respect to a circle in the image domain. This circle must be large enough to encompass all abnormal structures we wish to generate. Because this region is analogous to the dilated region of mass pixels used in Algorithm 8.1, we label it $M_d = \{(x,\, y) \in \mathbb{Z} \times \mathbb{Z}\}$.

The algorithm for modifying the appearance of a normal region N is given below. It is

initialised by decomposing N into a DT-CWT comprising subbands $\{N_{L,\theta}\}$.

---

for levels L = 5, 4, 3, 2, 1:

    Adjust the co-ordinates of the region to be synthesised to account for the factor of two down-sampling between DT-CWT levels

$$M_{d,L} = \left\{ \left(\left\lceil \frac{x}{2^L} \right\rceil, \left\lceil \frac{y}{2^L} \right\rceil \right) \mid (x, y) \in M_d \right\}$$

    for each orientation θ :

        Interpolate the coefficients in $N_{L+1,\theta}$ to lie on the same locations as $N_{L,\theta}$

        Discard the existing DT-CWT coefficients in $N_{L,\theta}(x, y)$, $(x, y) \in M_{d,L}$

    end

    for each $(x, y) \in M_{d,L}$, moving from the edge of the unfilled region inwards:

        Construct a feature vector $\mathbf{x}$ (see Algorithm 8.3 below)

        Sample the unknown values of $\mathbf{x}$ by conditionally sampling from the corresponding level model G_L given the known values of $\mathbf{x}$

        Reform the now complete windows $\{P_\theta \mid \theta = \pm 15, \pm 45, \pm 75\}$ and reinsert into each subband $N_{L,\theta}$

        Remove any coefficient locations from $M_{d,L}$ that have just been filled

    end

end

Algorithm 8.2: Synthesising mammographic texture using local models of DT-CWT coefficients

---

When all subbands $\{N_{L,\theta}\}$ are filled, we can reconstruct the synthesised mass background by inverting the DT-CWT. Note to sample each new coefficient in Algorithm 8.2 we need to extract a feature vector from the partially filled DT-CWT subbands. Algorithm 8.3 describes how this is achieved for a location (x, y) in the L-th level of a transform.

---

for each orientation θ

    Extract dual-tree coefficients from the 5x5 patch surrounding (x, y):

$$Q_\theta = \left\{ N_{L,\theta}\left(x \pm u, y \pm v\right) \mid u, v = 0,...,2 \right\}$$

    Having discarded the existing coefficients above, Q_θ will only be partially filled,

    with $N_{L,\theta}(x, y)$ in particular unknown.

    Extract the interpolated dual-tree coefficient from the next coarsest level:

$$Q_\theta' = N_{L+1,\theta}\left(x, y\right)$$

    Since $N_{6,\theta}$ remains unchanged, and synthesise proceeds from 5th to 1st level,

    $N_{L+1,\theta}\left(x, y\right)$ will always exist

end

Concatenate all the sampled windows $\left\{ P_\theta, Q_\theta \mid \theta = \pm 15, \pm 45, \pm 75 \right\}$ into a single sample

vector $\mathbf{x}$

Algorithm 8.3: Constructing a partially filled sample vector about each coefficient during the synthesis method described in Algorithm 8.2

8.6.5   Synthesis results

In Figure 8.23 we show four examples of regions modified using the Algorithm 8.2. The normal regions that have been modified are the same regions that were modified in Figure 8.18. A central circular area has been synthesised in each region – this area is marked by the dashed yellow circle in the top left image.

As with the results presented in section 8.5.2, the regions are assessed visually in the following section but not evaluated quantitatively.

8.6.6   Discussion

It is interesting to note that the modified regions generated from the model-based synthesis method described in 8.6.4 suffer from the different problems of those generated from the direct transfer method described in section 8.5. That is, the regions produced by directly transferring DT-CWT coefficients displayed the kind of abnormal structures (such as spicules) we would expect to find in a mass background but in addition contained unrealistic texture and image artefacts. In contrast, the regions produced by synthesising new DT-CWT coefficients appear to be free of image artefacts but do not convey the appearance of mass backgrounds we were hoping to achieve. The images exhibit the appearance of mammographic regions in which no clear linear structures are visible. Thus we have not managed to generate high-frequency mammogram structures such as spicules or indeed some of the finer structures in the parenchyma.
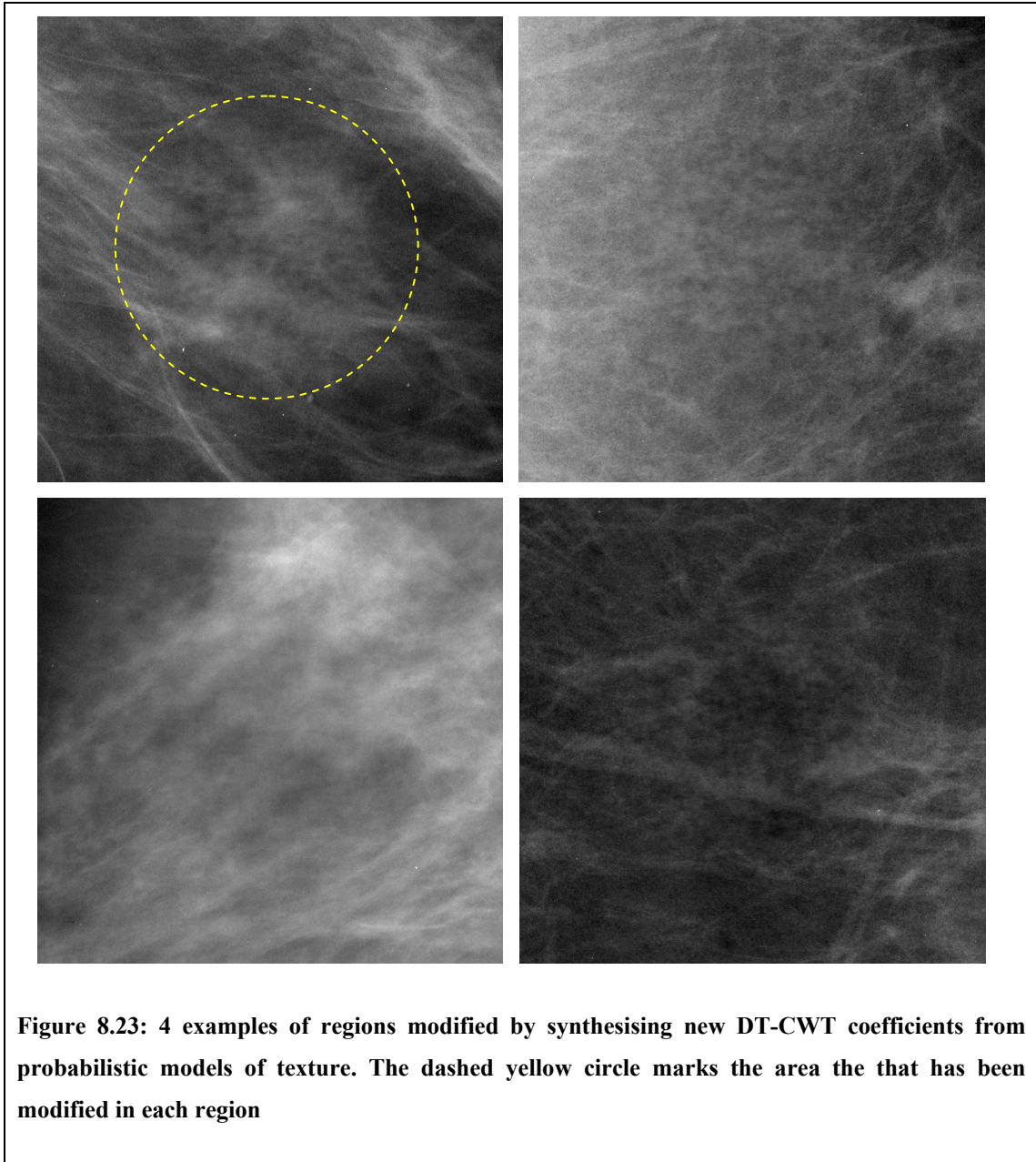
Figure 8.23: 4 examples of regions modified by synthesising new DT-CWT coefficients from probabilistic models of texture. The dashed yellow circle marks the area the that has been modified in each region

Below we discuss what has caused the failure of the synthesis model to produce structures in the modified region. We consider two contributing factors to the failure:

1) The local models are not generating patches of DT-CWT coefficients that correspond to mammographic structures (referred to below as the

correspondence problem)

2) The local models are generating DT-CWT patches that correspond to structures, but not with the correct spatial arrangement to generate structure when the transform is inverted (referred to below as the spatial alignment problem)

As an analogy, if the set of DT-CWT subbands was thought of as a jigsaw puzzle, then problem one states that we're not producing the right jigsaw pieces. In contrast, problem two states that we're able to produce the right pieces, but we're not putting them together in the correct order. We consider each problem in turn below.

The correspondence problem

We must first consider whether we have adequately modelled the probability distribution of the feature vectors within each frequency level. Each feature vector consists of 312 elements producing a complex multi-dimensional feature space. We expect feature vectors corresponding to different local patterns of mammogram regions (for examples segments of linear structures) to lie in clusters within this space. In fitting a Gaussian Mixture Model we hope to have found the centre of these clusters and in doing so describe the complete probability density of the feature space. However there are several ways the model fitting may fail. These include:

- The k-means algorithm getting stuck in local minima. This can be mitigated by repeating the clustering with different random seeds and selecting the result that produces the most compact cluster (as described in section 8.6.3)

- There not being enough feature points to adequately populate the feature space. Certainly with models fitted to coefficients from the higher frequency levels this

shouldn't be an issue, however because of the down-sampling in the DT-CWT we have much less data to work with in the lower frequency levels. If the data doesn't adequately populate the feature space we cannot learn the probability density throughout the space

- The feature vectors within the space may lie on a complex non-linear manifold, the density of which cannot be adequately modelled by mixtures of Gaussian components

For all of the points above (and especially the first two), model fitting would be easier if the number of dimensions in the feature space was dramatically reduced. Note that the synthesis method based on the models requires an explicit calculation of the joint probability between dimensions in the space, and therefore simply applying dimensionality reduction techniques (such as PCA) to the data prior to model fitting will not work. Instead, a better solution would be to build a more compact representation of the local patches of DT-CWT coefficients.

For example, we currently do not make the best use of the orientated subbands of the DT-CWT. The appearance of high-frequency structures in mammograms is not dependent on orientation, so the cross-sectional shape and intensity of a spicule is not affected by its orientation. As a result, at any location in the vicinity of a structure, the pattern of DT-CWT coefficients across the six subbands with respect to the maximal subband should be the same whatever the structure's orientation. Note this isn't true for lower frequency levels where the appearance of structures may depend on the global alignment in the mammogram. This means we should be able to construct rotation invariant features vectors. Although not used for synthesis, the work of Kingsbury [66]

has shown one way of constructing such features from DT-CWT coefficients.

Using rotation invariant features may not reduce the size of the feature vectors (and in fact we would have the additional task of recording the orientation elsewhere) but it should compact the space they lie in. In the current models, if we took six copies of a particular structure and aligned each to one of the six subband orientations then the feature patches we extract corresponding to the structure would lie in six different parts of the feature space. However, using rotation invariant features would result in just one part of the feature space being populated. Thus different parts of the feature space should correspond to different structure types, as opposed to the same structure at different orientations. This should make modelling the feature space easier.

In addition, we may be able to use the sparseness of the DT-CWT to a greater degree. Given a rotation invariant feature, it could be that all the important information regarding structure is encoded in a small number of coefficients of a local patch. As an example, Figure 8.24 shows the normal region and mass background previously depicted in Figure 8.10 and Figure 8.11. For each region we computed its DT-CWT, set all coefficients except the maximum across the six orientation bands to zero, and then inverted the transform. The resulting reconstructions of the regions are also shown in Figure 8.24.

Figure 8.24: Left column: the mass background and striated normal region depicted in figures 8.10 – 8.13. Right column: DT-CWT reconstructions of the regions using only the maximum of the 6 orientations for each coefficient
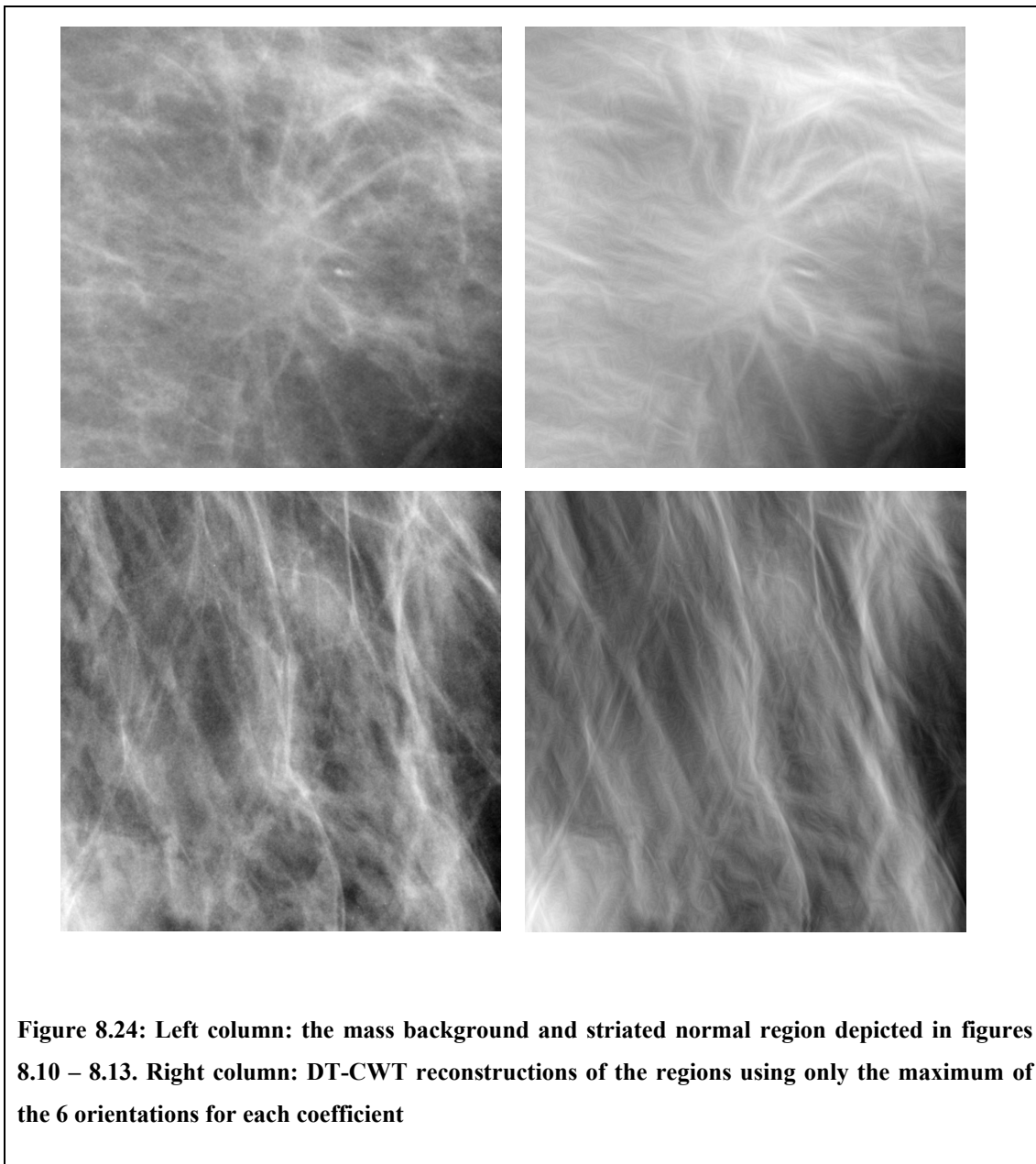
The reconstructions bear a close resemblance to the original regions, yet only a sixth of the DT-CWT coefficients we have previously attempted to model were used. Thus it may be that a model of local patches containing only the important coefficients would be enough to distinguish between the appearance of mass backgrounds and normal regions. This model would encode all the necessary information on the spatial

distribution of structure.

Of course to reconstruct a photo-realistic region we need to generate all the coefficients, but this may be achieved by having a separate model of the relationship between the important coefficients and the remaining coefficients. Because this model does not need to learn the spatial distribution of structure, feature vectors could be formed for locations as opposed to patches. Thus the combined size of the feature vectors for the two models would be much smaller than the complete 312-dimension vectors we have previously modelled.

A last point to consider on selecting feature vectors is whether transformations of the DT-CWT coefficients would be useful. For example, we saw in section 8.4.3 that in the vicinity of a structure there is a constant relationship between the phase of a DT-CWT coefficient and double the phase of the corresponding coefficient in the next coarser level. Thus modelling may be easier if DT-CWT phase was replaced by an inter-level phase difference in the feature vectors we extract from the data. Such a transformation would not give any further information to the model but it may cause the feature vectors to lie in a space that is more conducive to model fitting. An ideal model fitting method should be able to learn the phase relationships directly from DT-CWT phase, but as we discuss below, we do not yet have an ideal model fitting method!

Even if we do manage to construct far more compact feature vectors to encode mass background appearance, the feature vectors may still form a non-linear probability density that is not well approximated by mixtures of Gaussian components (at least without having to try and fit a prohibitively large number of components). As a topic of future work, we intend to assess whether a different method for learning non-linear

probability density functions in high dimensional spaces would be more suitable for our data. Given that a magnitude and phase based representation of our data may be most appropriate we are particularly interested in those methods that explicitly deal with circular data. This includes, for example, the work by Mardia [77] and Pennec [90]. We note however that finding a generic method for fitting density distributions to high-dimensional non-linear data remains an ongoing mission in the pattern recognition field.

The spatial alignment problem

In addition to generating DT-CWT patches that individually encode all the information required to reconstruct image structure, the models must also generate these patches in the correct spatial arrangement. If the correct spatial arrangement is not achieved, structure will not be realised when the region is reconstructed from the modified DT-CWT.

For example consider a DT-CWT patch that corresponds to a small segment of a line in a mammogram. For a line to be generated in the reconstructed region the models must produce multiple adjacent line patches aligned at the correct orientation. Our intention is that such patterns would be controlled by conditioning coefficients at each frequency level on the lower frequency levels. Thus if in a coarse level there is a 5×5 patch of coefficients corresponding to a line of some orientation, then in the next finer level there is a 10×10 patch of coefficients each of which is conditioned by a coefficient in the coarser patch. As a result the models should favour the generation of coefficients in the finer level that also correspond to a line at the same orientation. Simultaneous conditioning on neighbouring coefficients within the finer level acts to ensure these 'line' coefficients align correctly. This conditioning continues upwards through to the

finest DT-CWT levels.

Of course, this assumes we generate a suitable low frequency representation of mass background structure in the lowest level synthesised, otherwise there is no evidence of structure to pass upwards through the DT-CWT levels.

In addition, conditioning between levels enforces only a weak dependence on generating coefficients that correspond to structure. In the next section we look at whether the models can be forced to produce coefficients corresponding to structure at designated locations in the image.

8.7 Adding a model of global structure

In the previous section we described how the local texture models we have built do not adequately encode the long range interactions within an image that give rise to particular patterns of structure.

In this section we introduce a method designed to overcome this limitation.

8.7.1   Synthesising structure separately

The basis of the new method is to build local texture models for DT-CWT patches that correspond to structure, in addition to models that correspond to generic mammographic texture.

We assume that for every DT-CWT in the training data, we have a binary map for each frequency scale that specifies which locations within that scale have coefficients corresponding to structure. To build the models that correspond to structure we then simply repeat the processes described in section 8.6.3, but rather than populating the feature space with all possible feature vectors from the training data, we sample

from only those locations that correspond to structure. Similarly the models for generic mammographic texture are fitted to data sampled only from locations that do not correspond to structure.

To synthesise new coefficients we assume that we have a binary map for the target region analogous to the maps for each transform in the training data. Synthesis can then proceed as described by Algorithm 8.2. However now in each level we first sample coefficients for all locations that corresponding to structure from the appropriate structure model. We then fill in the gaps by sampling from the appropriate generic texture model, before proceeding to the next finer level. Finally, the modified transform is inverted as previously to reconstruct a synthesised mass background.

We refer to this method of synthesis as using a *hybrid model*. This is because the model combines a global map of structure with a local model of texture. In the next section we discuss how maps of structure could be generated.

8.7.2 Generating binary maps of structure

From the method described above, two questions are apparent. Firstly, how do we obtain the binary map of structure for the training data? Secondly, how do we obtain such a map for the target region?

To answer the first question, we note that we have already developed a line detection algorithm using DT-CWT coefficients, as described in section 8.4.5. This algorithm can be used to generate binary maps for each mass background in the training data

However the question of how we might obtain a binary structure map for a target region poses a much harder problem. First we consider what properties the map should have,

before discussing some theoretical ways of achieving these properties.

The binary map for the target region should contain, at each frequency level:

- all existing structure in the normal region, albeit with some of this structure distorted spatially

- additional structures corresponding to abnormal features such as spicules

As with our models of mass appearance and local mammogram texture, our aim is to be able to learn the valid set of constraints in which such maps vary, using the maps from real mass backgrounds as training data. New maps could then be sampled probabilistically from this model. To build such a model, we require a fixed length parameterisation that provides a compact representation of any given map. For the model to be generative, we must be able to reconstruct a complete mass from the compact representation.

However, because the set of structures within any one mammogram region may differ greatly, it is hard to define a set of consistent correspondences across a dataset of structure maps. As reviewed in section 3.3.1, the problem of learning patterns of mammographic structure has been studied [63, 88, 96]. We note however that previous work has concentrated on learning models to discriminate between normal and abnormal regions of mammogram, and do not provide a clear mechanism for synthesis[10].

---

[10] We will discuss the possibility of using the work developed in this thesis towards similar discriminative tasks in section 12.3

From a texture synthesis point of view, work such as that by Wu and Yu [127] has addressed the issue of combining explicit representations of structure with local texture models (see section 4.4). However in Wu and Yu's algorithm structure is imposed locally on a patch-by-patch basis in parallel to sampling from the texture model. Thus the algorithm does not provide the global representation of structure we desire.

In light of not being able to generate binary maps for target regions in a probabilistic framework, we use a simple alternative to show a proof of concept for the synthesis method described in 8.7.1. First we randomly select a mass background from the dataset. We then extract an equivalent sized region from a normal mammogram. We use the map of structure for the mass background to synthesise DT-CWT coefficients for the top five levels of the normal region, following the method described in section 8.7.1. Figure 8.25 shows synthesise results for the four normal regions that were depicted in Figure 8.23. In each case, we used the mass background depicted in Figure 8.24 as a template.

The most interesting aspect of these results is that the synthesised mass backgrounds contain no more structure than those produced using Algorithm 8.2 in section 8.6.4 (see Figure 8.23). Thus despite explicitly imposing a map of structure on the region, we are not generating the necessary patches of DT-CWT coefficients for the structure to be realised in the reconstructed region. This gives further evidence to support our belief that we have not yet solved the correspondence problem discussed in section 8.6.6.

Both the correspondence problem and the problem of generating a global map of structure for a region are the source of ongoing work, as discussed in chapter 12.
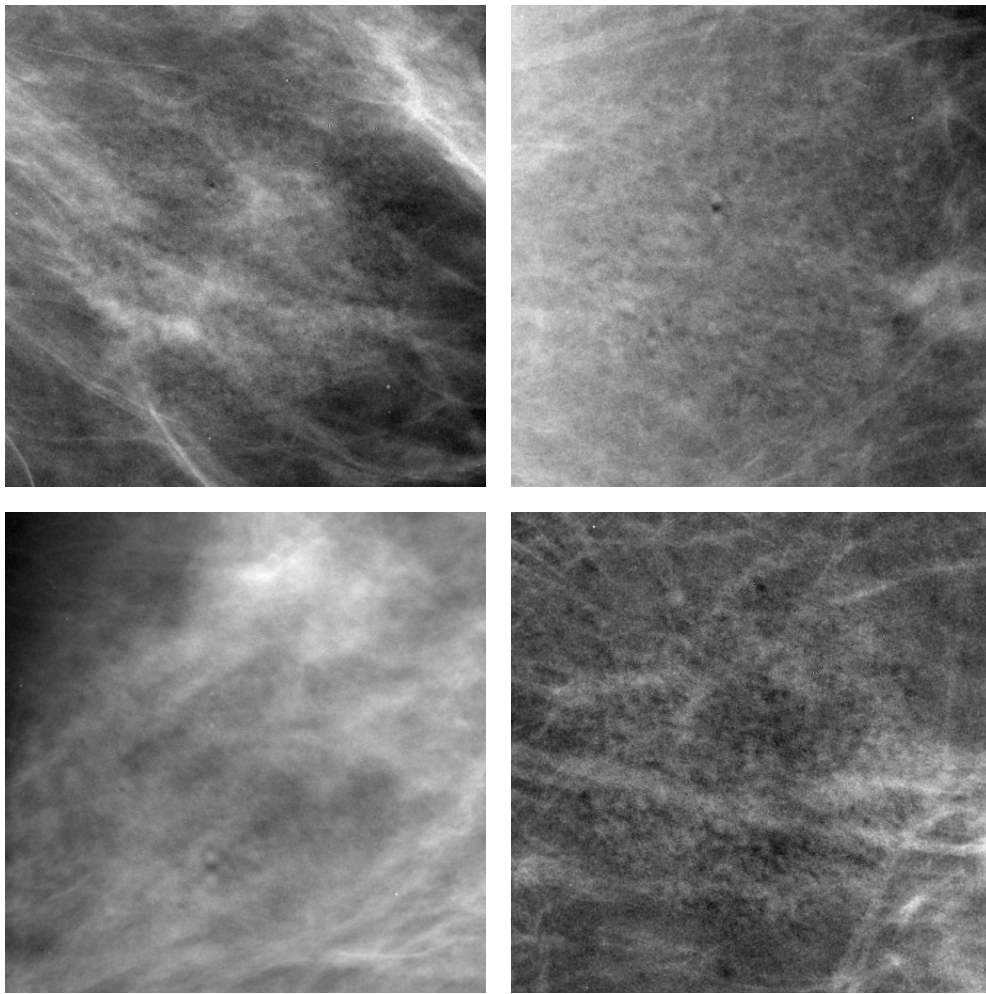
Figure 8.25: 4 examples of regions modified by synthesising new DT-CWT coefficients from probabilistic texture models. An explicit model of structure has been imposed on the regions, however this has not succeeded in generating more structure than the regions depicted in Figure 8.23

8.8 Discussion

We have discussed three different methods for generating synthetic mass background appearance in normal mammograms. In each case synthesis was achieved by modifying the coefficients of the DT-CWT of a normal region to match the properties of DT-CWT coefficients in mass backgrounds.

In section 8.5 the normal DT-CWT coefficients were modified by taking direct copies of coefficients from a real mass background. This generated the abnormal structures associated with a mass background as desired but also produced image artefacts that might betray the modified region as synthetic.

In section 8.6 we generated new DT-CWT coefficients by randomly sampling from local texture models. The models were built to encapsulate how DT-CWT coefficients in mass backgrounds were distributed on a patch-by-patch basis. By conditioning models for each frequency level on the adjacent coarser scale, the models were designed to capture the long-range interactions between structure that characterise the appearance of mass backgrounds and differentiate them from normal regions. However, as discussed in section 8.6.6, the synthesised mass backgrounds produced from the model did not contain the abnormal structures as intended. This suggested the models were inadequately learning the distribution of DT-CWT coefficients for mass backgrounds.

The synthesis method presented in section 8.7 was designed to overcome the failure to produce structure when synthesising from the texture models described in section 8.6. The method used a hybrid model that combined a global map of structure with local models of texture. The training data of mass background DT-CWT coefficients were split into two classes, those corresponding to linear structures and those not, based on the linear detection algorithm described in section 8.4.5. Separate models were fitted to each class of coefficients. To synthesis a new region, a template binary map of structure was required. Using a map taken from a real mass background to generate synthetic mass appearance in an equivalently sized normal region showed that further work was required in constructing the local texture models. In addition, we do not yet have a

method for generating new binary maps of structure that are not direct copies of real examples.

Given the potential of the methods shown in 8.6 and 8.7, our intention is to continue work in this area, implementing some of the changes suggested in 8.6.6 and 8.7.2. This is discussed further in the section on future work in chapter 12. However, to complete a quantitative evaluation of the mass synthesis method developed throughout this thesis, we require a method for synthesising mass backgrounds that generates abnormal texture associated with masses. Therefore despite its limitations, the method of directly transferring DT-CWT coefficients from an existing mass background to a normal region is the one we use to generate synthetic mass backgrounds in the remaining chapters of the thesis.

8.9 Summary

This chapter has presented a comprehensive body of work exploring the appearance properties of mass backgrounds. We have looked at how we can model mammographic texture, and in particular, how we can modify a normal mammogram region so that it matches the appearance of a mass background.

In section 8.2 we gave a brief review of previous attempts at synthesising the appearance of mass backgrounds. We noted that all previous work focused on simulating the appearance of individual spicules associated with a mass and did not propose a more detailed method for combining such structures – either with other spicules, the synthesised mass or the surrounding breast tissue. We stated our belief that limiting the focus to individual structures was not sufficient to realistically synthesise the complete appearance of malignant masses, and used this to motivate the work in the

remainder of the chapter.

Given our stated goal of modelling the complete appearance of mass backgrounds, in 8.3 we noted that the grey level pixel intensities of the mass backgrounds alone did not provide a rich enough description of the data. Instead we proposed decomposing the regions in the training set across scale and orientation. After reviewing previous work, we experimented with two wavelet based decompositions: the steerable pyramid and the dual-tree complex wavelet transform (DT-CWT). The key properties of both were that they separated images into subbands localised in scale and orientation whilst being invertible. The invertibility provided a natural path to synthesising new regions. However we observed that a real-valued steerable pyramid is not shoft invariant, whilst a complex valued steerable pyramid has a greater redundancy than the DT-CWT. Thus ultimately we chose to use the DT-CWT in the work presented in this chapter.

In section 8.4 we showed how further transformations of the DT-CWT could be used to gain greater information about structures, and particularly linear structures. We looked at the inter-level (ILP) and inter-coefficient (ICP) products, and presented our own modifications to the previously published transformations. The modified ILP and ICP transforms provided a further useful tool for modelling and analysis in the remainder of the chapter.

Having selected the DT-CWT as a suitable decomposition for mammographic regions, and developed tools for detecting linear structures, we proposed the following three stage framework for synthetically generating the appearance of mass backgrounds in normal regions:

1) Decompose a normal region into its DT-CWT

2) Modify some or all of the DT-CWT coefficients based on the distribution of coefficients in the transforms of real mass backgrounds from our training data

3) Invert the modified DT-CWT to reconstruct the now modified normal region

Given this framework, sections 8.5, 8.6 and 8.7 describe three different methods for completing step 2. As with the method for synthesising masses described in the previous chapter, all three methods for generating mass background appearance in normal regions can be run automatically, without the need for further user input for any individual region. As discussed both within the respective sections and in section 8.8, each of the three methods had its disadvantages. As a result, the direct transfer method presented in section 8.5 was chosen as the best method to use in the evaluation of the full synthesis method described in chapter 11. The more promising but complex methods based on synthesising from generative probabilistic texture models (described in sections 8.6 and 8.7) are the subject of ongoing research and will be discussed again in chapter 12.

In summary, in this chapter we have shown how we can convert regions of normal mammograms to have the appearance of mass backgrounds. These synthetically generated backgrounds complement the method for synthetically generating masses described in the previous chapter. Thus to complete our method for synthesising malignant masses in normal mammograms it remains to:

i) Match the properties of a synthetically generated mass to a synthetically modified target region

ii) Automatically select the location of a target region within a normal mammogram

In the following two chapters, we show how the two goals can be achieved.

Chapter 9 : Adding masses to mass backgrounds

9.1 Introduction

In the preceding three chapters we have presented a set of methods for modelling mammographic masses. In chapter 6, we described how each mass in a training set of data could be separated from its background, effectively producing two new sets of data: masses and mass backgrounds. In chapter 7 we showed how to build a combined model of mass appearance to describe the variation in shape, size and texture of the set of masses. In chapter 8, we investigated ways of modelling the set of mass backgrounds so that mass background appearance could be synthesised in normal mammogram regions. As a result of this work, we are able to synthetically generate:

1) A mass – by randomly sampling from the combined model of mass appearance built in section 7.8

2) A mass background – by modifying the appearance of a normal region as described in Algorithm 8.1

In this chapter we show we can combine the two, thus reversing the separation process described in chapter 6, and generating a complete mass. The appearance properties of the complete mass should match the properties of the real masses in our training data as they originally appeared in mammograms.

To begin the process of combining masses and mass backgrounds, we look again at the separation process described in chapter 6. Using the same labelling as previously, we can define a rectangular $X \times Y$ pixel region about any of the masses in our training data as

$$R(x, y), \quad for \quad x = 1, ..., X, \quad y = 1, ..., Y$$

Within this rectangular region, there will be a set of pixels $M = \{(x, y)\}$ that, due to x-rays passing through the mass, have a higher grey level than pixels at which x-rays passed only through normal breast tissue. Thus we can describe the region R as the sum of two sources of intensity, such that

$$R(x, y) = R_M(x, y) + R_B(x, y), \quad R_M(x, y) = 0 \ \forall (x, y) \notin M$$

Applying the separation process described in chapter 6 to each mass in the training data produced the set of separated masses $\{R_M\}$ modelled in chapter 7 and the set of mass backgrounds $\{R_B\}$ modelled in chapter 8.

If we now consider the synthesis methods described in the preceding two chapters, we recall we can sample from the mass appearance model to generate a synthetic mass S_M, and can modify any normal mammogram region into a synthetic mass background S_B.

Here S_M is rectangular region generated so that its centre is also the centroid of the mass (i.e. the centre of mass of non-zero pixels in S_M where each non-zero pixel is given an equal weighting). Meanwhile S_B is a rectangular region in which mass background appearance has been synthesised about some central pixel (x_c, y_c).

Providing S_B is sufficiently large[11], S_M can be padded with zeros to be the same size as S_B and such that the mass centroid in S_M has co-ordinates (x_c, y_c). As a result, if S_B is an $X \times Y$ pixel region, we can generate a complete mass S such that

$$S(x, y) = S_B(x, y) + S_M(x, y) \quad for \quad x = 1, ..., X, \quad y = 1, ..., Y$$

That is, we simply compute the pixel-by-pixel sum of intensities in the synthesised mass and mass background.

In the formulation above, we have treated S_M and S_B as independent of one another. However, for the complete region S to look realistic, the appearance properties of S_M must match those of S_B. In the following section we show how this can be achieved.

9.2 Sampling synthetic masses to match mass backgrounds

To develop a method for matching masses to mass backgrounds we first consider how synthetic mass backgrounds are formed. Recall, from the discussion in section 8.8, we have chosen to synthesise mass background appearance using Algorithm 8.1. In this algorithm a normal region is modified by directly transferring DT-CWT coefficients from a real mass background randomly selected from the training data.

Therefore if S_B is a synthetic mass background generated using Algorithm 8.1, there will be some real mass background R_B that has a strong correlation to S_B. Consequently, if R_M is the real mass that was separated from R_B in chapter 6, there will

---

[11] In section 10.3.3 we will show how we can ensure a large enough region for S_B is always selected

be a correlation between the appearance of S_B and R_M. This means that the synthesised

mass we combine with S_B should have a similar appearance to the real mass R_M.

If we now consider the mass appearance model built in section 7.7, we recall that the

appearance of each mass in the training data was parameterised by a combined

appearance vector. This vector specified co-ordinates along axes in the model feature

space defined by the set of principal modes $\mathbf{P}_{com}$. If we label the combined appearance

vector for a real mass R_M and a synthetic mass S_M as $\mathbf{b}_{com}$ and $\tilde{\mathbf{b}}_{com}$ respectively, then

stating that S_M should have a similar appearance to R_M is equivalent to saying that

$\tilde{\mathbf{b}}_{com}$ should lie close to $\mathbf{b}_{com}$ in the model feature space.

In section 7.8, new appearance vectors were sampled from the model feature space by

assuming a Gaussian distribution over the space. This distribution was assumed to have

no covariance between modes and so each element of a new vector could be sampled

independently from a univariate Gaussian distribution. Thus if there were D modes in

the combined appearance model, a new vector of mass appearance $\tilde{\mathbf{b}}_{com}$ could be

sampled such that

$$\tilde{\mathbf{b}}_{com}(i) \sim N(0, \lambda_i), \quad i = 1, ..., D$$

where each λ_i was computed as the standard deviation of model parameters in the

training data along the i-th principal mode.

In this chapter, to ensure that a sampled vector $\tilde{\mathbf{b}}_{com}$ lies close to a real appearance

vector $\mathbf{b}_{com}$, we fix the first d elements of $\tilde{\mathbf{b}}_{com}$ to be equal to the first d elements of

$\mathbf{b}_{com}$. That is, we generate $\tilde{\mathbf{b}}_{com}$ such that

$$
\begin{aligned}
\tilde{\mathbf{b}}_{com}(i) &= \mathbf{b}_{com}(i), & i &= 1,...,d \\
\tilde{\mathbf{b}}_{com}(i) &\sim N(0,\lambda_i), & i &= d+1,...,D
\end{aligned}
$$

Note that this doesn't force $\tilde{\mathbf{b}}_{com}$ to lie close to $\mathbf{b}_{com}$ along all dimensions of the space. Indeed for modes greater than d, $\tilde{\mathbf{b}}_{com}$ can lie infinitely far from $\mathbf{b}_{com}$. However, the modes of the appearance model are ordered in decreasing variance. So moving from 1 to D, the elements of $\tilde{\mathbf{b}}_{com}$ have a decreasing effect on the appearance of the reconstructed mass S_M. Thus by fixing only the first d modes of $\tilde{\mathbf{b}}_{com}$ we ensure the general shape, size and intensity of S_M will be similar to R_M, whilst allowing the details of the synthetic mass to vary.

The number of modes we fix controls the extent to which S_M is constrained to the appearance of R_M. For example, setting $d=0$ enforces no constraint, whilst setting $d=D$ results in S_M being equivalent to the full model reconstruction of R_M (see section 7.3). This effect is shown in Figure 9.1, in which the top row contains a real mass from the training data. In each of the five rows under the real mass, there are four synthetic masses sampled from the mass appearance model, with the number of fixed modes, d, increasing in each row. The values of d shown are 0, 1, 3, 5 and 10 respectively (in each case $D=32$). We can see that in the top row, where no constraints are used ($d=0$), the synthetic masses bear very little resemblance to the real mass. In particular we can see the size and overall intensity of the masses vary greatly. Moving through the subsequent rows we can see the masses become more similar in appearance

to the real mass.

To select a value for d for a particular mass appearance model, we can visually assess the change in appearance of the mean mass as a result of varying each mode through ± 2 standard deviations (whilst keeping all other modes fixed). This allows us to see how each principal mode changes the appearance of a mass.

For example, in an appearance model built from the masses in our training data using the method described in sections 7.3, 7.4 and 7.5, the first and second modes are associated with a large change in mass size, along with a change in the overall intensity of the mass. The third, fourth and fifth modes are primarily associated with a change in mass shape – specifically an elongation, a triangulation and a squaring of mass shape respectively. From the sixth mode onwards, the effects on mass appearance are more subtle. The size, overall intensity and low-frequency shape of the mass tend to be similar. Instead the modes are associated with changes in the distribution of intensity across the mass and higher frequency variations in mass shape.

In constraining the appearance of a synthetic mass to match a mass background, we want the general properties of the mass to be fixed, whilst allowing the details of the mass to vary. We therefore choose to set $d = 5$. We note however, that having a quantitative method for selecting d would be preferable. This issue is discussed further in section 9.4.

Figure 9.1: Top row: a real mass from the training data; Subsequent rows: four synthetic masses sampled from the mass appearance model after fixing the first 0, 1, 3, 5 and 10 modes of the sampled appearance vector to the real mass
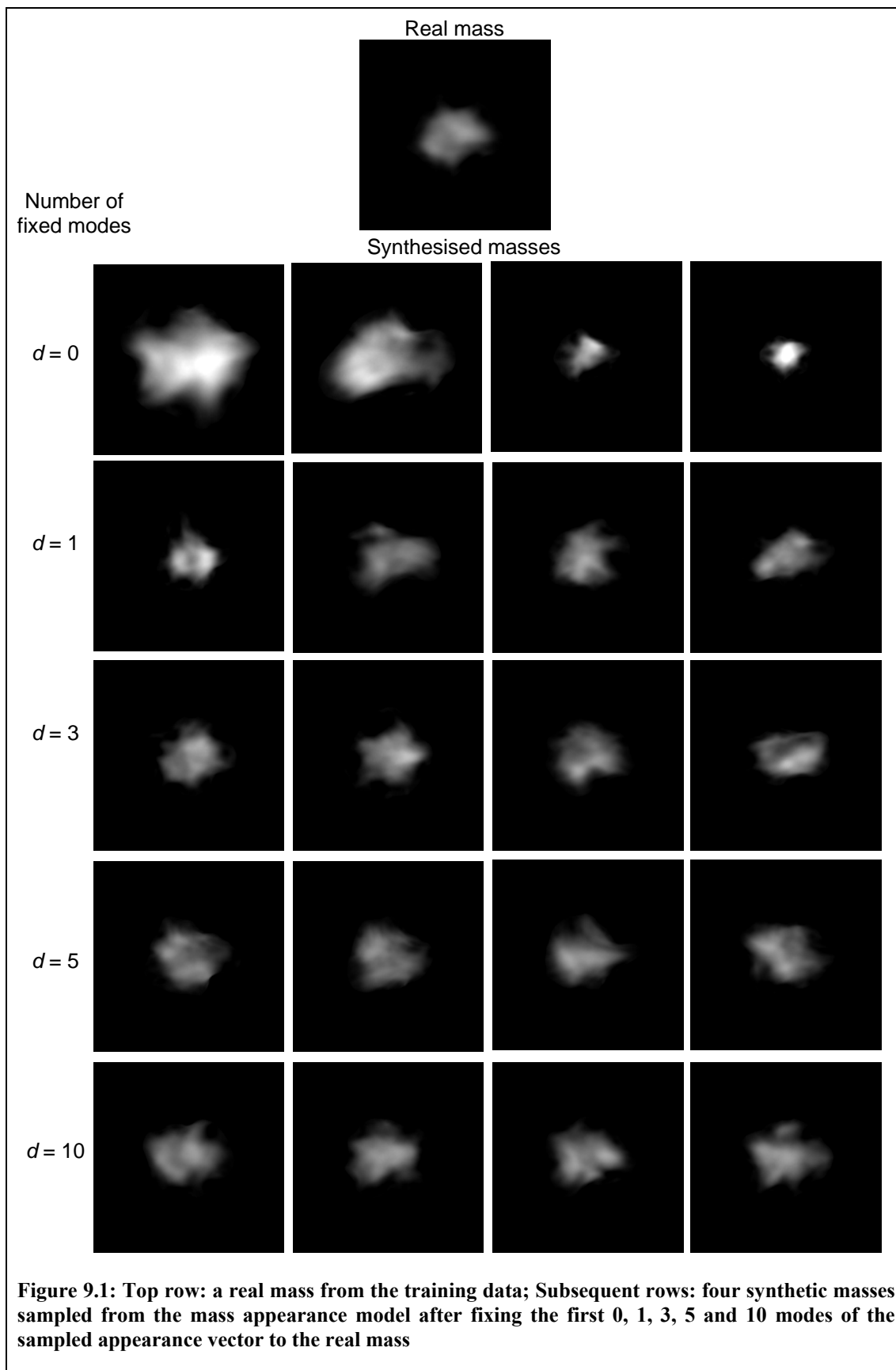
9.3 Aligning masses in mass backgrounds

In the previous section we showed how given a synthesised mass background, we can generate a synthetic mass to match the appearance properties of the region. However we have not yet considered how the synthetic mass should be rotated.

In the first stage of building an appearance model of masses, the shape of each mass was aligned using Procrustes analysis, as described in Algorithm 7.2. As a result of applying the Procrustes algorithm, the rotation, translation and scaling needed to align each mass to the mean mass shape was computed. The three parameters together describe the pose of mass shape in a region.

The scale parameter of each mass shape was included in the combined appearance model so that correlation between mass shape and texture and mass size was included in the final model (as described in section 7.6).

Meanwhile the translation of each mass shape encodes the location of its centre relative to the mammogram from which it was extracted. However, as discussed in section 9.1, the synthetic masses and mass backgrounds considered in this chapter are generated to have matching centres, and so we do not need to consider translation when matching the appearance of the two. Modelling the location of a mass with respect to the mammogram from which it is extracted, and thus properly accounting for translation, is the subject of the next chapter.

Therefore in matching the pose of a mass shape to a mass background it remains to consider rotation.

In Figure 9.1, each of the synthesised masses have been rotated to match the orientation

of the real mass to which they are matched. Because the real mass used in this example has a generally round shape, rotating each synthetic mass has made little difference in their similarity to the real mass. However a real mass that is particularly elongated only matches the properties of its mass background at the correct orientation. This is highlighted in Figure 9.2.

In Figure 9.2 (a) we show a real mass, R, that has an unusually elongated shape compared to the majority of masses in our data. If R_M and R_B are the mass and mass background generated by separating R in chapter 6, then we can generate a new region R' by rotating R_M through 90 degrees about its centre, before adding R_M back to R_B. Figure 9.2 (b) shows the modified region R'. In both R and R', the mass outline and the dominant spicules in the region have been annotated. In Figure 9.2 (b), we see that after rotating through 90 degrees, the mass no longer matches the spicules in the mass background.

We now consider how a synthetic mass S_M should be rotated within a synthesised mass background S_B. We assume S_B has been generating using Algorithm 8.1, in which a real mass background R_B was used to modify the appearance of a normal region *N*. As discussed in section 8.5.1, prior to applying the algorithm, R_B is rotated so that the dominant orientation of its breast tissue aligns with the dominant orientation of breast tissue in *N*. We label the angle of this rotation ϕ.

(a)   **(b)**
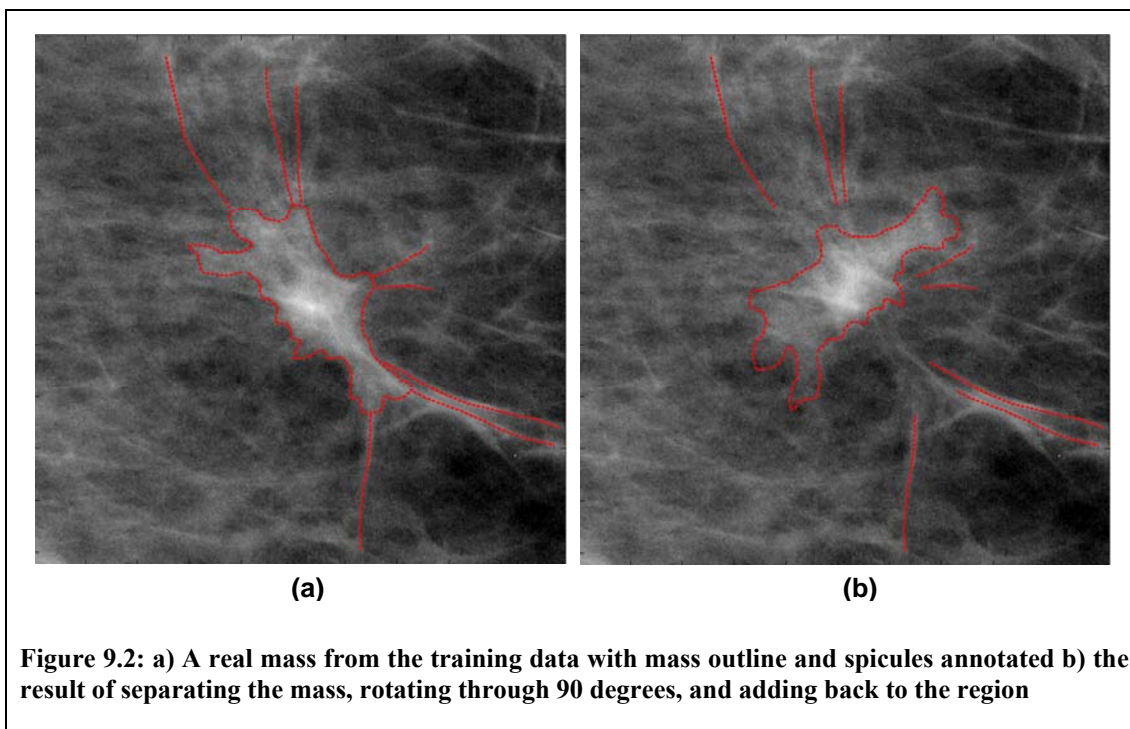
Figure 9.2: a) A real mass from the training data with mass outline and spicules annotated b) the result of separating the mass, rotating through 90 degrees, and adding back to the region

If R_M is the real mass associated with R_B, and R_M has a rotation parameter of θ, this means the mass shape in R_M was rotated by θ prior to constructing the shape model in section 7.4. So if R_M is reconstructed from its model parameterisation, its shape must be rotated by $-\theta$ to correctly match R_B. Therefore to match S_M to S_B, the mass shape in S_M must be rotated by $\phi - \theta$.

We can now write a complete algorithm for generating a synthetic mass in any region of normal mammogram *N*, as shown in Algorithm 9.1.

Randomly select a real mass background R_B from the training data

Rotate R_B to match the dominant tissue orientation of N and store the angle of rotation ϕ

Generate the synthesised mass background S_B from N and R_B using Algorithm 8.1

Let R_M be the mass separated from R_B, $\mathbf{b}_{com}$ be the parameterisation of R_M in the mass appearance model and θ be the rotation used to align R_M during Procrustes analysis

Generate a new mass appearance vector $\tilde{\mathbf{b}}_{com}$ such that:

$$\tilde{\mathbf{b}}_{com}(i) = \mathbf{b}_{com}(i), \qquad i = 1,...,5$$
$$\tilde{\mathbf{b}}_{com}(i) \sim N(0, \lambda_i), \quad i = 6,...,D$$

Compute $\tilde{\mathbf{x}}_{com} = \overline{\mathbf{x}}_{com} + \mathbf{P}_{com}\tilde{\mathbf{b}}_{com}$

Separate $\tilde{\mathbf{x}}_{com}$ and invert the weighting to form $\tilde{\mathbf{b}}_{shape}$, $\tilde{\mathbf{b}}_{tex}$ and $\tilde{\mathbf{b}}_{scale}$

Compute:

$$\tilde{\mathbf{x}}_{shape} = \overline{\mathbf{x}}_{shape} + \mathbf{P}_{shape}\tilde{\mathbf{b}}_{shape}$$
$$\tilde{\mathbf{x}}_{tex} = \overline{\mathbf{x}}_{tex} + \mathbf{P}_{tex}\tilde{\mathbf{b}}_{tex}$$
$$\tilde{\mathbf{x}}_{scale} = \overline{\mathbf{x}}_{scale} + \mathbf{P}_{scale}\tilde{\mathbf{b}}_{scale}$$

Rotate the new shape $\tilde{\mathbf{x}}_{shape}$ by $\phi - \theta$, and scale $\tilde{\mathbf{x}}_{scale}$

Map the new texture $\tilde{\mathbf{x}}_{tex}$ to the new shape $\tilde{\mathbf{x}}_{shape}$, to produce the synthetic mass S_M

Pad S_M with zeros so that it matches the dimensions of S_B

Compute the complete mass region $S = S_M + S_B$

Algorithm 9.1: Complete set of steps for generating a synthesised mass S in a normal region N

The algorithm described above was used to generate the three complete abnormal regions shown in the right-hand column of Figure 9.3, from the three normal regions shown in the left-hand column of Figure 9.3. In chapter 11, the realism of 30 masses generated using Algorithm 9.1 is evaluated in an observer study. The results of this evaluation, including observations on the appearance the masses from expert

mammography readers, are discussed in detail in section 11.6. Therefore the appearance of the masses shown in Figure 9.3 is not discussed here.

9.4 Discussion

In section 9.2 we described how we can generate a synthetic mass S_M so that it has similar general properties of appearance to a real mass R_M. If $\mathbf{b}_{com}$ is the D-element vector that parameterised R_M in the model of mass appearance, and $\tilde{\mathbf{b}}_{com}$ is the appearance vector from which S_M is reconstructed, a likeness between S_M and R_M was achieved by fixing the first d elements in $\tilde{\mathbf{b}}_{com}$ to be equal to the first d elements of $\mathbf{b}_{com}$.

In terms of sampling from the model feature space, fixing the first d modes of $\tilde{\mathbf{b}}_{com}$ effectively collapses the first d dimensions of the space. The remaining elements of $\tilde{\mathbf{b}}_{com}$ can then be randomly sampled from $(D - d)$-dimensional subspace by assuming a Gaussian distribution over this subspace. Generating $\tilde{\mathbf{b}}_{com}$ in this manner means that the final $D - d$ elements of $\mathbf{b}_{com}$ have no effect on $\tilde{\mathbf{b}}_{com}$ (other than the contribution $\mathbf{b}_{com}$ made when learning the global model). This allows the details of S_M to vary freely from R_M, whilst constraining the size, low-frequency shape and overall intensity of S_M.
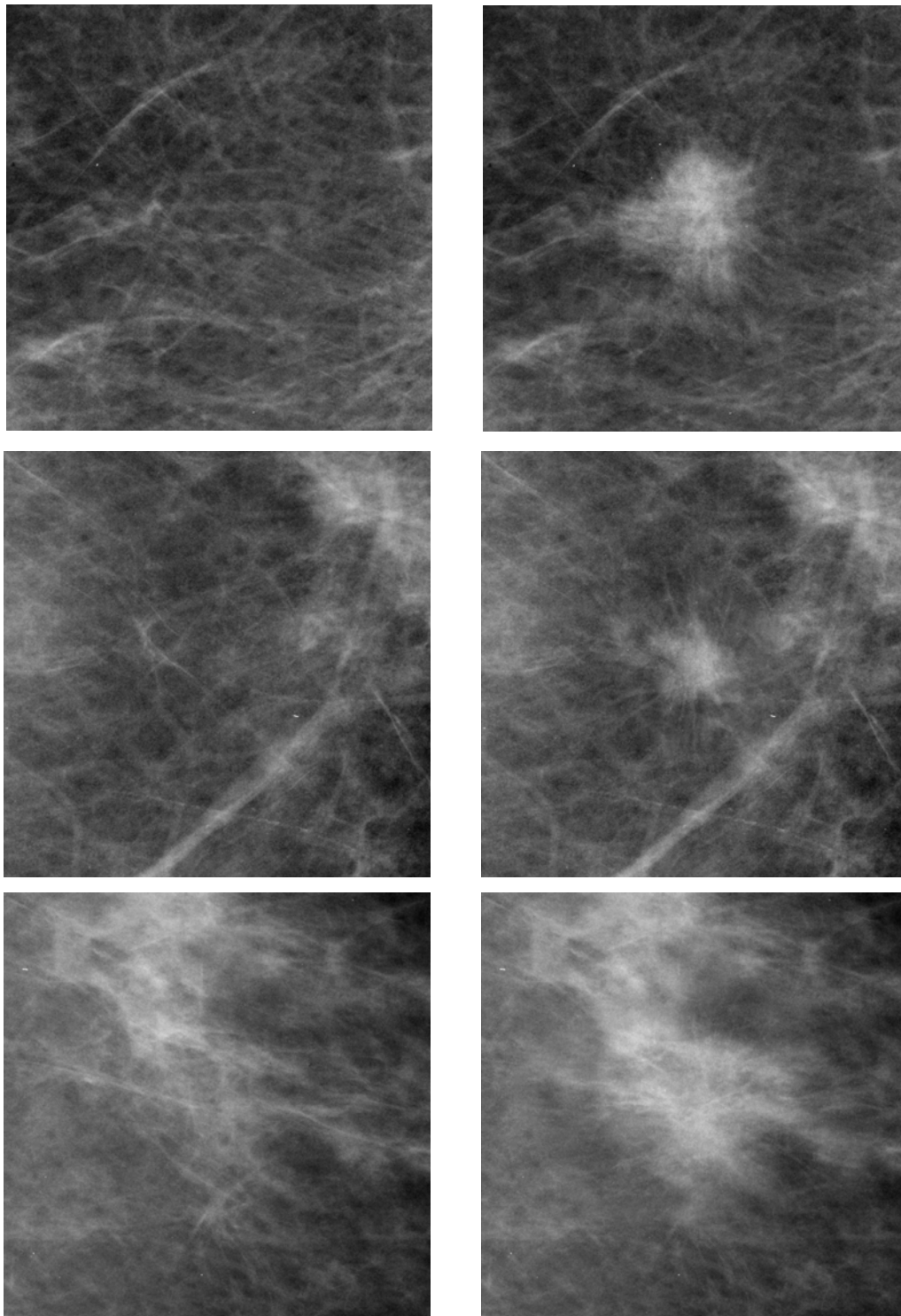
Figure 9.3: Synthesising masses in normal regions. The left-hand column contains three normal regions; the right hand column shows each normal region complete with a synthesised mass and modified backgorund, generating using Algorithm 9.1

In section 9.2, we described how we selected d given an appearance model built from our training set of masses. This involved visually assessing how the appearance of the mean mass was affected, by varying each combined appearance mode through ± 2 standard deviations. As a result we selected a value of $d = 5$ to produce a sufficient level of constraint between a synthetic mass and a real mass. Analysing the variance stored along each mode in the model, we see that the first 5 modes are responsible for 75% of the total variance. We suggest this threshold could be used to a value for d in models built from other datasets.

An alternative method for ensuring a likeness between a synthetic mass and a real mass would be to allow $\tilde{\mathbf{b}}_{com}$ to be randomly sampled across all D modes, but force the space from which each $\tilde{\mathbf{b}}_{com}(i)$ is sampled to lie in the vicinity of $\mathbf{b}_{com}(i)$. For example, rather than assuming a Gaussian distribution across the entire model space, we sample from a local Gaussian distribution centred on $\mathbf{b}_{com}$. In this case, the extent to which S_M resembles R_M is controlled by the standard deviation along each mode of the local distribution. Increasing the standard deviation along each mode would tend to generate sampled vectors $\tilde{\mathbf{b}}_{com}$ lying further from $\mathbf{b}_{com}$, thus increasing the difference in appearance between S_M and R_M. However, as with selecting a value for d (described in section 9.2), we do not have a quantitative method for determining the standard deviation along each mode.

In the absence of a quantitative method for selecting the standard deviations, we prefer the method of fixing a set number of modes in a sampled appearance vector to ensure a

likeness between synthetic and real masses. Whilst this method still requires a subjective decision to select d, the effect of fixing specific modes in the model is easier to visualise than the effect of allowing a limited variation across all modes. In addition, having selected the number of modes to fix for a given model, we can compute the proportion of variance this constrains in the model. As discussed above, setting a variance threshold of 75% could provide a measure to quantitatively select the number of fixed modes in mass models built from different datasets.

In both the methods for constraining a synthetic mass discussed above, we have assumed we have a real mass to which it should be similar. This is because the method chosen to synthesise mass backgrounds (Algorithm 8.1) uses a real mass as a template when modifying a normal region. However, as discussed in sections 8.6 and 8.7, we are working on a method for synthesising mass backgrounds that does not rely on using individual real masses as templates. Thus we conclude this section by considering how synthetic masses and mass backgrounds could be matched using the region synthesis method proposed in section 8.7.

In the method proposed in section 8.7, mass background appearance was synthesised with respect to a map of structure. The map specified the location of structures (both normal and those associated with a mass such as spicules and microcalcifications) in the region, and was parameterised by a fixed length vector in a global model of mass background structure.

If both mass background maps and masses can be parameterised as fixed length vectors, then the joint probability of maps and masses can be learnt from our training data. When a new mass background map is sampled to synthesise a mass background, we can

use the joint distribution to sample a mass appearance vector. As a result the synthetic mass reconstructed from the appearance vector should match the synthesised region generated from the mass background map. Thus in addition to providing a probabilistic framework in which to synthesise mass backgrounds, the method proposed in section 8.7 leads naturally to a quantitative, probabilistic method for matching masses and mass backgrounds.

9.5 Summary

In this chapter we have described how a synthetic mass can be generated to match a synthesised mass background.

In section 9.2, we noted that using Algorithm 8.1 to generate a synthetic mass background caused a strong dependence between the background and a single real mass from the training data. Using the parameterisation of masses in the combined appearance model (generated in section 7.7), we constructed a synthetic mass so that its general shape, size and texture matched those of the real mass.

In section 9.3, we showed that to complete the match between a synthesised mass and a mass background we must ensure the mass shape is correctly rotated. Once this was achieved we were able to formalise the set of steps required to synthesise a complete mammographic mass within any normal region. These steps were described in Algorithm 9.1. Again we note that no parameters need to be manually adjusted during Algorithm 9.1, and thus the algorithm can repeated without the need for user input. The method can therefore be used to generate large sets of synthetic mammographic masses automatically.

However, in both this and the previous chapter, we have so far talked about synthesising a mass background from an existing normal mammogram region. We have not yet discussed how we obtain a normal region from a mammogram. As stated in section 3.2.2, the aim of the work in this thesis is to demonstrate a method for automatically generating a set of mammographic masses in a set of real normal mammograms. Thus to complete this method we must show how a region can be automatically selected from a normal mammogram. Achieving this is the subject of the next chapter.

Chapter 10 : Locating Synthetic Masses in Normal Mammograms

In the previous chapter we presented a complete method for synthesising mammographic masses. As described in Algorithm 9.1, given any normal mammogram region as input, we can generate the appearance of a malignant mass within the region.

However this does not yet meet the goal of our thesis. As stated in section 3.2.2, we aim to develop a method for synthetically generating mammographic masses in normal mammograms. Thus in addition to showing how we can synthesise a mass within a normal region, we must show how the region can be selected from a normal mammogram.

In this chapter we present a method for selecting a suitable region within any normal mammogram in which to generate a synthetic mass. The remaining sections of this chapter show how this can be achieved.

10.1 Introduction

In chapters 8 and 9, we referenced a normal region with respect to some centre pixel, about which mass appearance was generated. Therefore selecting a region in which to synthesise a mass from a normal mammogram is equivalent to finding a suitable location for the centre of a mass.

In this section we give a brief review of a previous method designed to select mass centres in normal mammograms and use this as a basis for developing our own method.

In section 3.2, we reviewed several methods for synthesising mammographic masses. Of the methods that synthesised a mass directly in a 2-D mammogram[12], only the work of Caulkin [23] described a detailed method for selecting a location within the mammogram at which to generate the mass.

Caulkin's method proposed the use of a set of mammograms containing malignant masses as training data. The breast shape of each mammogram was extracted and used to learn a mean breast shape. The mean breast shape was then taken as a common co-ordinate frame, and the location of each mass centre relative to this shape calculated. A probability density function was fitted to describe the distribution of locations within the mean breast and was used to sample new locations in normal mammograms.

The advantage of this approach is that if it is used to select locations for synthetic masses in a set of normal mammograms, the distribution of these locations should match the distribution of mass locations in a real set of data.

We note though, that given a normal a mammogram, a location sampled using Caulkin's method is based only on the shape of the breast. The method does not consider the appearance of breast tissue at the location selected. Because of this, Caulkin noted that the local texture properties of a region selected using the method did not always match the appearance properties of a mass synthesised in the region. For this reason, Caulkin manually selected the locations of synthetic masses used in the evaluation of the work in his thesis.

---

[12] As opposed to synthesis methods in which a 3-D breast model is projected into 2-D such as in Bliznakova's method [17]

Using our mass synthesis method, we have shown in chapters 8 and 9 how the local appearance properties of a normal region can be modified and a synthetic mass generated to match these properties. This allows a mass to be generated in any normal region and therefore a method for selecting mass centres using only breast shape is sufficient for our work.

However, Caulkin's method for selecting mass centres is not adequate for our needs. In particular, there are three areas in which the method must be improved:

1) The model of mass location described by Caulkin is computed using manual segmentations of the breast in each mammogram. Using this model to select a region from a new normal mammogram requires a manual segmentation of the breast in the new mammogram. This contradicts our goal of creating a fully automatic synthesis method, and therefore we develop a new model based on automatic segmentations of the breast

2) The model described by Caulkin applies only to MLO mammograms. However our synthesis method applies to both CC and MLO mammogram, and therefore we extend the model of mass location to cover masses in CC mammograms

3) The model described by Caulkin allows masses to be sampled in the immediate vicinity of the skin-air boundary of a breast. Such locations are not realistic (or indeed physically viable if the mass crosses the skin-air boundary) and thus we describe how the model can be adapted to prevent this from happening

In sections 10.2 and 10.3 we describe how we have implemented the improvements to Caulkin's method outlined above. We start by showing an automatic algorithm to

segment the breast in each mammogram of our training set.

10.2 Segmenting the breast in a mammogram

In this section we formalise what we mean by segmenting the breast. Figure 10.1 depicts an MLO mammogram of a left breast and a CC mammogram of a right breast, with various features of the mammograms labelled. The aim of a breast segmentation algorithm is obtain an accurate outline for the breast in any given mammogram. We are particularly interesting in finding the skin-air boundary which marks the boundary beyond which x-rays pass only through air. In the two mammograms depicted in Figure 10.1, the skin-air boundary is marked by the green dashed line.

In general the skin-air boundary of a CC mammogram begins and ends at the chest wall edge of the mammogram, forming an approximate semi-circle. In MLO mammograms, the skin-air boundary should begin at the top edge of the image and ends either at the bottom of the image or at the chest wall edge (depending on whether or not the infra-mammary fold was captured in the mammogram).

Because of the physical properties of the compressed breast, the skin-air boundary should be approximately smooth. So for the purpose of building a model of breast shape, we require a segmentation algorithm that produces a smooth contour. However, obtaining a smooth boundary may not be necessary for other applications such as computing breast density where the segmentation is used as a mask to determine the set of pixels to be included in the calculations.
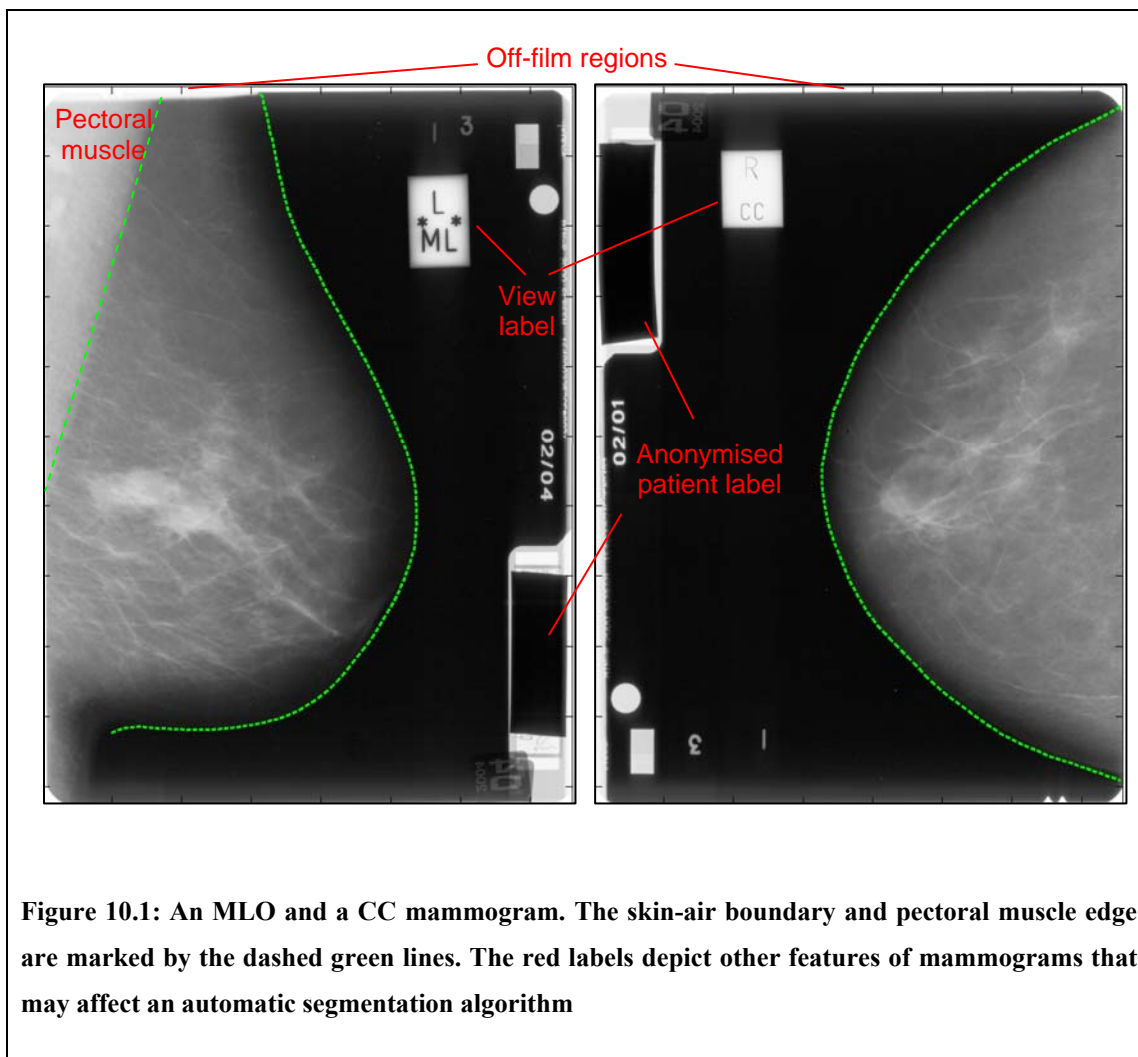
Figure 10.1: An MLO and a CC mammogram. The skin-air boundary and pectoral muscle edge are marked by the dashed green lines. The red labels depict other features of mammograms that may affect an automatic segmentation algorithm

In addition to the breast, there are several other features of a mammogram that must be taken into account when finding an outline for the breast. These features are labelled in Figure 10.1, and discussed briefly below.

- Off-film region: the off-film region in a mammogram is generated when the mammogram is digitised. It usually forms a near white border at the edge of the digitised image. Because of its extreme gey-level and characteristic straight edges it is usually easy to determine and is excluded from the image prior to locating the

breast boundary

- Patient label: The patient label always appears in the top left corner of right mammograms and the bottom right corner of left mammograms. In the original films it is an area of high intensity, however all the mammograms in our data have been anonymised prior to digitisation and thus a near black rectangle covers all but the edges of the patient label. Because of its fixed position, the patient label can be excluded at the start of any segmentation algorithm and plays no part in determining the breast border

- View labels (and other image markers): the view label in each mammogram specifies whether the mammogram is a CC or MLO view of the left or right breast. It does not have a fixed position, and in some mammograms may be located close to the skin-air boundary. In such cases the view label may cause problems because it generates features (for example a strong edge) that may be confused with the breast border in automated algorithms. In addition to the view label, there may be other labels and stickers in the mammogram. As with the view label, if any of these are located too close to the breast they may cause problems in an automated segmentation algorithm

- Pectoral muscle: In MLO mammograms the pectoral muscle appears as an approximately triangular region of high intensity (relative to breast tissue) in the upper corner of mammograms. We find the boundary between the pectoral muscle and the breast (as marked in Figure 10.1) as part of our segmentation algorithm, so that the boundary can be used to define an axis of alignment for the breast in MLO

mammograms.

In the following sections, we describe a method for segmenting the breast and pectoral muscle in each of the mammograms in our training data. The breast segmentation algorithm we present is based upon one developed by Ferrari et al. [44]. We then discuss why Ferrari et al.'s method was chosen as a base and discuss the changes we have made. We also discuss how our algorithm relates to other methods previously reviewed in section 3.3.2.

As with many of the segmentation algorithms reviewed in section 3.3.2, our algorithm is divided into two main stages:

1) Obtaining an approximate estimate of the main breast region that lies inside the skin-air boundary

2) Finding the skin-air breast boundary in the region surrounding the approximate breast region

We describe in detail our method for achieving each of these two steps in the following sections.

Before proceeding we note that all mammograms were initialised to be upright with the chest wall on the left side of the image. Thus prior to segmentation, mammograms of the right breast were reflected about the vertical axis. All mammograms were resized to have a spatial resolution of 250μm per pixel. This represents a loss of resolution relative to the images used to model mass appearance; however the additional spatial resolution is not required to segment the breast. As reported in section 5.1.1, all mammograms

were 8-bit images with grey level pixel intensities ranging from 0 to 255.

10.2.1 Obtaining an approximate breast region

Ultimately the aim of a breast segmentation algorithm is to determine the skin-air boundary of the breast. However, the breast pixels in the immediate vicinity of the skin-air boundary attain their intensity due to x-rays passing through only a thin layer of skin and fat. Because of this, the grey levels of these pixels are usually much closer to the grey levels of background pixels than to the pixels in the remainder of the breast and the contrast between breast pixels at the skin-air boundary and the background is very low. This makes finding the skin-air boundary directly (that is, without prior knowledge of its approximate position) difficult.

Our solution is to compute an approximate region for the breast, comprising the set of pixels that can be differentiated from the background by setting a threshold on pixel grey level. The edge of this region can be used to initialise a search for the skin-air boundary. By construction this edge should lie inside the skin-air boundary, and thus we label it as the *inner breast edge* in the remainder of this chapter.

To compute a threshold for each mammogram, we generated a grey level histogram (with one bin per grey level from 0 to 255) for each mammogram. Whilst there may be local fluctuations in the bin counts, the histograms of all mammograms in our training data share a similar shape. There is a large peak at low grey levels associated with background pixels, a second peak associated with breast pixels (this second peak may have several local maxima depending on the composition of the breast) and a final spike at very high grey levels associated with the near-white pixels in the off film region. In Figure 10.2 we show the combined histogram of all the mammograms in our training

data, along with the two histograms of individual mammograms. In each case, the histograms share the characteristics described above.

Figure 10.2: Pixel intensity histograms for the combination of all mammograms in the training data (red) and for two individual mammograms (blue)
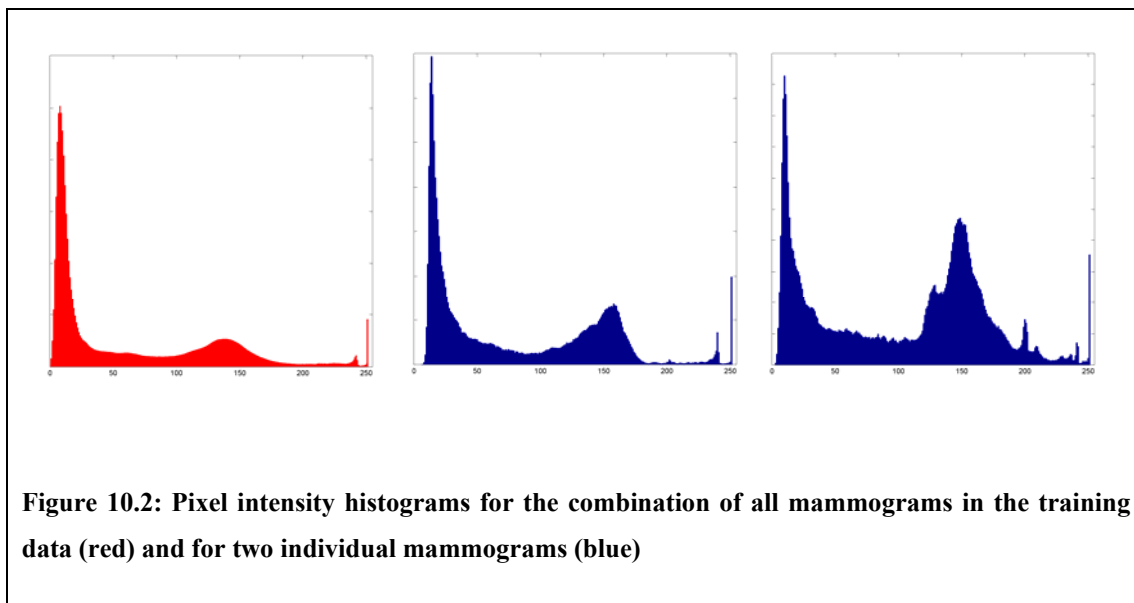
For each histogram we attempt to find the midpoint between the peak associated with background pixels and the peak associated with breast pixels. This is achieved by applying the Lloyd-Max algorithm [75] to compute the centroid of the two peaks. So that the spike associated with off-film pixels does not confound the algorithm, the algorithm was applied to the first 240 histogram bins. Computing the midpoint of the background and breast pixel peaks provides a grey level g with which we can threshold the mammogram.

In applying the threshold we generate a binary map such that all pixels with grey level greater than g are defined as breast pixels and assigned a value of one. All pixels less than or equal to g were assumed to be background and assigned a value of zero. Depending on the composition and position of identification labels in the mammogram,

some labels may not have been set to zero along with some bright "speckle" pixels in the background. To remove these, all pixels except those belonging to the largest object in the binary map are set to zero. Additionally, all zero pixels in the interior of the largest object are set to one.

(a)   **(b)**

Figure 10.3: Binary masks of an MLO mammogram: a) After the initial threshold is applied b) After applying morphological opening and discarding all but the largest region
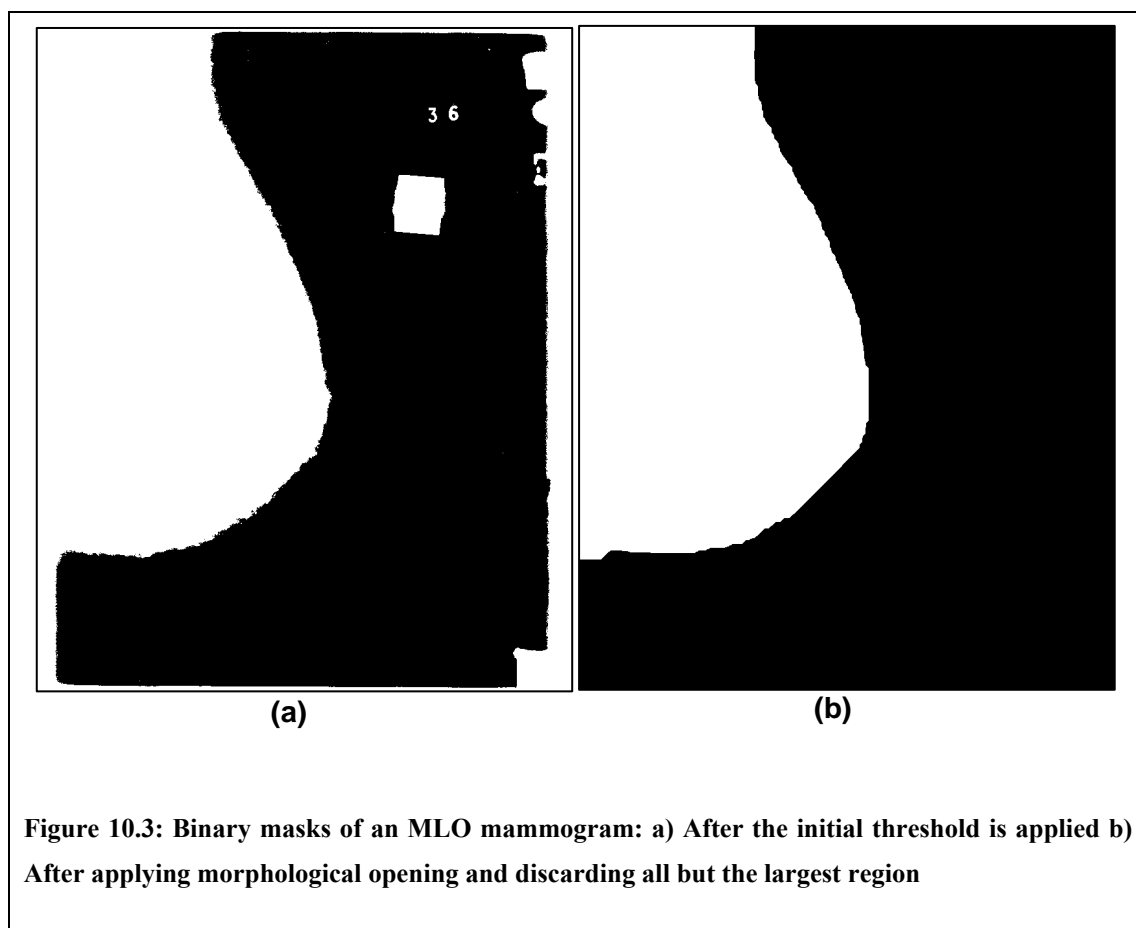
The remaining breast object may be attached to the off-film region, as depicted in Figure 10.3 (a). Although the main characteristic of the off-film region is its high (near white) grey level, it cannot usually be removed simply by thresholding the original mammogram. This is because there is often a diffuse region adjacent to the off-film area where the intensity fades from near white to the near black background. Instead, we apply morphological opening to the binary map, thus removing (that is, setting to zero)

the off-film region as depicted in Figure 10.3 (b). Given the typical width of the off-film regions in our data, we found opening with a 75 pixel disc was appropriate. However the size of the structuring element may need to be varied for films obtained from a different digitiser (which may introduce off-film regions of a different size).

The resulting map is used as an approximate region for the breast. The inner-breast edge can then be computed as the perimeter of this region (ignoring where the perimeter lies on the top, bottom and chest wall edges of the image).

10.2.2   Finding the skin-air boundary

Determining the inner breast edge in a mammogram, as in the previous section, provides an approximate location for the true skin-air boundary of the breast. In general, the skin-air boundary should run approximately parallel to the inner breast edge, with the two edges separated by a small distance. However, the exact distance is determined by the size and composition of the breast, whilst local irregularities in the inner breast edge mean that the skin-air boundary won't always be parallel to the inner edge. Due to physical properties of the compressed breast though, we expect the skin-air boundary to be a smooth curve, with a possible irregularity at the nipple.

To find the skin-air boundary we first smooth the inner breast edge by computing a moving average along the edge. This ensures the local orientation of the inner edge is smooth. We then extract grey level profiles from the mammogram along lines normal to the direction of the smoothed inner breast edge. To avoid profiles that overlap, the profiles were sampled at every 25 pixels (~6mm) along the smoothed edge.

Position of inner breast edge (50th column of image)

(a)   (b)

Skin-air boundary

Figure 10.4: Image formed by stacking the set of normal profiles sampled at intervals from the inner-breast edge of the mammogram shown in Figure 10.1: a) displayed using a grey-scale map b) displayed using a contrast enhanced RGB colour map
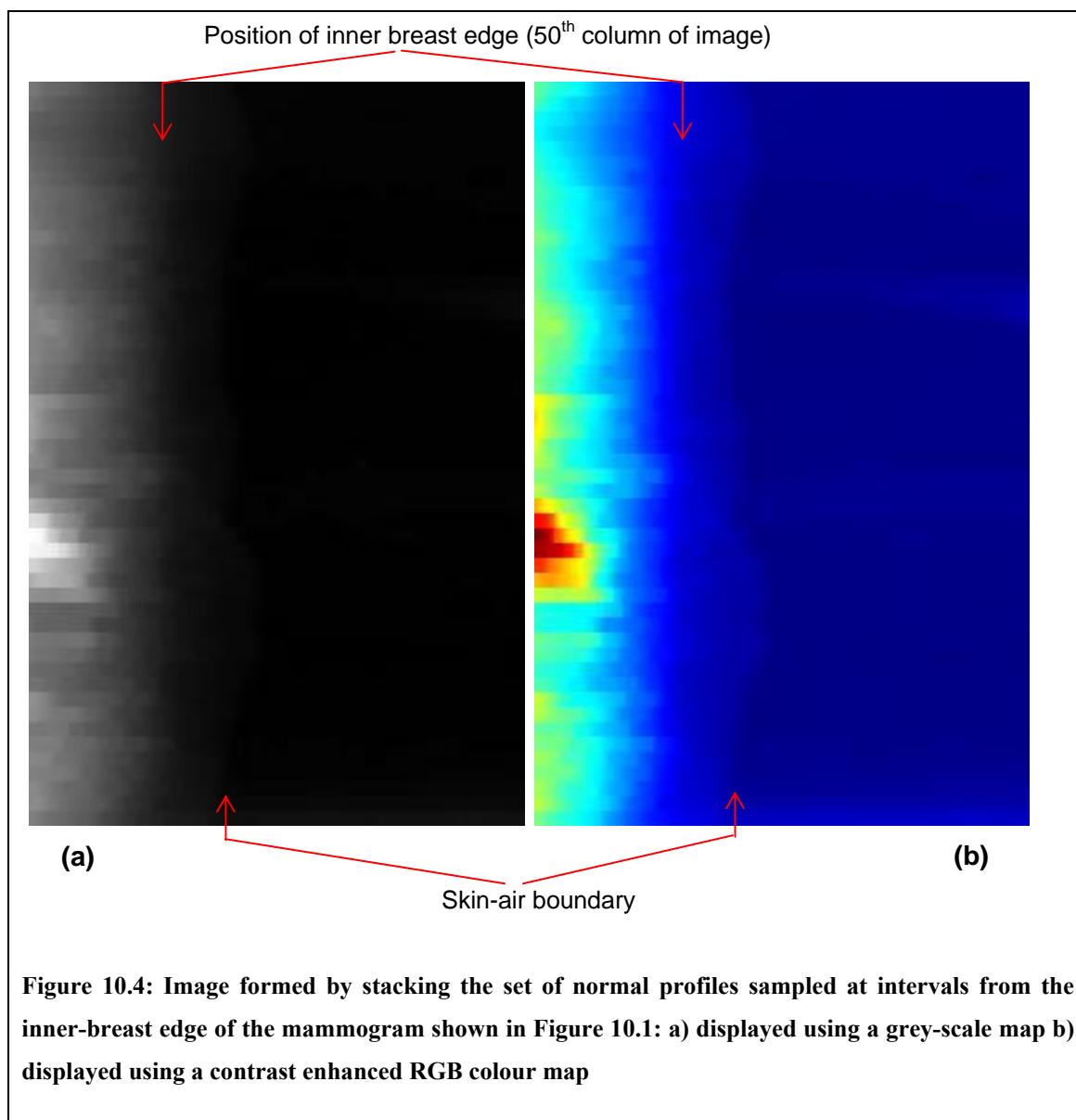
The coordinates along each profile define the set of positions from which the final skin-air boundary is computed. Although the skin-air boundary should lie outside the inner edge, there may be local deviations in the inner breast edge that cause this not to be the case. For example, this could occur if a patient label was located too close the breast. Thus we extend the normal profiles 50 pixels (~12mm) inside and 150 pixels (~37mm) outside the inner edge (the values were empirically chosen – however experiments show

that providing they are large enough, the algorithm is robust to changes in the profile length – thus the profiles can be chosen to be conservatively large).

In Figure 10.4 (a) we show an image generated by stacking the set of normal profiles extracted from the mammogram depicted in Figure 10.1. By construction, pixels lying on the smoothed inner edge now form the 50th column of the normal profile image. The strong inner breast edge originally detected by thresholding the mammogram is clearly visible in the immediate vicinity of this column. Meanwhile the much lower contrast skin-air boundary lies to right. The weakness of the skin-air boundary makes it virtually impossible to see in the grey-scale image depicted in Figure 10.4 (a), however it can be seen in Figure 10.4 (b) by using a contrast enhanced RBG colour map.

To detect the skin air boundary we make use of three of its key properties:

i)   In the image of normal profiles it forms a near vertical edge that changes from higher intensity to lower intensity moving from left to right

ii)   Of the edges in the normal profile image matching the properties defined in point (i), it has the lowest grey level values

iii)   In the original mammogram it forms a smooth curve

(a)

(b)
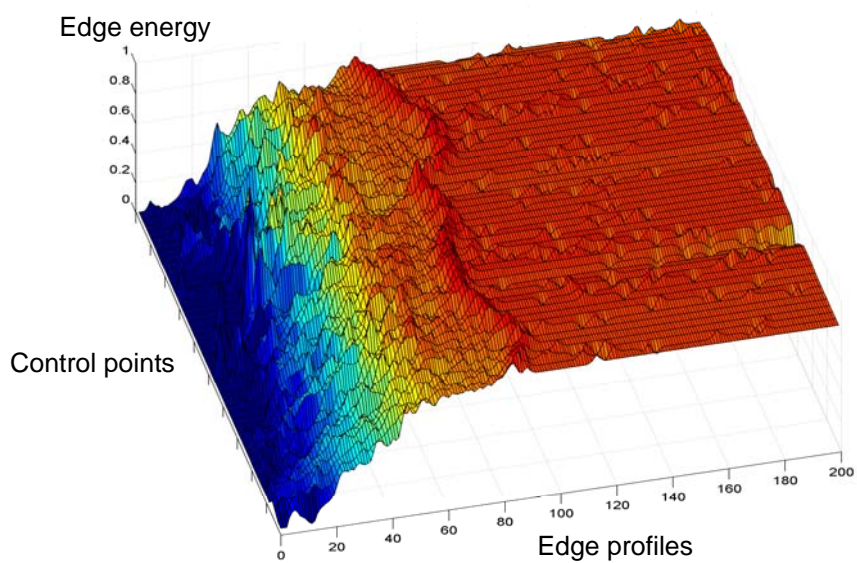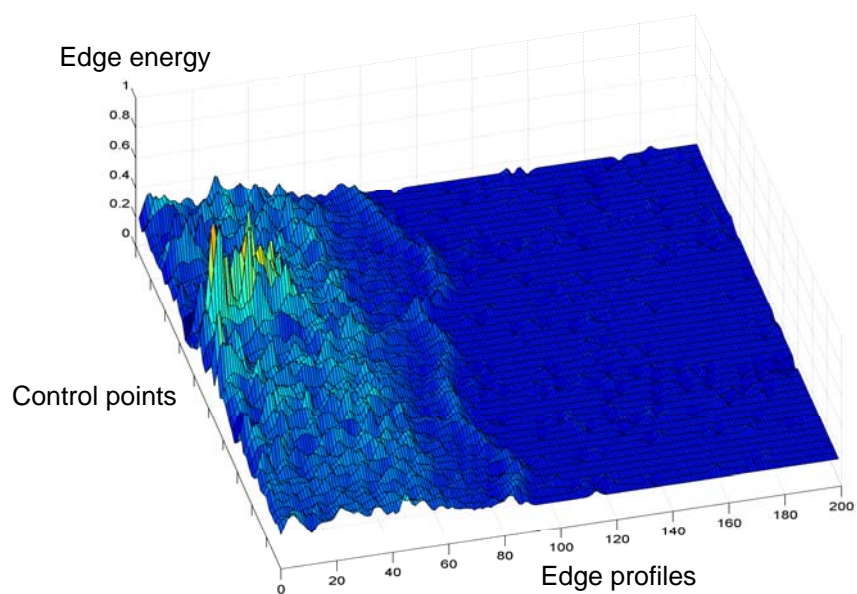
Figure 10.5: Edge energies of a set of normal profiles: a) After applying the filter [1 2 4 0 -4 -2 -1]; b) After applying the filter [1 2 4 -2 -4 -2 -1]

We combine the first two points to form an energy image for the skin-air boundary, such that pixels lying on the boundary should have the highest value in the energy

303

image. This is achieved by filtering the rows of the normal profile image. Applying a vertical edge detector, such as the filter [1 2 4 0 -4 -2 -1] accentuates left-right edges as defined in property (i). Subtracting a multiple of intensity on which the filter is centred accentuates low contrast edge as described in property (ii). The two steps can be combined by applying a single filter [1 2 4 -φ -4 -2 -1] for some positive integer φ. A value of $\varphi = 2$ was empirically determined. The result of applying the edge filter $\varphi = 0$ is shown in Figure 10.5 (a). We see that whilst the skin-air edge is clearly identifiable, it is not the strongest feature in the region. The result of applying the filter with $\varphi = 2$ is shown in Figure 10.5 (b). Note how the skin-air boundary is now a stronger feature than inner breast edge, and indeed is the strongest feature in the region.

However due to noise in the image, the energy image alone is not sufficient to detect the skin air boundary. Whilst the maximal pixel in the majority of the rows lies on the skin-air boundary, in some rows a completely unrelated feature may have generated a maximal response. To overcome this we make use of property (iii) and apply an adaption of the active contour algorithm [64]. Active contours provide a way of detecting a smooth edge such that local image energy is maximised. As described in section 3.3.2, several previous breast segmentation algorithms have used variants of the active contour to find the skin-air boundary.

The key feature in our adaption of the active contour algorithm is that local energy is maximised in the edge energy image whilst the smoothness constraints are applied to coordinates with respect to the original mammogram – thus satisfying the third property of the skin-air boundary outlined above.

Algorithmically, this is efficiently implemented by storing arrays of the *x* and *y* co-

ordinates of the points at which normal profiles are sampled in the mammogram. For an example, we label a mammogram *M* and consider a case in which *n* normal profiles are sampled along the smoothed inner breast edge. The *i*-th normal profile N_i consists of 200 points sampled from *M*, at points $\left(x_{i,j}, y_{i,j} \right)$, $j = 1,\ldots,200$. Thus we can form the normal profile image *N*, and the associated arrays of sampling points *X* and *Y* as:

$$
N = \begin{bmatrix} N_1 \\ \vdots \\ N_n \end{bmatrix} = \begin{bmatrix} I(x_{1,1}, y_{1,1}) \cdots I(x_{1,200}, y_{1,200}) \\ \vdots \qquad\qquad \vdots \\ I(x_{n,1}, y_{1,1}) \cdots I(x_{n,200}, y_{n,200}) \end{bmatrix}
$$

$$
X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_{1,1} \cdots x_{1,200} \\ \vdots \qquad \vdots \\ x_{n,1} \cdots x_{n,200} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} y_{1,1} \cdots y_{1,200} \\ \vdots \qquad \vdots \\ y_{n,1} \cdots y_{n,200} \end{bmatrix}
$$

We can then generate the edge energy image *E* by filtering the rows of *N* as discussed above. The final step prior to running the active contour algorithm is to initialise a set of control points. We set one control point per row, thus there are *n* control points. We position the *i*-th control point p_i on the column that returns the maximum edge energy in that row, thus $p_i = (c,i)$ where $c = \arg\max_j \left(E(j,i) \mid j = 1,\ldots,200 \right)$. Having initialised the contour control points we can run our active contour algorithm.

In the standard active contour algorithm the total energy of the contour is minimised by iteratively adjusting the position of the control points. At each control point p_i, three energies are computed:

$$e_{\alpha,i} = \left\| p_i - p_{i-1} \right\|$$
$$e_{\beta,i} = \left\| (p_i - p_{i-1}) - (p_{i+1} - p_i) \right\| = \left\| 2p_i - p_{i-1} - p_{i+1} \right\|$$
$$e_{image,i} = E(p_i)$$

And thus the total energy a point contributes to the contour can be computed as:

$$e_i = \alpha e_{\alpha,i} + \beta e_{\beta,i} - e_{image,i}$$

The total energy of the contour is computed as the sum of the individual control point energies.

From the equations above we see that the alpha energies $e_{\alpha,i}$ are minimised by minimising the distance between control points (and consequently the length of the contour). Meanwhile $e_{\beta,i}$ is minimised when a control point lies on the midpoint between the preceding and succeeding control points. Thus minimising the beta energies favours continuity in the contour. The constants α and β weight the importance of minimising contour length and discontinuity against maximising the edge energy, encoded by $e_{image,i}$, at each control point.

Our adaption of the algorithm is implemented by changing the computation of the alpha and beta energies to:

$$e_{\alpha,i} = \left\| (X(p_i), Y(p_i)) - (X(p_{i-1}), Y(p_{i-1})) \right\|$$
$$e_{\beta,i} = \left\| 2(X(p_i), Y(p_i)) - (X(p_{i-1}), Y(p_{i-1})) - (X(p_{i+1}), Y(p_{i+1})) \right\|$$

Thus the length and continuity of the contour are computed with respect to its position in the original mammogram.

At each iteration in the active contour algorithm, the control points are constrained to move only within each row. Thus in the mammogram, the control points move in or out along each normal profile until the final skin-air breast boundary is selected.

Figure 10.6 (a) and (b) show the final skin-air boundary detected in a CC and MLO mammogram respectively.

(a)   **(b)**
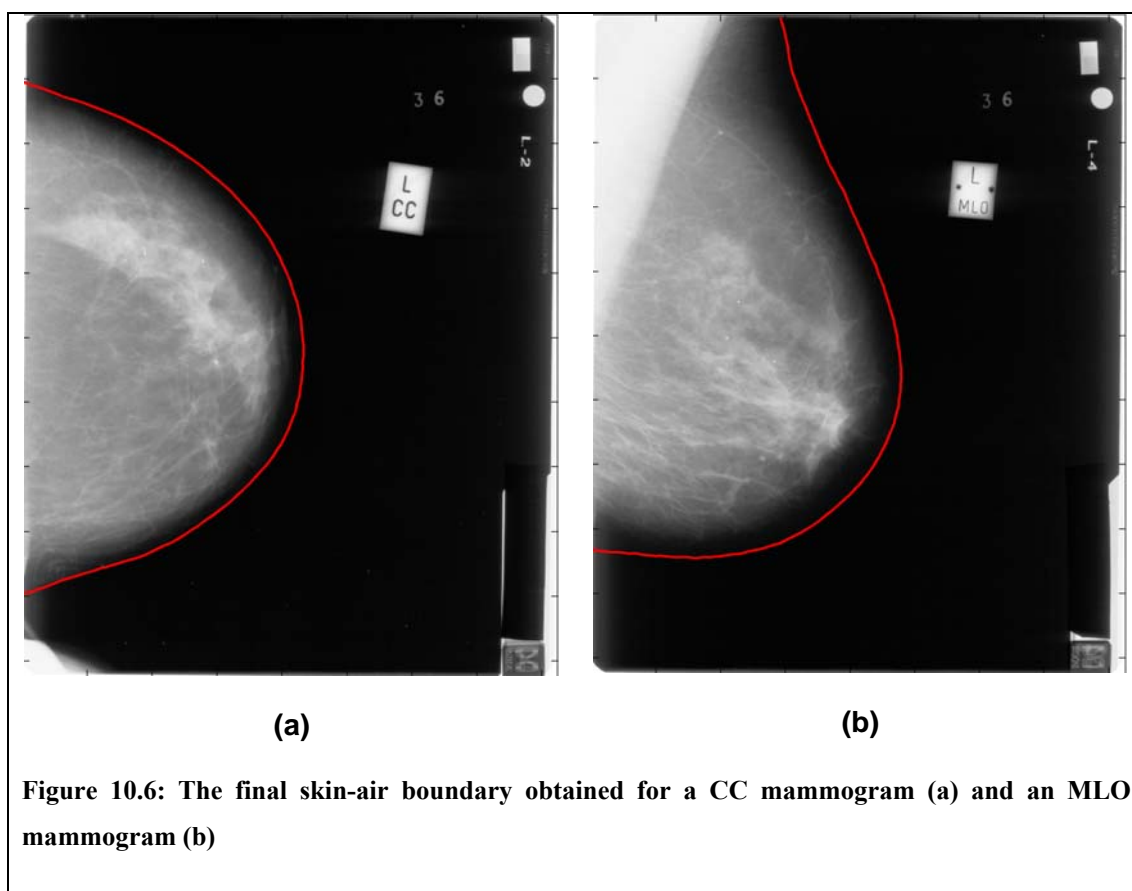
Figure 10.6: The final skin-air boundary obtained for a CC mammogram (a) and an MLO mammogram (b)

10.2.3   Detecting the pectoral muscle in MLO mammograms

To fit the model of mass distribution in MLO mammograms described in section 10.3.1, we require a demarcation of the pectoral muscle. The pectoral muscle occupies an approximately triangular region in the top left corner of each MLO mammogram (remembering that right mammograms have been reflected prior to segmentation). The location of the edge of the pectoral muscle will be used to fix a set of common axes for

MLO mammograms, and thus only a straight line approximately marking the edge of the pectoral muscle is required.

Figure 10.7: Pectoral muscle segmentation: the dashed blue line shows the region searched to find the edge of the pectoral muscle; the red line depicts the straight line fitted to the pectoral muscle
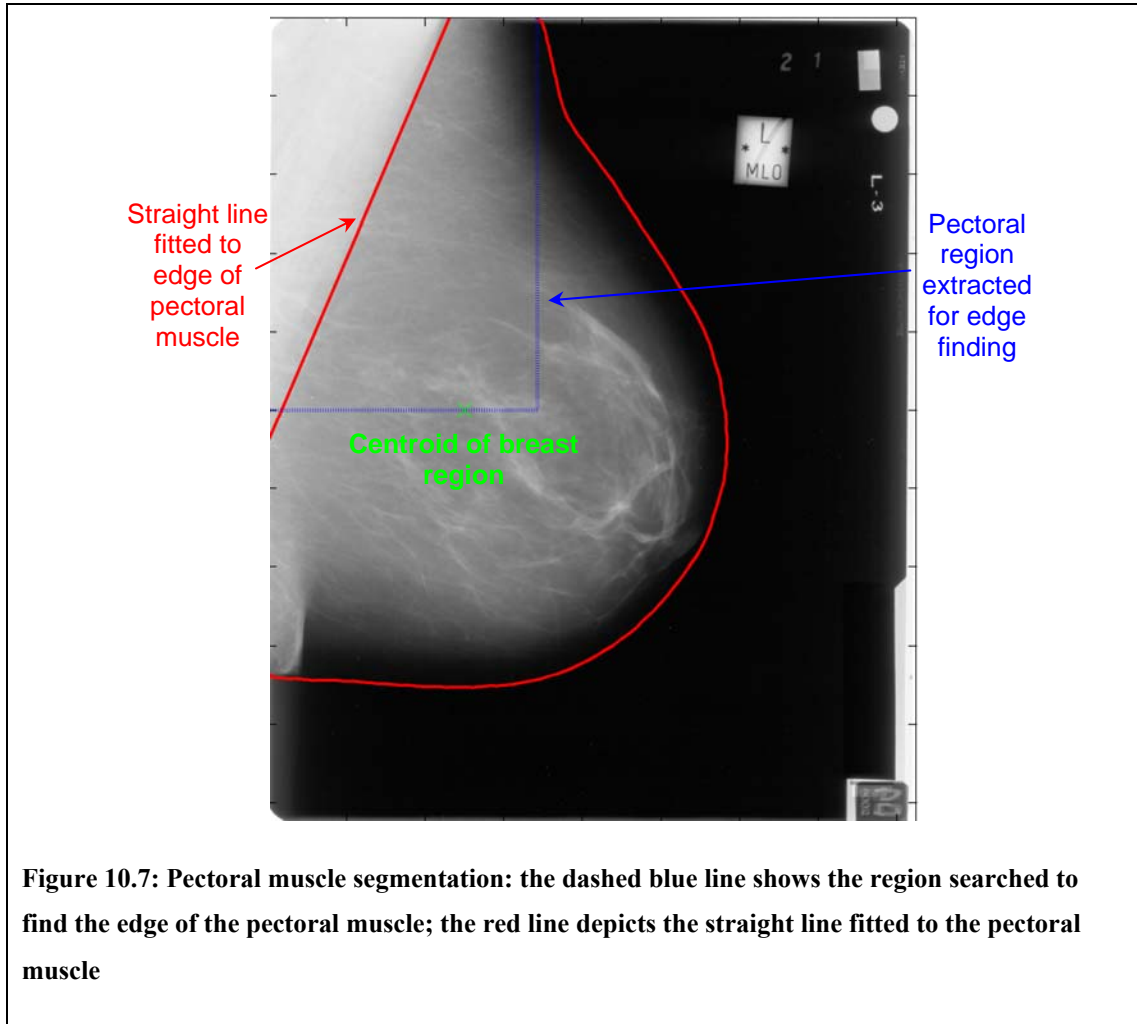
We fit a straight line to the edge of the pectoral muscle in each MLO mammogram using an adaption of the method described by Karssemeijer [62]. First a rectangular region containing the majority of the pectoral muscle is extracted from the upper-left corner each mammogram. The rectangle extends width-ways to the first point on the skin-air boundary, and downwards to the centroid of the breast region, as depicted in Figure 10.7. Note this region does not need to contain the pectoral muscle in its entirety,

providing it contains enough of the pectoral edge to determine a suitable straight line. The region extracted from the mammogram depicted in Figure 10.7 is shown in Figure 10.8 (a).

(a)   (a)

Figure 10.8: a) Pectoral muscle region extracted from the mammogram shown in Figure 10.7; b) Result of applying an orientation specific Canny edge detector to the region in (a)
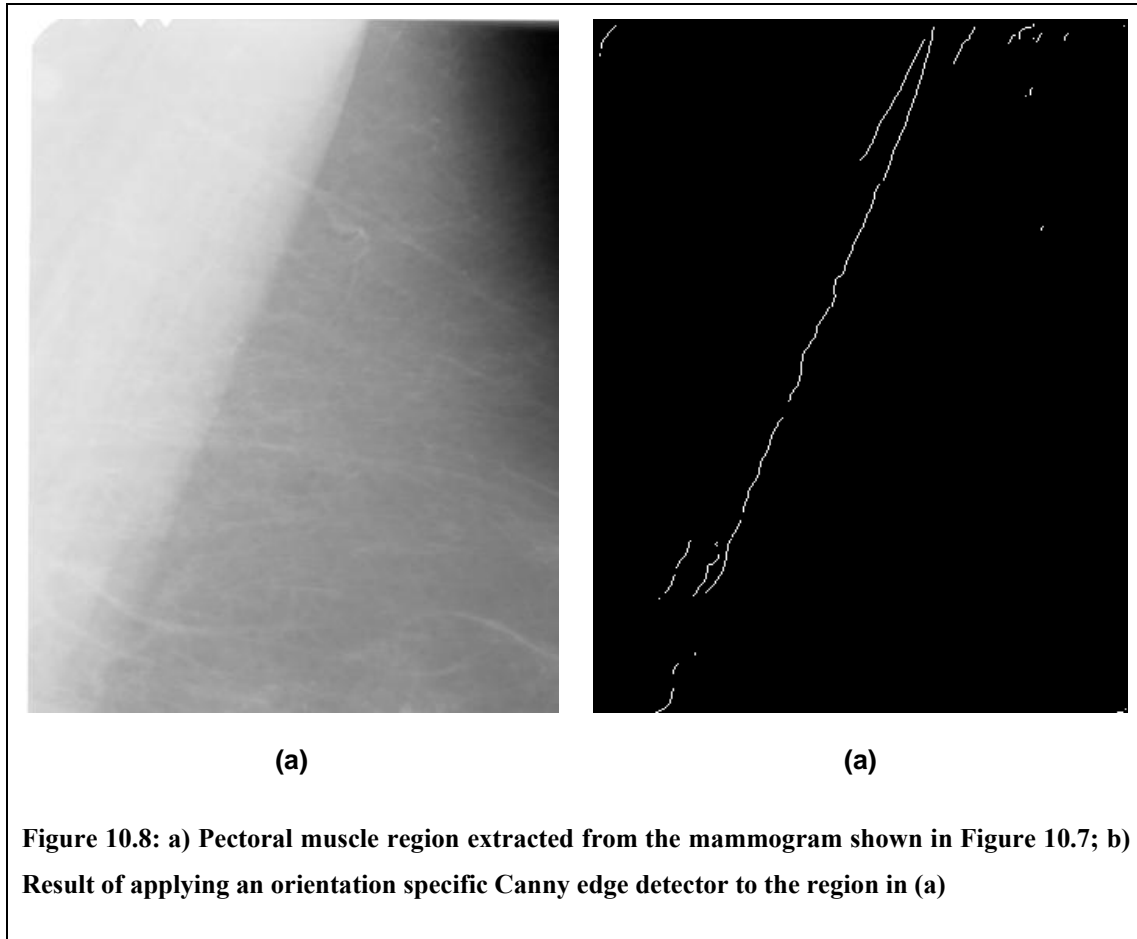
We then apply a Canny edge detector [22] to the region, with the additional constraint that only edges orientated between 30º and 90º are considered. The result of this edge detection is shown in Figure 10.8 (b). Finally a Hough transform [37] is used to select the dominant straight line in the region. Figure 10.7 shows the result of the straight-line fitting in the regions depicted in Figure 10.8.

10.2.4   Discussion

In sections 10.2.1 and 10.2.2 we described our algorithm for segmenting the breast in each mammogram of our data. The key requirement of the algorithm was that a smooth contour was obtained at the skin-air boundary. Provided this requirement is met, any segmentation algorithm could be used.

In section 3.3.2 we reviewed several methods for segmenting the breast in mammograms. Of these we attempted to implement an algorithm described by Ferrari et al. [44]. We chose this method at the commencement of this project as it appeared to represent the state-of-the-art. Moreover, by using an active contour model, it explicitly defined a method for obtaining a locally smooth breast border.

However, whilst implementing Ferrari et al.'s algorithm, we observed that certain aspects of it were not suited to our data. For example, the size and grey level of off-film regions (which are dependent on the scanner used to acquire digitised mammograms) in our data are markedly different from those in the MIAS database (used in Ferrari et als method). Thus having computed an initial threshold for each mammogram, the morphological operations we applied to obtain an approximate breast region, and subsequently the inner breast edge, differed.

In addition, Ferrari et al.'s active contour algorithm was designed to find a continuous contour for the entire breast. This may have been because the images used in their work had an additional area of background at the chest wall side and therefore both the skin-air boundary and a boundary on the chest wall had to be detected. With the mammograms in our data, only the skin-air boundary needed to be detected, and thus a non-continuous contour was used.

Ferrari et al. used their own variant of the traditional active contour model. During each iteration, control points were allowed to move in any direction in a local 7x7 pixel neighbourhood. In our active contour model, control points were allowed to move only along a fixed line normal in direction to the inner-breast edge. Constraining the movement of control points in this way means that to move the contour through a similar distance in the mammogram requires far fewer computations than if control points can move locally in any direction.

Whilst the adapted active contour model developed by Ferrari et al. operated entirely within the mammogram domain, our algorithm computed image energy in a separate image. This image was forming by applying an edge filter to the set of normal profiles extracted from the inner edge. The filter used was designed so that not only would it respond to edges, but that it would respond maximally to the edge of lowest overall intensity in the image. This solved the problem of the active contour locking onto edges formed by areas of high density inside the breast, as was noted in Ferrari et al.'s results.

Of the other breast segmentation algorithms reviewed in section 3.3.2, the work by Pan et al. [86] is particularly interesting. In this the monogenic signal was used to decompose mammograms. Like the DT-CWT, applying the monogenic signal to images generates measures of amplitude, complex phase and orientation across multiple scales. Various thresholds based on phase and orientation were used to determine the inner breast edge, before the final skin-air boundary was obtained using a phase and amplitude based threshold.

The grey level based thresholding used to determine the inner breast edge in our algorithm requires less computation than calculating the monogenic signal for a

mammogram. Because simpler grey level thresholds worked for our set of mammograms we did not attempt to apply measures based on phase and orientation. However, thresholds based on phase and orientation may be more consistent for different sets of mammograms. Given that we already compute the DT-CWT of mammograms in our data when modelling mass backgrounds, it would be interesting though to see whether the steps in Pan et al.'s algorithm could be applied using measures of phase and orientation obtained using the DT-CWT. This may provide an efficient solution of obtaining the benefits inherent in Pan et al.'s algorithm (that is, robustness against changes in intensity and contrast), without having to also compute the monogenic signal of each mammogram.

We also note that the algorithm presented by Pan et al. uses a local threshold to determine the final skin-air boundary. As a result this boundary may not be smooth as noise in the image may cause local fluctuations in the boundary. As discussed in the introduction of section 10.2, obtaining a smooth boundary may not be necessary for all applications of breast segmentation, but is required for our models of breast shape. Thus if we were to use the algorithm presented by Pan et al. to obtain the skin-air boundary in our mammograms we may need to apply a method for smoothing the final boundary.

Because phase and orientation may be more robust to changes in image set, combining the features used to identify the skin-air boundary in the segmentation algorithm described by Pan et al. with our adapted version of the active contour model described in section 10.2.2 could prove the most flexible segmentation algorithm to obtain a smooth skin-air boundary for any generic set of mammograms.

To locate the pectoral muscle in MLO mammograms, a simple procedure was

implemented to fit a straight-line to the dominant edge of the muscle, as described in section 10.2.3. Such a procedure is adequate for the model of mass distribution described in section 10.3.1. However, the edge of the pectoral muscle may not always form a straight line and if a more precise segmentation is needed then further processing steps may be applied. For example, normal profiles could be sampled along the straight edge and searched to find the exact border of the muscle. Such methods were discussed in section 3.3.2.

Finally we note that our segmentation algorithm does not return the position of the nipple. We attempted the methods for locating the nipple described by Mendez et al. [81]. These included searching the skin-air boundary for a region of raised grey level and/or changes in curvature. However, we found that in many mammograms (particularly CC mammograms) the nipple was not clearly visible as an image feature. In these cases automated detection gave unreliable results that could not be used to define correspondence between a set of breast shapes. In cases where the nipple is not a visible image feature, a radiologist can make a subjective decision on the nipple's position based on the position and orientation of ducts in the mammogram. Attempting to mimic this decision in automated process was beyond the scope of this thesis, and in section 10.3.1 we discuss an alternative choice of marker.

In the next section we describe how we use the breast segmentation algorithm to build a model of mass location.

10.3 Modelling mass location

As stated in section 10.1, the aim of this chapter is to develop a method for selecting a suitable location within a normal mammogram at which to generate a synthetic mass.

Given a set of normal mammograms, we want the set of possible locations for synthetic masses generated by the method to be consistent with the locations of real masses in the training data.

To achieve this we build a model of the location of masses within the training data. As a result, sampling from the model generates new locations for masses consistent with the training data. The model describes the probability of finding a mass in a particular location in a mammogram based only on the co-ordinates of the location and the breast shape.

As noted in section 10.1, a method has been developed by Caulkin [23] for building a model of mass location within MLO mammograms that have been manually segmented. In section 10.3.1 we show how we have adapted this method to work with the automatic segmentation algorithm described in the previous section, and show the location model obtained by applying the method to our training data. We then show how we have adapted the method to work with CC mammograms and again show the results of applying the method to our data. Finally, we describe how new locations can be sampled from the models and make an important amendment to ensure that synthetic masses are not located too close to the skin-air boundary.

10.3.1   Modelling mass location in MLO mammograms

The method described in this section is based on a method for modelling the distribution of masses in MLO mammograms described by Caulkin [23]. Caulkin's method used a manual delineation of the skin-air breast boundary and the edge of the pectoral muscle. In addition, the position of the nipple in each mammogram was marked by a radiologist (in cases were the nipple was not clearly visible, the position was estimated by the

radiologist). In this section we review the key points of Caulkin's method, describe the changes we have made to make the method compatible with automatically obtained delineations of the breast and pectoral muscle, and show the model of mass location obtained by applying the new method to our data.

Computing a mean breast shape

Before the locations of masses within different mammograms can be compared, a coordinate frame within a common breast shape must be defined. A location for each mass relative to the common breast shape can then be computed, and thus the distribution of the mass locations modelled.

Learning a common breast shape requires fixing a set of landmark points consistently on the breast of each mammogram in the training data. In Caulkin's method, the skin-air boundary in each mammogram was rotated and translated so that the straight line marking the pectoral muscle formed a vertical axis and the perpendicular line between the pectoral muscle and the nipple formed the horizontal axis. As discussed in section 10.2.4, the nipple is not a feature found reliably by our segmentation algorithm and thus cannot be used to define landmark points on each breast border. Instead we use the centre of mass of the breast region (in which each pixel is assigned an equal weight regardless of grey level) and fix the horizontal axis to be the horizontal line connecting this point and the pectoral muscle, as depicted in Figure 10.9 (a).

Given the new axes, landmark points were spaced evenly along the skin-air boundary. Start and end points for the landmarks were defined by computing the points of maximum negative curvature in the sections of the skin-air boundary above and below the nipple respectively. At any point on the skin-air boundary, curvature was calculated

using the chord-length method described in appendix A.3. Using this method positive curvature was associated with a convex shape relative to the pectoral muscle and negative curvature associated with concave shape. The two points of maximum negative curvature computed for a skin-air boundary from the training data are shown in Figure 10.9 (b). In addition to start and end points on the skin-air boundary, start and end points were defined on the pectoral edge at the same height (on the rotated axes), also shown in Figure 10.9 (b).

Once a standard set of correspondence points have been assigned to each breast shape, the set of shapes are aligned using Procrustes alignment (see sections 4.2.1 and 7.4.1). Note that unlike aligning mass shapes in section 7.4, the points on the breast shape have a fixed physical correspondence, and thus the origin of the points on each breast shape is not allowed to vary in the Procrustes Algorithm.

Following alignment, the mean shape of all the breasts can be computed. However, before proceeding to building the location model we must make sure no unrealistic segmentations have been used in generating the mean breast shape. When using manual segmentation, this can be taken as given, however using an automatic segmentation there may be some breast borders that have been incorrectly delineated, producing shapes that are not true instances of breast shape. One solution could be to manually inspect the breast border generated for each mammogram in the training set. However given both the size of the training set and the fact we need to perform similar error checking during mass synthesis we prefer an automated solution.

Figure 10.9: a) Axes fitted to the skin-air boundary and pectoral muscle for an MLO mammogram b) Correspondence points fitted to the breast shape shown in (a)
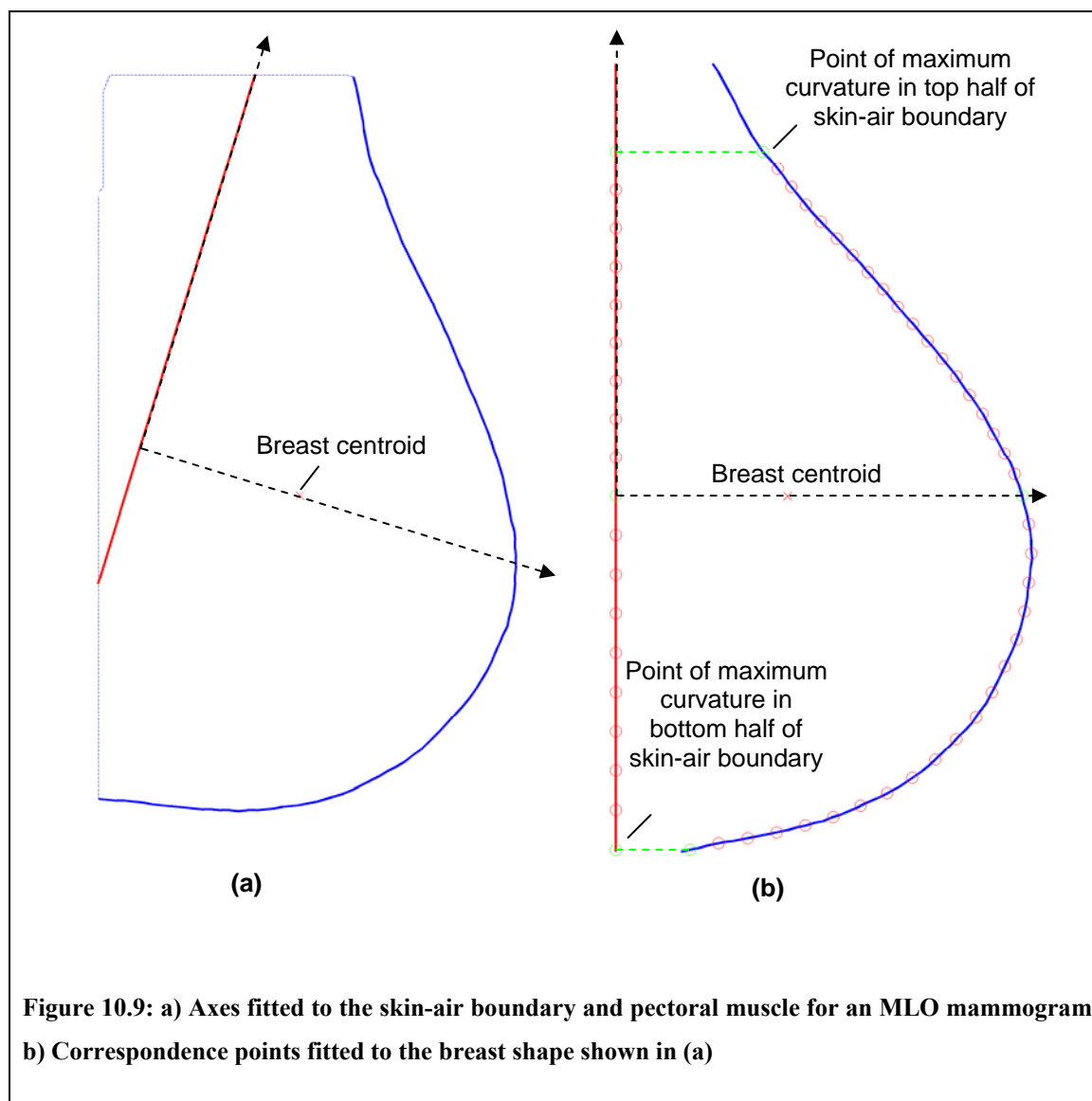
This is achieved by computing a point distribution model (see section 4.2.1) from the set of aligned breast shapes. The model reconstruction error can be computed for each shape, as previously described in section 7.3.3. The mean and standard deviation of the set of model reconstruction errors is computed. Breast shapes with an error greater than three standard deviations of the mean are regarded as outliers and discarded from the training data. We then realign the remaining shapes and recalculate the mean breast shape. The mean shape error threshold at which shapes are discarded is stored and will

be used again when target locations are sampled from the location model, as described in section 10.3.3.

Figure 10.10: Mass centres transformed into the mean breast shape: a) MLO mammograms; b) CC mammograms
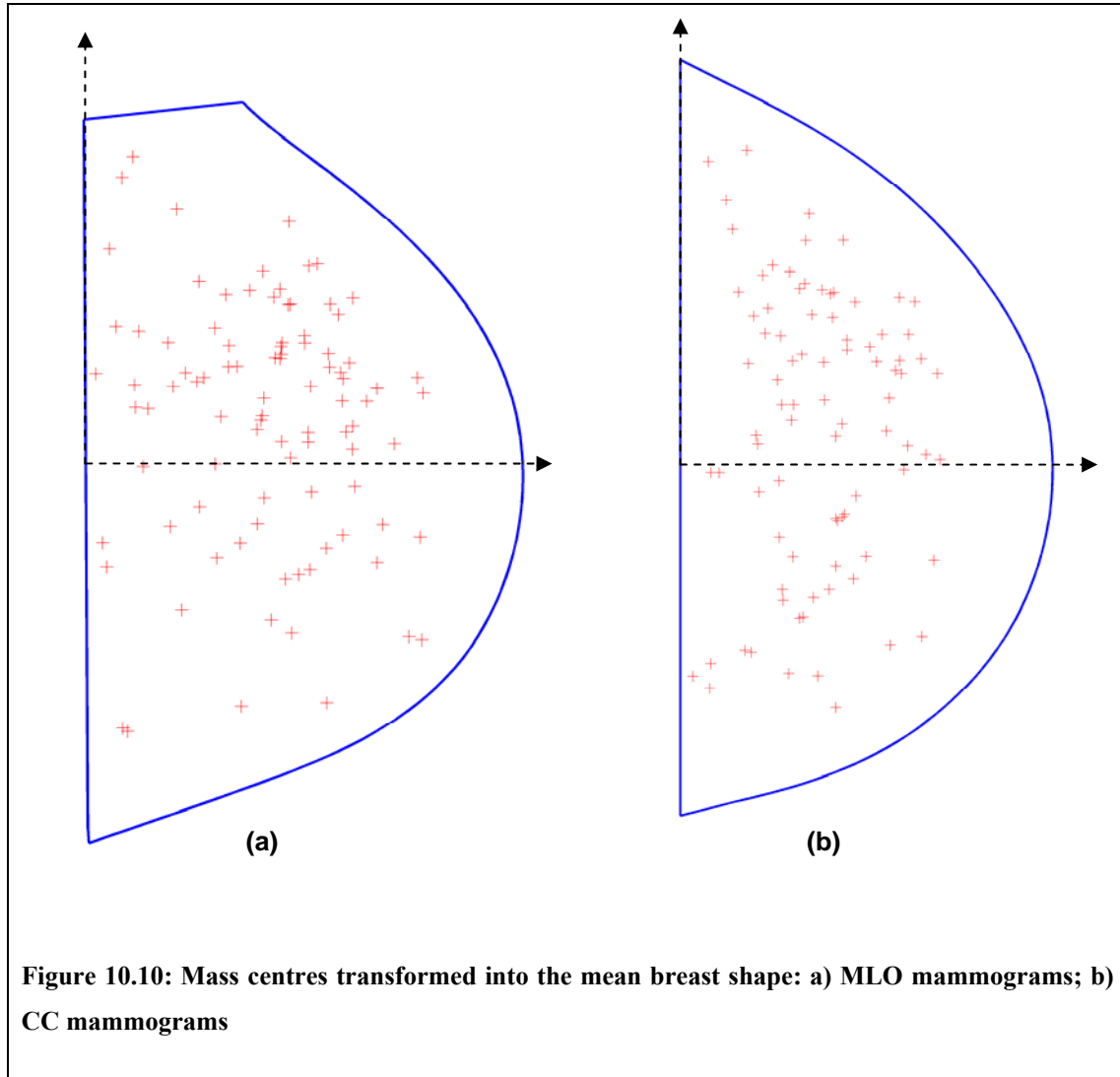
Note that because breast shape is a property of all mammograms regardless of whether they contain malignant masses, all 270 MLO mammograms from our training data were used in computing the mean breast shape. We note that a large malignancy near the skin may change the shape of a breast, however given that our mammograms were acquired from a screening population we would not expect such masses to present in our data.

Thus we decided it would be acceptable to use both normal and abnormal mammograms when computing mean breast shape. Including the normal mammograms gave a large set of data, improving the generality of the breast shape model.

Estimating mass location probability density using fixed Gaussian kernels

Having established a mean breast shape to act as a common co-ordinate frame for all MLO mammograms, we can build a model of mass location following the method described by Caulkin [23].

The first step in the method is to compute the position of the centre of each mass in the training data with respect to the co-ordinates of the mean breast shape. This is achieved by warping the breast shape containing each mass to the mean breast shape, using the thin-plate spline warping algorithm described in appendix A.2.1. This is applied to all masses present in MLO mammograms. As a result there are 103 locations identified as mass centres within the mean breast shape, as depicted in Figure 10.10 (a).

To compute a probability density map of mass location we fit a Gaussian kernel to each mass centre. The probability density at any location in the mean breast shape can then be computed as the normalised sum of the kernels. In the final probability distribution, the standard deviations of the Gaussian kernels are adaptively scaled at each mass centre, however to compute the scaling factors we must first compute a fixed kernel distribution.

Assume we have N mass centres and label the co-ordinates of the i-th centre (x_i, y_i). Then if we fit fixed Gaussian kernels with standard deviation σ_x and σ_y (with no cross

covariance between x and y) at each mass centre, the probability density at any position (u,v) in the mean breast shape can be computed as:

$$D(u,v) = \sum_{i=1}^{N} A \exp\left[-\frac{(u-x_i)^2}{2\sigma_x{}^2}\right] \exp\left[-\frac{(v-y_i)^2}{2\sigma_y{}^2}\right] \quad (10.1)$$

where A is a normalisation constant so that D sums to unity.

Values for σ_x and σ_y are chosen by considering the x and y dimensions to be two separate 1-D kernel density estimation problems in which the smoothing parameter σ must be optimised. Following the method described by Hall [51], we compute σ_x as:

$$\sigma_x = 1.06 N^{-1/5}\left[(N-1)^{-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2\right]^{1/2}$$

Substituting y for x produces an equivalent formula for σ_y.

Using equation (10.1), the probability density for the location of masses within the mean MLO breast shape was computed. This is depicted in Figure 10.11 (a).

Estimating mass location probability density using adaptively scaled Gaussian kernels

Calculating a probability density map for mass location using fixed Gaussian kernels, as described above, suffers from a drawback. In regions of the mean breast shape where there are few mass centres, the kernel placed on each mass centre causes a local spike in the probability density. To overcome this problem, we fit scaled Gaussian kernels to each mass centre, where the scaling is adapted depending on the local density at each mass centre. Thus mass centres in low density regions will have a larger (and

consequently smoother) Gaussian kernel fitted than mass centres in regions of high density.

To compute the scaling factors used at each mass centre, we again follow the method described by Hall [51]. This uses the initial fixed kernel density estimation calculated above to compute a scaling factor λ_i for the i-th mass centre:

$$\lambda_i = \left[\frac{D(x_i, y_i)}{g} \right]^{-1/2}$$

Where g is computed as the geometric mean of the initial density estimate such that:

$$g = \exp\left[N^{-1} \sum_{i=1}^{N} \log D(x_i, y_i) \right]$$

Having computed scaling factors for each mass centre, the final probability density for any point (u, v) within the mean breast shape can be computed as:

$$D_a(u, v) = \sum_{i=1}^{N} A \exp\left[-\frac{(u - x_i)^2}{2(\lambda_i \sigma_x)^2} \right] \exp\left[-\frac{(v - y_i)^2}{2(\lambda_i \sigma_y)^2} \right]$$

where again, A is a normalisation constant so that D_a sums to unity.

The result of applying this method to compute a probability density map for the location of MLO masses using the 103 masses in our training data is depicted in Figure 10.11 (b). The distribution of this density is discussed in section 10.3.4.
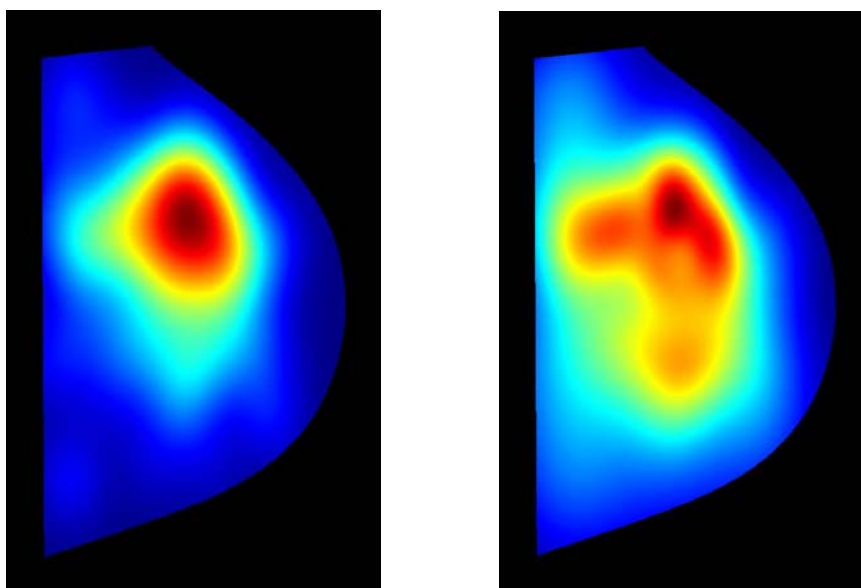
Figure 10.11: Probability maps for the location of masses in MLO mammograms a) Using Gaussian kernels of fixed standard deviation b) Using adaptively scaled Gaussian kernels
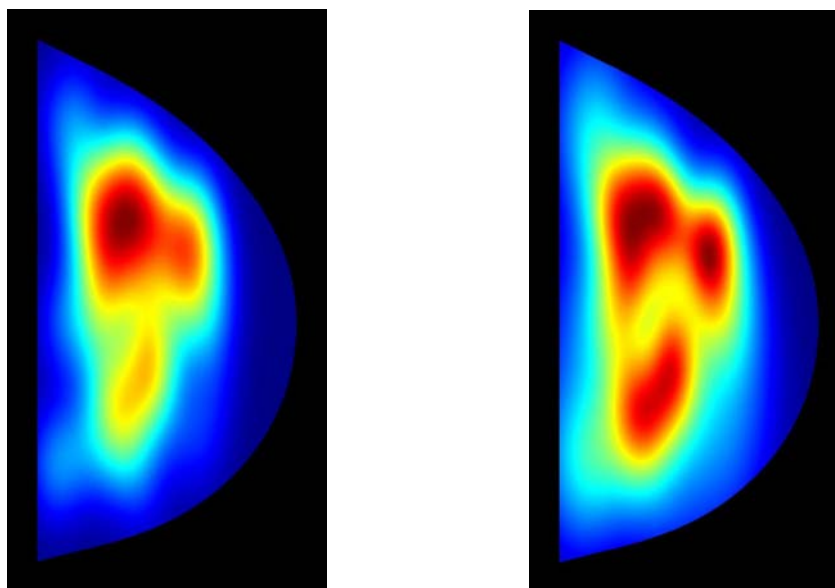
Figure 10.12 Probability maps for the location of masses in CC mammograms a) Using Gaussian kernels of fixed standard deviation b) Using adaptively scaled Gaussian kernels

10.3.2   Modelling mass location in CC mammograms

We also want to build a model of mass location for CC mammogram, so that synthetic

masses may be automatically generated in a mammogram of either view. In our training

data we have 93 masses from CC mammograms and a total of 270 CC mammograms

(regardless of whether a mass is present or not). Using these data we can build a model

of mass location in CC mammograms based on the method described in the previous

section.

Breasts in CC mammograms do not have the same features as breasts in MLO

mammograms: the pectoral muscle is not present, and the curvature of the skin-air

boundary at the top and bottom of each mammogram differs. Thus we must define a

new set of correspondences to define the shape of each breast in a CC mammogram.

In most CC mammograms, the breast forms an approximate semi-circle so that the skin-

air boundary starts and end at the chest wall (as seen, for example, in Figure 10.6 (a)).

However, in some mammograms of larger breasts, the skin-air boundary may reach

either the top or bottom (or both) of the mammogram before reaching the chest wall. In

these cases we extrapolated the ends of the skin-air boundary until they met the chest

wall. Thus for every CC mammogram we have start and end points for the skin-air

breast border located at the chest wall. We then calculate the centroid of the breast

region, and fit a set of axes so that the vertical axis runs along the chest wall and the

horizontal axis passes through the breast centroid. Again, if it could be determined

automatically, we would rather use the nipple as a marker for the horizontal axis (see

section 10.2.4). The intersection of the horizontal axis with the skin-air boundary is

taken as a third feature point on for each breast, thus dividing the skin-air boundary into two segments. Correspondence points may then be spaced equally along each segment of the skin-air boundary and along the chest wall to define each breast shape.

Having fitted a common set of correspondence points to each CC breast shape we can implement the remainder of the model building method as described in section 10.3.1. To summarise, the steps in this process are:

i)   Align the shapes using Procrustes analysis

ii)   Compute the principal components of a shape model and compute the reconstruction error for each breast shape

iii) Discard any outlier shapes, realign, and calculate the mean CC breast shape

iv)   Compute the position of the centre of each CC mass in the training data relative to the mean breast shape by thin-plate spline warping each breast to the mean shape

v)   Fit a fixed Gaussian kernel to each mass centre in the mean breast shape to compute an initial probability density of mass location

vi)   Fit adaptively scaled Gaussian kernels to each mass centre based on the initial density estimate. The resulting density is normalised and taken as the probability density for the location of a mass at any position within the mean breast shape

The probability density for the location of CC masses computed by applying the method

described above to our training data of 93 masses is shown in Figure 10.12 (b). As with the probability density map computed for MLO mammograms, the distribution of this density will be discussed further in section 10.3.4.

10.3.3   Sampling locations from the models

Using the models of probability density for mass location built in sections 10.3.1 and 10.3.2, we can sample a location at which to generate a synthetic mass in any normal mammogram. The steps required to do this are described below:

i)   Apply the segmentation algorithm described in section 10.2 to obtain the breast border and, if the normal mammogram is an MLO, the pectoral muscle edge

ii)   Fit the applicable CC/MLO correspondence points to define the breast shape of the mammogram. Reconstruct the breast shape from the appropriate CC/MLO breast shape model to check a valid segmentation of the breast has been obtained

iii)   Transform the breast shape to the CC/MLO mean breast shape

iv)   Sample a new location from the CC/MLO probability density map

v)   Calculate the co-ordinates of the sampled location with respect to the breast shape of the normal mammogram, based on the transform computed in step (iii)

However, before this method can be used to generate normal regions in which to generate synthetic masses, we must make a further adjustment to ensure that unrealistic

325

mass locations are not sampled from the models.

We know that masses must be contained within the breast, and therefore must not cross the skin-air boundary (masses may however be located across the chest wall in CC and MLO mammograms or overlapping the pectoral muscle in MLO mammograms).

The maps of probability density developed in sections 10.3.1 and 10.3.2 extend to the skin-air boundary. Because no masses in the training data were located at the skin-air boundary, the probability density in both MLO and CC maps in the immediate vicinity of the skin-air boundary is very small. However, because it is non-zero, there is still a chance that a location could be sampled in this region.

For example, the average radius of a circle bounding the mass regions in our training data is approximately 20mm[13]. The sum of probability densities in a region of this radius inside the skin-air boundary of the CC distribution is 0.074. Thus approximately 1 in 14 locations sampled from the model would be an unrealistic location for an average sized mass in our training data.

Our solution is to set the density to zero in a region of some radius inside the skin-air boundary, with the radius depending on the size of the mass for which we are sampling a location. The dependence on the size of the mass is a critical point. Locations near the skin-air boundary may be suitable for small masses, whilst for a synthetic mass of size similar to the largest masses in our training data, there is only a small region of suitable locations near the centre of the breast shape.

---

[13] By bounding radius, we mean the greatest distance from the centre of the mass to the end of any structure associated with the mass

Thus given a normal mammogram and a radius that bounds the mass to be generated, we amend the sampling procedure presented at the start of this section as described below:

i)   Apply the segmentation algorithm described in section 10.2, to obtain the breast border (and pectoral muscle edge if necessary)

ii)   Fit the applicable CC/MLO correspondence points to define the breast shape of the mammogram and check the validity of this shape

iii)   Transform the breast shape to the CC/MLO mean breast shape

iv)   Rescale the radius with respect to the mean breast shape based on the transform learnt in step (iii)

v)   Set the probability densities of the CC/MLO location model to zero in the region inside the skin-air boundary, based on the radius computed in step (iv)

vi)   Re-normalise the probability densities and sample a location from the model

vii)   Calculate the co-ordinates of the sampled location with respect to the breast shape of the normal mammogram, based on the transform computed in step (iii)

If we want to obtain a normal region in which to place a synthetic mass based on the sampled location, we simply extract a square region with the sampled location at its centre. The dimensions of the square are selected to be just bigger than the bounding

radius of the mass to be generated.

The method described above can now be used to automatically select new regions in which to generate synthetic masses from a set of normal mammograms.

10.3.4   Discussion

In sections 10.3.1 and 10.3.2 we have shown how we can build models of mass location for MLO and CC mammograms respectively.

The method for building the model for the location of MLO masses was based on an earlier method presented by Caulkin [23]. Our method adapted the earlier model so that it could be used with automatic (as opposed to manual) segmentations of the breast.

Applying the new method to the MLO mammograms in our training data produced the probability density map shown in Figure 10.11 (b). We note that the distribution of densities in this map shows the most likely area for masses to be located is in the upper half of the mean MLO breast shape. This is consistent with the probability density generated by Caulkin.

Looking at the probability density for CC mammograms, depicted in Figure 10.12 (b), we see that again the greatest regions of density are found in the upper section of the mean CC shape.

Using the inner/outer upper/lower quadrant terminology for describing regions of the breast used by radiologists (as discussed in section 2.1.1), we see that the most likely location for masses given both the MLO and CC density maps is the upper-outer

quadrant of the breast[14]. This is consistent with the literature on where malignant masses are most frequently located – as for example, described in the review by Sickles [108] and confirmed by latest statistics on breast cancers in the UK [4].

In section 10.3.3, we described how the models of mass location could be used to sample new locations for generating synthetic masses in normal mammograms. To ensure that unrealistic location near or on the skin-air boundary were not sampled from the models, we adapted the probability density map depending on the size of synthetic mass to be generated. The size of a mass was specified in terms of the radius of a circle required to bound the mass.

Note that if we are using the method of directly transferring DT-CWT coefficients to synthesise the appearance of a mass region in a normal region (as described in section 8.5), then the bounding radius for the mass to be generated can be obtained from the real mass used as a template in the algorithm. Thus we would execute the first step of Algorithm 8.1, prior to selecting a normal region using the method described in the previous section. If we are using a model-based algorithm to modify a normal region (see sections 8.6 and 8.7), then we must sample a bounding radius for the synthetic mass prior to modifying the region and before selecting the normal region.

Note also that because the model of mass appearance (described in chapter 7) learned the shape and texture of masses in conjunction with mass size, the adapted models of mass location implicitly link mass appearance with mass location.

---

[14] Although we note that without using the nipple as a marker, it is perhaps not valid to make a direct comparison with breast quadrants

The method described in this chapter for sampling synthetic mass locations in normal mammograms can be used with any mass synthesis method. The only requirement of the synthesis method is that it supplies a suitable measure of mass size so that the location probability density map can be adapted accordingly. Thus the work in this chapter may be a useful addition to other synthesis methods. Similarly it could easily be adapted to describe the location of other disease signs in mammograms (for example microcalcifications).

It is important to note, however, that the models in this chapter only take into account the global shape of the breast in which an abnormality is sampled. How real the synthesised abnormality appears will depend on the degree to which a synthesis method takes into account the local appearance of a location sampled from the model. Our method for synthesising malignant masses has been designed explicitly to adapt to local appearance properties of the region in which masses are generated.

Even with a method for adapting local appearance, it could be argued that the global model of location should consider more information than mass location relative to the breast shape. For example, an attempt could be made to include information on the shape of the fibroglandular disc, as this may influence where masses are most likely to be located in the breast. However, the fibroglandular disc is not a consistent feature across all mammograms. For example, particularly fatty breasts may have no obvious internal structures but can still develop masses. For this reason, using information other than breast shape to construct a global model of mass location has not been explored further in this thesis.

10.4 Summary

In this chapter we have shown how we can build a statistical model to describe the probability of finding a mass at a particular location within a breast. This probability is based only on the relative location of masses relative to the shape of the breast.

The breast shape within any mammogram was defined with respect to the skin-air boundary of the breast, and, for MLO mammograms, the edge of the pectoral muscle. These features were found automatically for the mammograms in our training set using the segmentation algorithm described in section 10.2.

Given the skin-air boundary and pectoral muscle edges for a set of MLO mammograms, section 10.3.1 showed how a model of mass location could be constructed. The model was based on one previously described by Caulkin [23]. Section 10.3.2 then adapted the model for MLO mammograms to be compatible with CC mammograms.

In section 10.3.3 we described how the models could be used to sample new locations at which to generate synthetic masses in normal mammograms. The set of locations at which a mass could be sampled depended on the size of the mass. This ensures that when synthetic masses are generated they will not encroach on the skin-air region of mammograms.

As with the methods presented in previous chapters, an important feature of the method for sampling new mass locations is that it can run without user input. Given only a normal mammogram, and information on whether the mammogram is an MLO or CC

view of the left or right breast, the complete process described in section 10.3.3 can be run automatically to select a normal region in which to generate a synthetic mass.

Combining the mass location method described in this chapter with our method for synthesising mass appearance described in Algorithm 9.1 produces a complete method for automatically generating malignant mammographic masses in normal mammograms.

In following chapter we show how we have evaluated the realism of synthetic masses generated by the complete method.

Chapter 11 : Evaluating the Appearance of Synthetic Masses

11.1 Introduction

In chapter 10 we described a method for sampling suitable locations within a normal mammogram at which to synthesise a malignant mass. In chapter 9, we described a method for synthesising a mass in a normal mammogram region. Combining the two approaches means we have a complete method for generating synthetic malignant masses in normal mammograms that can be run without requiring user input. Thus given a set of normal mammograms, we can automatically generate a set of synthesised malignant mammographic masses.

As discussed in section 2.2.7, the principal application in which we envisage synthetic masses could be used is as part of training software for reading mammograms. The synthetic masses would be used in place of real examples so that despite having only a fixed set of real masses, the software could generate a potentially infinite range of masses to display to trainees.

Therefore the ultimate test of whether our method has been successful is to evaluate if masses generated using the method appear realistic to experts at reading mammograms. In this chapter we describe details of an observer study in which mammography readers at the Nightingale Breast Centre, Manchester attempted to distinguish between real masses and synthetic masses generated using our method.

In the next section we present a general overview of the observer study. Sections 11.3 to

11.5 provide full details of the study.

11.2 An observer study of synthetic mass realism

To assess the realism of synthetic masses generated from our method, we designed an observer study in which mammography readers at the Nightingale Breast Centre, Manchester attempted to distinguish between real masses and masses we had synthesised.

Ten expert mammography readers were asked to rate a randomised set of 30 real and 30 synthetic masses on 5-point scale ranging from 'definitely synthetic' to 'definitely real'. Having rated all 60 masses, the readers were invited to give feedback on masses they had rated as either definitely or probably real. The readers completed the task individually and independently, using software we had developed for the study.

Statistical analysis of the pooled ratings was performed to assess whether synthetic masses were identified at a rate significantly better than chance. Our hypothesis was that readers would not be able to identify synthetic masses, and thus there would be no statistically significant difference between the ratings assigned to synthetic masses and the ratings assigned to real masses.

In addition to the pooled analysis, the ability of individual readers to identify synthetic masses was measured and used to assess if a reader's experience or job type affected their performance in the task.

The feedback provided by the readers was used to make a qualitative assessment of our method for generating masses.

The study was carried out in collaboration with a 4th year medical student, David Barbosa Da Silva, who undertook data collection by training readers to use the software and supervising reading sessions. We were responsible for the design of the experiment, production of the software and the data analysis presented here.

In the following sections we provide full details of the design, implementation and results of the study.

11.3 Data

In this section we describe the data used in the observer study. In particular, the construction of a test set of 30 synthetic and 30 real mammographic masses is described, and details of the mammogram readers that participated in the study are presented.

11.3.1 Dataset of real and synthetic masses

To generate a test set of synthetic masses, 30 normal mammograms were selected at random from the 277 normal mammograms in our training data. As described in section 5.1.2, these mammograms showed no signs of disease, and this was confirmed by a breast radiologist.

In each normal mammogram a single location was chosen at which to generate a synthetic mass by sampling from the model of mass location, as described in section 10.3.3. A 1024 by 1024 pixel region was then extracted symmetrically about the sampled location. In each of the normal regions a single synthetic mass was generated using Algorithm 9.1, thus producing a final set of 30 synthetic masses.

In addition to the set of 30 synthetic masses, 30 real masses were selected from our

training data. For each real mass, a 1024 by 1024 pixel region was extracted from the original mammogram containing the mass; this was symmetrical about the centre of the mass.

11.3.2 Mammogram readers

All qualified mammogram readers at the Nightingale Breast Centre, Manchester were invited to participate in the study. 10 readers accepted the invitation, of which 4 were breast radiologists and 6 were either breast physicians or radiographers that had undergone further training to enable them to read and report on mammograms. The experience of the readers varied broadly, from a consultant radiologist of 20 years experience to a radiographer with just 1 year experience reading films. All participants were reading at least 5,000 mammograms per year, up to a maximum of 10,000 films.

11.4 Method

In this section we describe the design and implementation of the study, and methods for the analysis of the results.

11.4.1 Study design

The 60 mass regions were shown to each reader participating in the study in a random order. The readers were asked to rate each mass as either "definitely synthetic", "probably synthetic", "possibly real", "probably real" or "definitely real". The ratings provided by each reader were recorded and used to quantitatively evaluate the realism of the set of synthetic masses.

The ratings assigned by each participant in the study were recorded anonymously, but the experience of each participant and their job type were recorded along with the ratings they gave. This allowed us to check for any correlation between ability to

identify synthetic masses and reader experience. In addition, the time taken by a reader to analyse each mass was recorded, to assess whether this affected reader performance.

Having confirmed a rating for each mass, the reader was not able to review or amend their answer. However, at the end of the task, readers were subsequently presented with masses they had described as definitely or probably synthetic (regardless of the true status of the mass) and asked to identify why they made that rating. Whilst not used in the quantitative evaluation, this feedback was used to provide detailed information on what, if any, aspects of our method for generating masses needed improving.

For each mass included in the feedback section, a reader was able to select as many of the following options as they felt were appropriate:

- Mass located incorrectly with respect to breast tissue

- Not enough spicules

- Too many spicules

- Mass should be more radio-opaque given its size

- Unrealistic texture at mass boundary

- Mass border too regular

- Mass border too sharply defined

We also invited each reader to make any additional comments on the appearance of a mass. The options were designed to gain as informative feedback as possible.

As an alternative to providing feedback, the readers could state that they no longer thought a mass was synthetic. However this was not used to amend their earlier rating.

11.4.2   Study implementation

The readers participated in the study individually and independently. Thus data collection consisted of ten separate reading sessions, each involving a single reader. During each session, the 60 test masses were displayed using software we had developed for the study.

The interface of the software was designed so that each mass region was displayed at maximum resolution on the screen, alongside a simple set of radio buttons with which to rate the mass. This allowed the reader to concentrate their efforts on identifying the synthetic masses, and complete the task unaided. A screen shot of the software is shown in Figure 11.1.

Given the spatial resolution of mammograms in our data, regions of 1024 by 1024 pixels correspond to approximately 5cm by 5cm regions of film. The area of screen occupied by each mass region was approximately 15cm by 15cm, and thus each mass was effectively enlarged by a factor of 3. Because the masses were already enlarged, the readers were not provided with further zoom functionality.

All mass ratings, feedback and timings were recorded automatically by the software. In addition, for each reader, we recorded the approximate number of films they read per year and the number of years they had been qualified as a mammography reader. These data were collected and recorded by the collaborating student (see section 11.2).
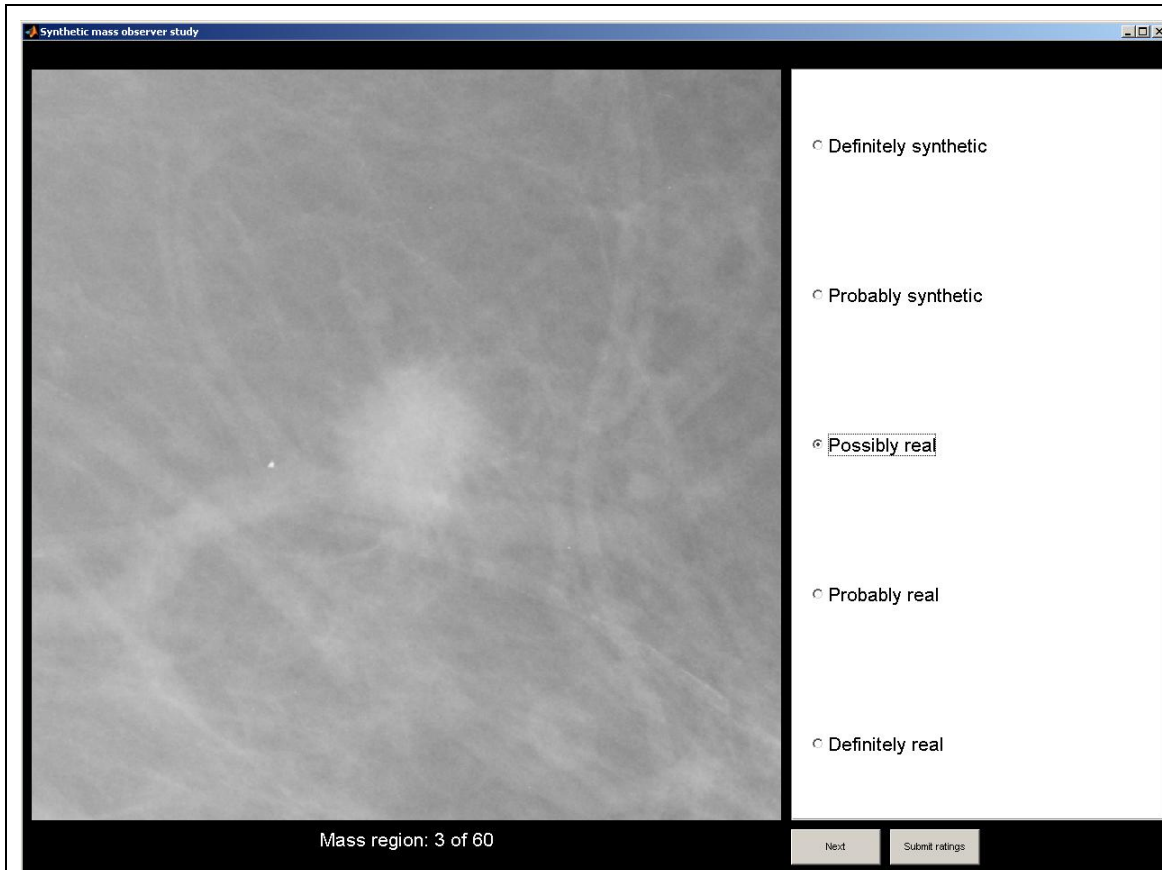
Figure 11.1: Screen shot of the software used by readers during an observer study. For each mass display to a reader they selected one of the options on the right, before clicking 'Next'. When all 60 masses had been rated the could submit their ratings

11.4.3 Statistical analysis

To enable statistical analysis of the ratings given by readers in the study, values were assigned to each rating so that "definitely synthetic" scored 1 through to "definitely real" scoring 5.

In the primary statistical analysis, the ratings from all readers were pooled. Labelling the rating assigned to the *i*-th real mass by the *j*-th reader as $\Phi_{R,j}(i)$, we computed the pooled mean score for each real mass

$$\overline{\Phi_R}(i) = \sum\nolimits_{j=1}^{10} \Phi_{R,j}(i)\big/10$$

Similarly, we computed the pooled mean score for each synthetic mass

$$\overline{\Phi_S}(i) = \sum\nolimits_{j=1}^{10} \Phi_{S,j}(i)\big/10$$

If readers were able to distinguish between real and synthetic masses we would expect the scores in $\overline{\Phi_R}$ to be higher than the scores in $\overline{\Phi_S}$. This was tested statistically by applying the Mann-Whitney U-test against the null hypothesis that $\overline{\Phi_R}$ and $\overline{\Phi_S}$ are samples from equivalent distributions.

In the secondary analysis, we returned to the individual ratings assigned by each reader. We fixed operating points $0 \leq \tau \leq 5$ to make a binary split on the ratings assigned by a reader, such that a rating greater than τ was taken to mean real and ratings less than or equal to τ were taken to mean synthetic. We defined a true positive (TP) as a real mass that has been rated as real, and a false positive as a synthetic mass that has been rated as real and computed the true positive rate (TPR) and false positive rate (FPR) of the j-th reader's scores as

$$TPR = \frac{\left|\Phi_{R,j} > \tau\right|}{30}$$
$$FPR = \frac{\left|\Phi_{S,j} > \tau\right|}{30}$$

where the operator $\left|\Phi > \tau\right|$ means "count the number of elements in Φ that are greater than τ".

By plotting TPR against FPR for values of τ varying from 0 to 6, we generated a receiver-operating-characteristic (ROC) curve for each reader. The area under each ROC curve (AUC) was computed as a measure of the overall ability of each reader to distinguish between real and synthetic masses. The closer a reader's AUC value was to 1, the better they were able to differentiate between real and synthetic masses. In contrast, an AUC of 0.5 would suggest a reader identified synthetic masses at a rate no better than chance.

Using the AUC value for each reader, we tested whether reader experience or job type had a significant impact on their ability to identify synthetic masses. As stated above, for each study participant, we recorded the approximate number of films they read per year and the number of years they had been qualified as a mammography reader. To get an overall measure of experience for each reader we multiplied the two numbers to obtain an approximate total number of films read. Where a range was given for the number of films read per year, the central value in the range was used in the multiplication. Due to the approximate nature of this measure, we chose to use it only as a method for ranking the readers in terms of experience, with rank 1 being the most experienced. We then computed Spearman's rank correlation coefficient to assess with there was any significant correlation between reader experience and reader AUC.

Similarly, the AUC value for each reader was used assess whether there was any significant correlation between a reader's ability to identify synthetic masses and the time they spent analysing each mass in the study. Labelling the time (in seconds) it took for the the j-th reader to rate the i-th mass (whether real or synthetic) as $T_j(i)$, we computed the mean rating time for each reader as

$$\overline{T_j} = \sum\nolimits_{i=1}^{60} T_j(i)/60$$

We then computed Spearman's rank correlation coefficient between each reader's AUC value and their mean rating time.

The summary statistics of the set of reader AUC values were also used to compare the results of our study to other previous studies in the literature, as discussed in section 11.6.3.

11.5 Results

In this section we present the results of the observer study, including: a quantitative analysis of the ratings assigned to each mass (section 11.5.1) and an analysis of the qualitative feedback provided by readers during the study (section 11.5.2).

11.5.1   Statistical analysis of mass ratings

As described in section 11.4.3, the primary statistical analysis performed in the study was a test on the pooled ratings assigned to each mass by the participating readers. The null hypothesis in our test was that the pooled ratings assigned to real masses were not significantly higher than ratings assigned to synthetic masses. The test strongly rejects the null hypothesis (Mann-Whitney U-test, $p < 0.0001$) giving evidence that readers were able to identify synthetic masses at a rate statistically better than chance.

Following the primary analysis we computed ROC curves each individual reader. The set of ten ROC curves along with their AUC values are depicted in Figure 11.2. The mean AUC of all readers was 0.70 with a standard deviation of 0.09.

The AUC values, along with the data on reader job type, experience and films read per

year are summarised in Table 11.1.

Table 11.1: Summary of reader experience and AUC value for participants in the observer study

	Radiologists				Radiographers/Breast Physicians					
Years experience	20	17	7	1	8	6	5	7	1	1
Films read per year	7,000	5,000-6,000	7,000	7,000	8,000-10,000	6,000-7,000	7,000-8,500	>5,000	5,500	5,000
Experience ranking	1	2	4	8	3	5	6	7	9	10
AUC	0.55	0.88	0.58	0.77	0.67	0.70	0.70	0.73	0.70	0.67
	0.70 ± 0.16				0.71 ± 0.03					
	0.70 ± 0.09									

Figure 11.3 shows the AUC values for readers ranked in terms of experience. There was no trend in AUC values with increasing or decreasing experience (Spearman's rank correlation coefficient $\sigma = 0.47$, $p = 0.18$). This implies a reader's experience did not have an impact on their ability to identify synthetic masses.

In Figure 11.3, the AUC bars are coloured red for radiologists and blue for radiographers or breast physicians. Comparing the AUC values for the two groups showed no significant difference, implying job type was not a factor in the observer study.
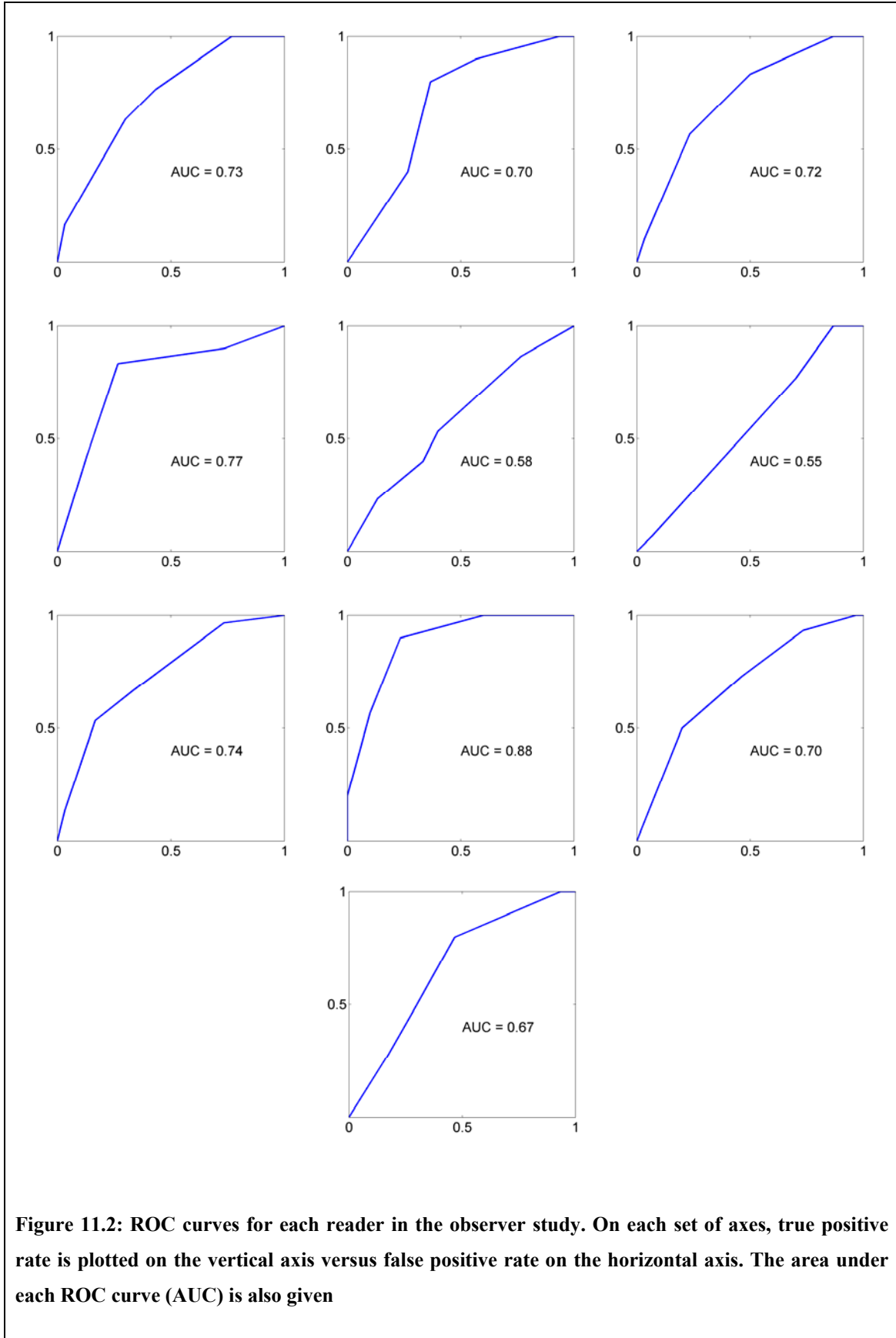
Figure 11.2: ROC curves for each reader in the observer study. On each set of axes, true positive rate is plotted on the vertical axis versus false positive rate on the horizontal axis. The area under each ROC curve (AUC) is also given

Finally, there was no significant correlation between AUC and the mean time in seconds a reader took to analyse each mass (Spearman's rank correlation coefficient σ = 0.58, p = 0.90).

Figure 11.3: AUC values for readers ranked in order of experience (1 = most experienced, 10 = least). Red bars show AUC values for radiologists; blue bars for radiographers or breast physicians
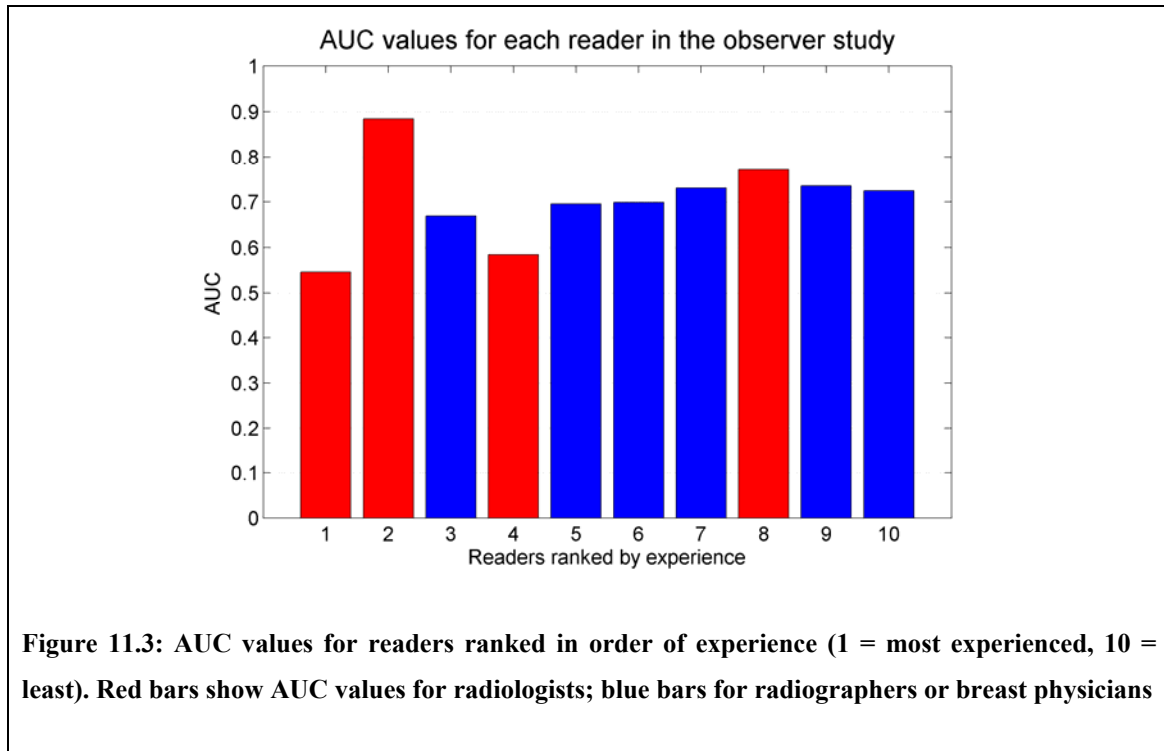
11.5.2 Analysis of reader feedback

As described in section 11.2, when each reader had provided a complete set of ratings for the 60 masses, they were asked to provide feedback for masses they had rated as either definitely or probably synthetic. Although this part of the task was voluntary, it was completed by all ten readers in the study.

For each mass, the reader could either select that they now believed the mass was real or they could give reasons why they believed the mass was synthetic. For the latter, the software listed seven reasons for identifying a mass as synthetic, of which the reader could select as many as they felt appropriate (the list of reasons was given earlier in

section 11.4.1). In addition a reader could enter any additional comments in an open text box.

Because the set of masses rated as definitely or probably synthetic differed for each reader, and because a reader could select as many reasons as they wished for each mass, the feedback does not provide a consistent set of data from which we can make statistical inferences. However analysing the frequency with which particular reasons were selected in the feedback data is useful in understanding what aspects of synthetic masses were unrealistic.

If we sum the number of synthetic masses identified as either probably or definitely synthetic across all readers we compute

$$\sum_{j=1}^{10} \left| \Phi_{S,j} \leq 2 \right| = 156$$

This is the number of times readers were asked to provide feedback on a synthetic mass during the study. On 20 occasions, a reader then stated they no longer believed the mass was synthetic. As a result, each reason for correctly identifying a mass as synthetic could be selected a maximum of 136 times.

By far the most commonly selected reason was "Unrealistic texture at mass boundary" chosen on 77 occasions (52%). The next most commonly selected reason was "Mass should be more radio-opaque given its size". This reason was selected 17 times (13%), however we also note it was selected 14 times when readers were giving feedback on masses they had rated as synthetic but were actually were real. Further analysis shows that one of the ten readers selected this reason 14 times (7 for synthetic masses, 7 for real masses), suggesting that this reader had some perception of how they expected

masses to appear that wasn't evidenced in the real data. No other reason for correctly identifying a mass as synthetic was selected more than 10 times.

Of the additional comments added by readers to synthetic masses, the majority expanded on the notion of unrealistic texture in the locality of the mass. For example, one comment stated, "fat appearance at periphery of mass has poor match with background parenchyma". In addition, one user noted that in three masses a blood vessel appeared to be interrupted at the mass boundary.

Coupled with the statistical analysis performed in the previous section, this feedback is useful in determining what areas of our method for generating synthetic masses need to be improved on. This is discussed further in the following section.

11.6 Discussion

In this section we discuss the choices we made in designing the observer study presented in this chapter. We discuss the results of the study and compare our study to previous evaluations of synthetic mammographic masses reported in the literature.

11.6.1   Study design

To increase the power of our study we wanted to recruit as many mammogram readers as possible. Thus all qualified readers at the Nightingale Breast Centre were invited to participate in the study. The number of readers who actually took part in the study (10) was determined by the availability of readers during the study period.

The total number of masses used in the experiment was chosen to be as large as possible, whilst still allowing each reader to complete the study in a reasonable length of time. Ensuring the study did not take up too much of each participant's time was

essential given the readers' already busy schedules. However we did not want to force the readers to analyse each mass within a fixed period of time. This might cause a reader to rush a decision they were not sure on, and thus bias the results. After initial testing, a total of 60 masses were chosen so that each participant could easily complete the experiment within one hour.

A standard size region for all masses (synthetic and real) was chosen so that there was consistent reference on which to judge the size of synthetic masses generated by our method. The mass regions were displayed at maximum resolution to the readers, and thus were approximately enlarged by a factor of 3. We chose to display enlarged copies of each region so that every reader in the study made a detailed assessment mass appearance. An alternative would have been to display masses synthesised directly in whole mammograms. However, the readers would then have had to zoom in to properly assess each mass. This would have taken up more of the readers' time and could have led to some readers assessing a mass without zooming in – in which case pooling reader ratings to assess mass realism may not have been valid.

The software used in the experiment was designed to allow readers to complete the task unaided. No problems using the software were reported.

11.6.2   Study results

In section 11.5.1 we showed there was statistically significant evidence that readers could identify synthetic masses at a rate better than chance. This implies our method has not achieved our goal of generating synthetic mammographic masses that are indistinguishable from real masses.

Average score = 1.2

Average score = 4.4

Average score = 1.4

Average score = 4.2

Average score = 1.8

Average score = 4.0

Figure 11.4: The three least realistic (left column) and most realistic (right column) synthetic masses by average reader rating. In each case the pooled average score of the 10 readers is shown (1 = definitely synthetic, 5 = definitely real)
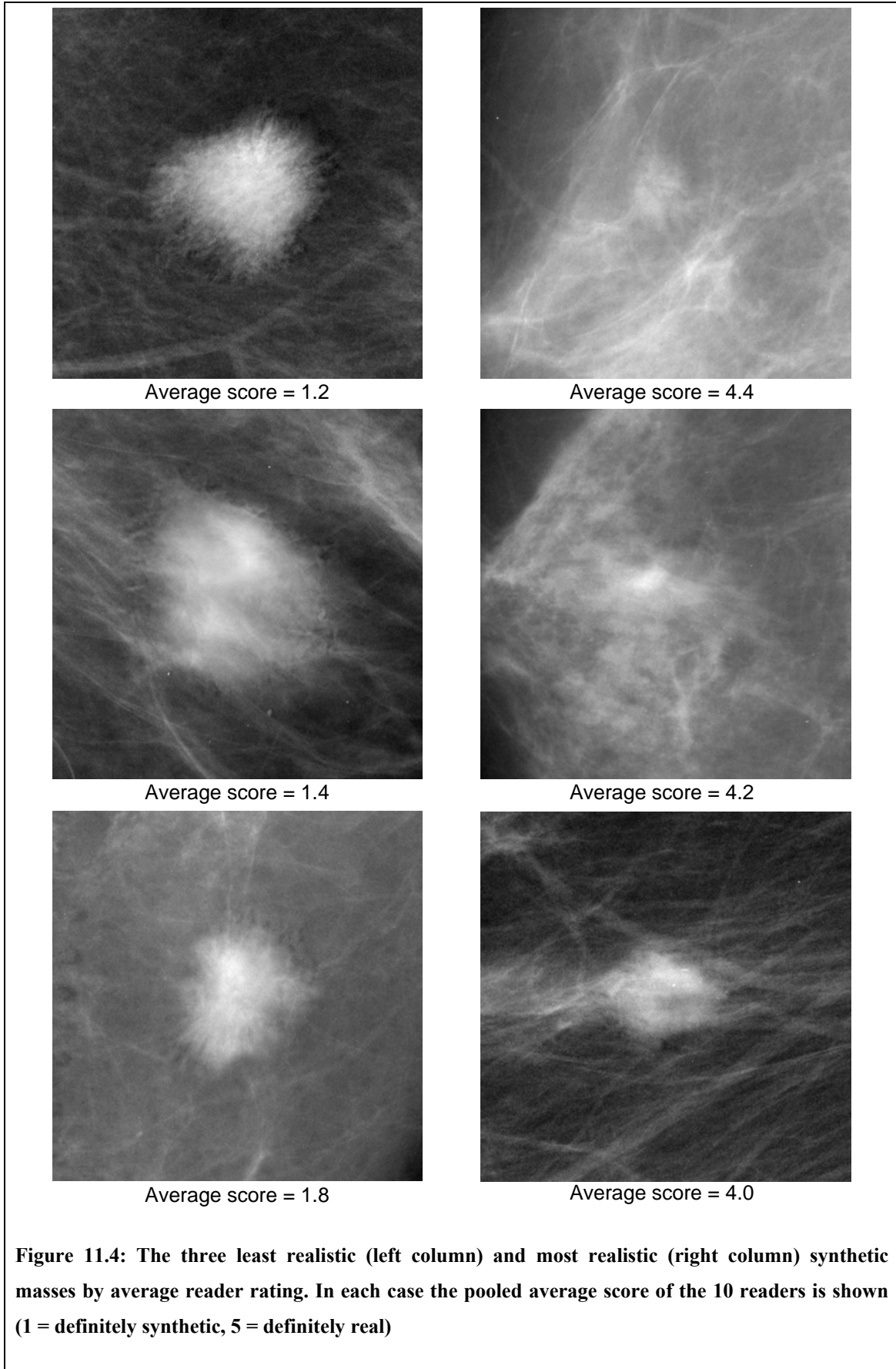
In particular, it appears our method for synthesising mass background appearance requires significant improvement. In section 8.5.3, we noted that Algorithm 8.1 may generate patches of unrealistic texture when synthesising mass background appearance in normal regions. Given the feedback described in section 11.5.2, it would appear this was the major reason that led to masses being correctly identified as synthetic. This can be observed in the left column of Figure 11.4 in which the three least realistic masses by average reader rating are shown. Note how in each case there is a clearly identifiable region of unrealistic texture at the mass border. In contrast, in examples where our texture synthesis algorithm has succeeded, the masses blend seamlessly with the background, as evidenced in the three most realistic masses depicted in the right column of Figure 11.4. As noted in section 8.7, work to improve a method for synthesising mass background appearance is ongoing, and this is discussed further in the following chapter.

It was interesting that the AUC values computed for each reader did not imply that radiologists were any better at detecting the synthetic masses than radiographers or breast physicians, and that there was no correlation between experience and AUC. That the least experienced readers performed at a similar level to radiologists with several decades of experience suggests all the participants in the study could justifiably be termed experts. Additionally, the lack of correlation between the time taken to analyse each mass and AUC implies that where a reader had a lower rate of identifying synthetic masses, this was not simply a result of rushing their judgement for each mass. From this we conclude the readers participating in our study provided a fair assessment of the realism of masses generated by our method.

11.6.3 Comparison to other studies

In the remainder of this section we compare the results of our observer study to results reported for earlier studies evaluating the realism of synthetic masses. In section 3.2, we reviewed several methods for synthesising mammographic masses [17, 23, 54, 99, 101, 102, 112]. We recall that the work presented by Skiadopoulos et al. [112], Saunders et al. [101, 102], and Caulkin [23] included results of an observer study to assess the realism of masses generated by their respective methods. We look at each study in turn below.

Skiadopoulos et al. described a method for synthesising circumscribed breast lesions [112]. The realism of masses generated by their method was tested in an observer study in which six radiologists attempted to distinguish between 60 real and 60 simulated lesions. The area-under the ROC curve of pooled responses was reported as 0.55±0.03. This implied there was no statistically significant difference between real and simulated lesions. However circumscribed masses do not have spicules, and therefore do not significantly change the appearance of breast tissue in the region surrounding the mass. Thus the part of our method that most commonly identified masses as synthetic is not required when generating circumscribed masses. This makes a direct comparison between the results of our observer study and those reported by Skiadopoulos et al. inappropriate.

Saunders et al. did synthesise mammographic masses, complete with spicules [101]. In an observer study, 25 real and 50 synthetic malignant masses[15] were assessed by three radiologists, and assigned a rating between 0 and 100 (0 corresponding to definitely synthetic, 100 corresponding to definitely real). After fitting ROC curves, an average AUC of 0.65±0.07 was reported. Although it is impossible to make strong statistical inferences given only the summary statistics presented by Saunders et al., we note that the mean AUC computed in our observer study (0.70) lies within the error bounds on the reported mean AUC of Saunders et al.'s study. We can also, for example, apply the Wilcoxon signed-rank test to show there is no statistical evidence to suggest our set of AUC values were from a distribution with median greater than 0.65 ($p = 0.13$). This suggests there was no statistical difference in how realistic masses generated from our method appeared compared to masses generated using the method of Saunders et al.

Caulkin evaluated the realism of masses generated by his method using an observer study in which 15 radiologists attempted to distinguish between 25 real and 25 synthetic masses [23]. A mean AUC of 0.69±0.13 was reported. Again we note that the mean AUC computed in our observer study lies within the error bounds of the mean AUC reported by Caulkin, and that a Wilcoxon signed-rank test implies our set of AUC values are drawn from a distribution that has a median not statistically different to 0.69. This suggests there was no statistically significant difference in how realistic masses generated by the two methods appeared.

---

[15] The 50 synthetic malignant masses included 25 masses with synthesised spicules and 25 without. Also included in the study were 25 real and 25 synthesised benign lesions

Because we used Caulkin's method for generating synthetic masses as our starting point for our own method, it is disappointing not to have generated masses that appeared more realistic than masses generated by Caulkin's method. Given the feedback received from radiologists, we believe this is because in attempting a more complete synthesis of the interaction between a mass and its surroundings, the method we developed in chapter 8 generated regions of unrealistic texture in a significant number of mass backgrounds. This offset the improvements in modelling a training a set of masses described in chapters 6 and 7, so that the overall realism of masses generated by our method was approximately equal to those generated from Caulkin's method. However, as described in section 8.2, we believe that whilst it is more difficult than modelling individual spicules, modelling the texture and structure of whole mass regions is the only way in which malignant mass appearance can properly be synthesised. Thus we propose continuing with the improvements to the mass background model described in sections 8.6.6 and 8.7.2.

In addition, as discussed in chapter 10, whilst Caulkin proposed a model of mass location, the model was not used in sampling new locations for synthetic masses. Instead, the masses tested in the observer study implemented by Caulkin were manually located in mammograms. We also that the synthetic masses evaluated by Saunders et al. were manually located in normal mammogram regions. As such, of the sets of synthetic malignant masses that have been tested in observer studies, only those generated by our method have been synthesised fully automatically within normal mammograms.

Finally, with regards to the work of Saunders et al. and Caulkin, we note that care must be taken in making direct comparisons between the results of observer studies. This is

because the conditions of each study may differ. Factors such as the spatial resolution at which masses are synthesised and the quality and resolution of the screen on which masses are displayed, may affect how easy it is to identify synthetic masses in the study. For example, the resolution of mammograms in which we generated synthetic masses was 40μm per pixel, as compared to 50μm used in Saunders et al.'s work and 100μm in the work presented by Caulkin. Other factors, such as the rating system used between the studies, and whether additional synthetic lesions were tested simultaneously, may also cause differences in the results obtained for malignant masses.

For these reasons, and given the similarity of summary statistics reported, we do not believe it is appropriate to draw definite conclusions on which of the three methods generated more realistic masses. Analysing the results of each study individually however, implies that neither the methods presented by Caulkin and Saunders et al., nor our method generated synthetic masses that were indistinguishable from real masses to mammography experts.

11.7 Summary

In this chapter we have evaluated our method for generating synthetic malignant mammographic masses by assessing how realistic such masses appear to expert mammography readers.

We described an observer study in which 10 mammography readers at the Nightingale Breast Centre, Manchester attempted to distinguish between 30 real masses randomly selected from our training data, and 30 synthetic masses generated using our method for synthesising the appearance of malignant masses in normal mammograms.

The readers assigned each mass in the study a rating on a 5-point scale, varying between definitely real and definitely synthetic. This allowed a quantitative analysis of the extent to which readers were able to identify synthetic masses. This analysis gave strong evidence that the readers could identify synthetic masses at a rate significantly better than chance. Consequently we conclude our method does not consistently generate masses that are indistinguishable from real masses.

In addition to the 5-point rating, radiologists provided feedback on masses they identified as definitely or probably synthetic. An analysis of this feedback suggested that failures in the method for synthesising mass background appearance were by far the largest factor in betraying masses as synthetic. Improving the method by which mass background appearance is modelled and synthesised is the principal area in which we intend to continue work.

We compared the results of our study to the results of earlier studies assessing the realism of synthetic malignant masses. The summary statistics computed in our study were similar to those reported by Caulkin [23], and Saunders et al. [101]. Whilst we do not think it is appropriate to make strong statistical inferences between the different studies, this suggests masses generated using our method have a similar level of realism to masses generated using previously published synthesis methods.

Chapter 12 : Future Work and Conclusions

In this chapter we summarise the work presented in this thesis, discuss possible areas for future work and outline the key contributions our work has made to the field of mammographic research.

We first review our project to synthesise malignant mammographic masses. This includes a brief discussion of:

- the clinical motivation for the project and the requirements this imposed on our method for generating synthetic mammographic masses

- the methods we have developed to meet the project goal

- the observer study used to evaluate the extent to which our method met the project goal

We then examine the limitations of our work and describe how the methods we have developed could be extended to overcome these limitations. In addition we look at other areas to which work in this thesis could be applied, outline the contributions made by our work, and lastly, present our final conclusions on the project.

12.1 Summary

The goal of the work described in this thesis was to develop a method for synthesising the mammographic appearance of malignant breast masses. Large sets of mammograms, displaying a wide range of abnormalities associated with breast disease, are essential for a variety of tasks inlcuding training and assessing mammogram readers, testing imaging equipment and validating image processing techniques. In such tasks, the ground truth

356

of any abnormality present in a mammogram must be known. Synthesising abnormalities overcomes the problem of obtaining a sufficient volume of real data with known ground truth.

However, to be clinically useful the synthetic abnormalities must be indistinguishable from real examples to experts in the field. A review of previous synthesis literature indicated that whilst methods for synthesising microcalcifications and benign breast masses had proved successful, as yet a method for generating the realistic appearance of malignant (and particularly spiculated) masses had not been developed.

The review also indicated that modifying the appearance of real normal mammograms produced more realistic results than attempting to simulate the appearance of whole mammograms. It was observed that due to their relative rates of occurrence, normal mammograms were far more readily available than mammograms containing signs of disease, and so obtaining large sets of normal mammograms in which to synthesise mass appearance would not be a problem.

At this stage, we formalised two requirements of a method for synthesising malignant mammographic masses:

i)   Given a set of normal mammograms, the method should run without user input to generate a set of synthetic masses, thus enabling large sets of data to be generated automatically

ii)   The set of masses generated by the method should be indistinguishable from real mammographic masses, allowing us to substitute synthetic data for real data when training mammography readers

To meet the two requirements, we constructed statistical models to encode the variation

in appearance and location of a set of real malignant mammographic masses

Throughout the project we used a dataset comprising 540 digitised mammograms supplied by the Nightingale Breast Centre, Manchester. The mammograms were grouped in cases each containing the MLO and CC views of the left and right breast of one woman. The cases included 101 malignant masses, the majority of which were visible in both CC and MLO views. To enable further work, the border of each mass was demarcated by a breast radiologist. A rectangular region was extracted about each mass in one mammogram projection. The 101 mass regions were then used as training data from which to build models of mass appearance.

To model mass appearance, in each region we needed to isolate the contribution a mass had made to grey levels in the image. We used a thin-plate spline interpolant to estimate what the grey level of background tissue underlying the mass would have been, had the mass not been present. The background estimates were modified to include high-frequency structures in the region, as it was assumed these were not part of the central mass. As a result of applying the separation method, we effectively generated two new sets of image data: a set of *masses*, and a set of *mass backgrounds*.

We then worked with the set of masses to construct a statistical appearance model. The model encapsulated how masses in the training set varied in size, shape and texture. By fitting a distribution to the model parameterisation of masses, we were able to describe probabilistically how a global set of masses could vary in appearance. Thus sampling from this model generated synthetic instances of masses that matched the characteristics of real masses.

Having built a model of mass appearance, we focused on the set of mass backgrounds. By their construction, each background contained any spicules belonging to the mass that had been separated, in addition to any existing structures in the breast tissue. In many of the backgrounds the presence of a mass had distorted the surrounding breast tissue. Our aim was to model this distortion so that given a normal region we could modify the region in a way that not only generated mass spicules, but also accounted for how structures in the region would be altered by the presence of a mass. To work with the complex textures and structures present in mammographic tissue, we used the dual-tree complex wavelet transform (DT-CWT) [65] to decompose the regions in our training data. We proposed the following three stage process to synthesise mass background appearance:

1) Compute the DT-CWT of a normal region

2) Modify coefficients in the decomposition to match the properties of DT-CWT coefficients in real mass backgrounds

3) Invert the modified DT-CWT to reconstruct a region in which mass background appearance has been synthesised

We experimented with three methods based on this framework described. Ultimately, we chose to use a method in which DT-CWT coefficients from a real mass background (randomly selected from the data) were directly transferred into the transform of a normal region.

Having developed a method for modelling both masses and mass backgrounds, we showed how the two could be combined. Thus having synthesised mass background

appearance in a normal region, we could sample a mass with matching characteristics. This process was summarised in Algorithm 9.1, and gave us a complete method for generating the appearance of a malignant mass in a mammogram normal region.

The final stage in our method was to automate the process of selecting a location within a target mammogram at which to synthesise a mass. To achieve this we constructed a model that described how masses were spatially located within the breast. To build the model we needed to obtain the breast shape in all the mammograms in our training data. Thus we developed a method for automatically segmenting the breast border and, in MLO mammograms, the pectoral muscle. Given a normal mammogram, we used the segmentation algorithm in conjunction with the mass location model to sample co-ordinates at which to generate a synthetic mass. The sampling procedure took account of the size of mass to be generated, so that unrealistic locations near the skin-air boundary of breasts would not be selected.

Coupling the method for sampling a mass location with the method described for synthesising the appearance of a mass within a normal region gave us a complete method for synthesising a malignant breast mass in any normal mammogram. The set of modelling and synthesis steps that comprise the complete method are summarised in the flow charts depicted in Figure 12.1.

By using statistical models to encode mass appearance and location, we were able to generate sets of synthetic masses automatically in normal mammograms by sampling from the probability distributions encoded in the models. Further, by learning the model distributions from real training data, we aimed to ensure that any set of synthetic masses

generated using the method have the same appearance characteristics as real masses.

To assess whether our method had consistently generated synthetic masses of realistic appearance, we implemented an observer study involving ten expert mammogram readers from the Nightingale Breast Centre, Manchester. During the study a reader was shown a mixed set of 30 real and 30 synthetic masses, and asked to rate each mass on a scale varying from definitely real to definitely synthetic. Analysis of the results showed that the synthetic masses were identified at a rate significantly better than chance and implied that a proportion of synthetic masses generated from our method are not indistinguishable from real masses. In addition to the ratings assigned to each mass, the readers in the study were asked to provide feedback on masses they had identified as synthetic. This feedback suggested that in the majority of masses correctly identified as synthetic, failures within the mass background synthesis algorithm were responsible. This feedback gave further weight to our intentions to improve how mass regions were modelled, as discussed in the next section.

Figure 12.1: Flow chart summarising the process for modelling a set of malignant masses and subsequently using the models to generate synthetic masses in normal mammograms
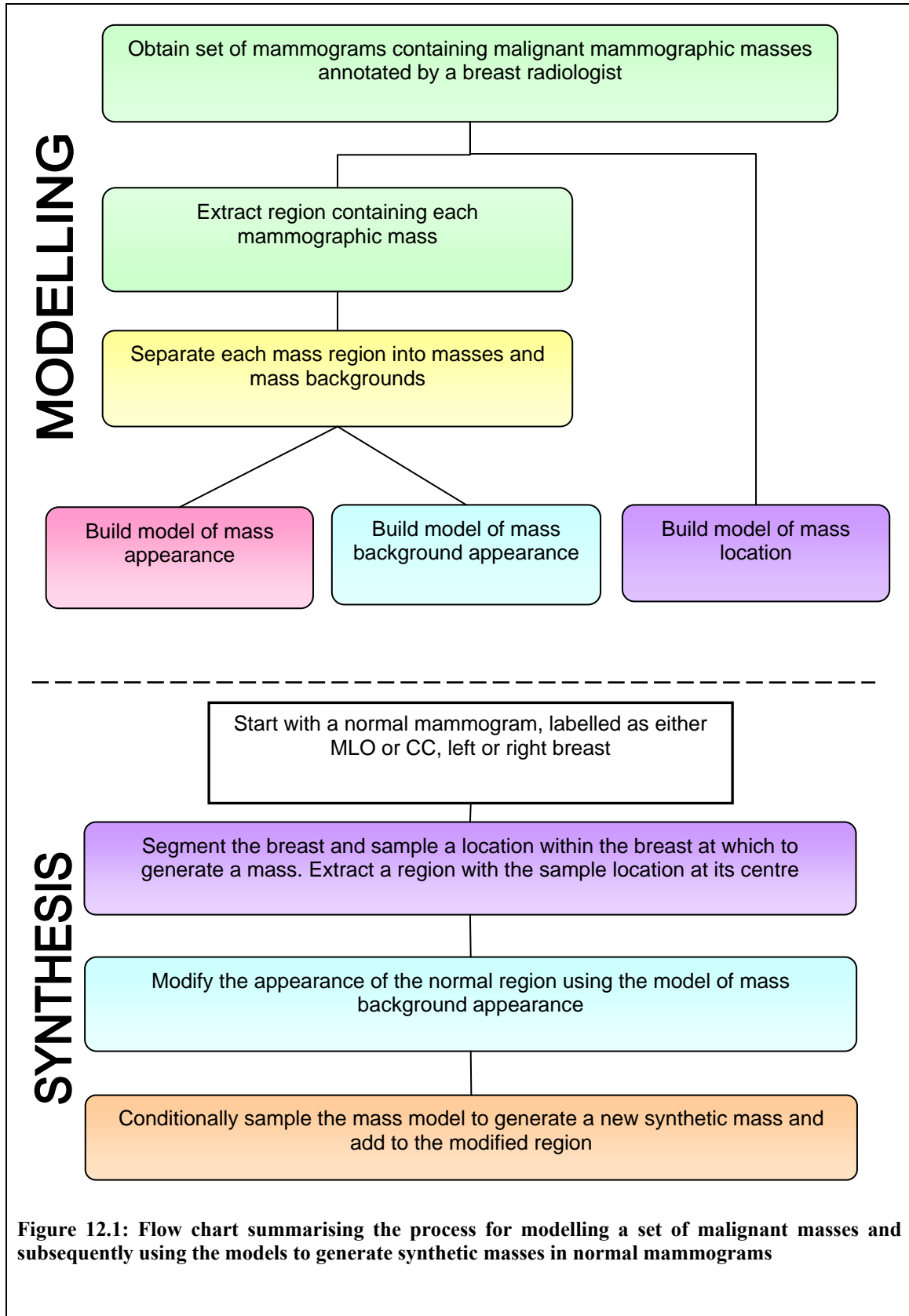
12.2 Further development of our work

In discussing the limitations of work presented in this thesis there are two areas to consider: firstly we look at specific limitations of the method by which we have synthesised the appearance of malignant breast masses in mammograms; secondly we consider the more general limitations imposed by the scope of the project we have presented. In each case, we discuss how the methods we have developed could be extended to overcome the stated limitations.

12.2.1   Improving our synthesis method

Given the results of the observer study outlined above, it is clear we have not yet developed a method for consistently generating synthetic masses that appear sufficiently real to be mistaken for real masses by mammography experts. This limits the extent to which masses generated by our method could be used in software designed to aid reader training in mammography. Thus it is necessary to review which aspects of our method contributed to the unrealistic appearance of some masses.

Synthesising mass background appearance

Given the feedback provided by readers in the observer study, we first consider the way in which mass background appearance was synthesised in normal regions.

As discussed in section 8.2, we developed a method for synthesising the way in which the presence of a mass affected the local appearance of breast tissue in a mammogram that was more ambitious than previous attempts at synthesising spiculated malignant mammographic masses. Whilst previous authors had attempted to synthesise spiculated masses, their focus was confined to modelling individual spicule structures and adding these to a central mass. How such spicules interacted with surrounding structures in the

breast had not been addressed, nor had the way in which any malignant mass (with or without spicules) distorts local breast tissue been considered. Therefore we proposed a synthesis method that, given a normal mammogram and a synthetic mass, would generate spicules in addition to modifying the appearance of existing structures in the breast tissue.

However, we were unable to construct a probabilistic model from which such synthetic appearance could be sampled. Seeking to describe mass backgrounds using only the local correlations of coefficients at varying scales of a wavelet transform was not sufficient to encode the long-range interactions that govern the appearance of mammographic structures. Meanwhile we have not yet identified a suitable global model to parameterise the structures present in a mass region as a fixed length vector. Thus in contrast to masses, we have not been able to develop a statistical model to describe the variation in appearance of mass backgrounds across the set of training data.

As a compromise we used the real mass regions in our training data as templates to synthesise appearance in normal regions, as described in Algorithm 8.1. This has two major limitations: firstly, the DT-CWT coefficients in a mass region were not always compatible with coefficients in the normal region to which they were transferred and thus unrealistic appearance was generated; secondly, it reduces the generality with which mass region appearance can be synthesised.

From the feedback provided by mammogram readers it would appear the first limitation was the single biggest cause of masses being identified as synthetic in the observer

study. In particular, our method struggled to produce realistic spicules, both in terms of matching spicule appearance to surrounding breast tissue and matching the spatial distribution of spicules to the appearance properties of the synthetic mass (despite the conditional sampling algorithm proposed in chapter 9). As a result, where spicules were generated, they did not always appear to be growing from the mass as intended.

Moreover the second limitation (coupled with the restrictions this placed on sampling a mass to match a synthesised background discussed in chapter 9), reduces the likelihood of our method being able to generate the full range of appearance we might expect to see in a global set of real masses.

To overcome these limitations we proposed a hybrid model of mass background appearance combining a global map of structure with a set of local texture models (see section 8.7.1). The global model controlled the content and spatial arrangement of structures in a mass region. Thus sampling from the global model given a normal region would control the generation of spicules and the spatial interaction between a mass, its spicules and existing structures in the region. The local model would then fill in the textural details given the map of structure to generate realistic mammographic appearance.

The proposed hybrid model was not used during the generation of the synthetic masses evaluated in this thesis, because both the global model of structure and local models of texture require further development. These issues were discussed in detail in sections 8.6.6 and 8.7.2. Our intention is to continue working on this model as we believe it is crucial in obtaining a proper understanding of malignant mass appearance.

Synthesising mass appearance

Given the data available, we are satisfied that the appearance model we constructed in chapter 7 provided a good description of the set of separated masses. However, as with any model based on learning variation in real data, the more training examples we have the more confidence we have in the ability of our model to provide a general description of data in a global population. Because expert annotation of masses in the original mammogram data was one of the most time consuming aspects of the modelling process, obtaining larger datasets for further work may benefit from an automatic segmentation of masses. Automatically segmenting masses would also have the advantage removing the only element of the modelling and synthesis methods we have developed that requires manual intervention.

We also observe that with more training examples we would expect to populate the feature space over mass appearance more densely. This would provide us with better information with which to assess the validity of our assumption of a Gaussian distribution of the feature space. Given the data we had available it was difficult to judge if there were non-linear correlations in the mass appearance parameters. However, if such non-linearities were present (and therefore not captured by the Gaussian distribution we assumed over the space) we do not believe these significantly contributed to unrealistic appearance in masses synthesised from the model.

Modelling mass location

As with the model of mass appearance, having more data would benefit the probability distribution learned in the model of mass location.

We also note the limitation imposed by not having a method for automatically detecting the nipple in mammograms. Without the nipple as a landmark, we may have reduced the physical correspondence between each breast shape when computing the mean breast shape in both CC and MLO mammograms.

As discussed in section 10.2.4, previously used methods for locating the nipple [81], proved unsuccessful in mammogram where the nipple was not obviously in profile. In these mammograms, a radiologist would make a subjective decision on the location of the nipple given the arrangement of parenchymal tissue in the breast. We have not attempted to mimic such a decision in an automated algorithm. However in developing a sufficient description of tissue structure with which to model mass region appearance, we may be able to use spatial arrangement of tissue to compute an approximate location for the nipple in any mammogram. The nipple could then be used as a marker in future mass location models.

As a final point with regards to locating masses, we note that our method considers only the location of masses in the training data relative to breast shape. We do not attempt to model the location of masses relative to internal structures in the breast (for example, the fibroglandular disc). This was because such structures are not a consistent feature of all mammograms. There may be an argument however, that where a breast contains a well defined glandular region, the location of masses within the breast should be considered relative to the glandular structures.

However, despite the limitations outlined above, the feedback from readers during the observer study (see section 11.5.2) indicated that the locations selected for synthetic

masses were not a factor in identifying them as unrealistic.

12.2.2   Extending the scope of our synthesis method

In this section we consider ways in which our method for generating synthetic mammographic masses could be extended beyond the scope of the project set out in section 3.2.2. In particular, we consider synthesising mammographic abnormalities other than malignant masses; synthesising masses in four-view mammogram cases (as opposed to single mammograms); and synthesising breast masses in 3-D image modalities. Each topic is discussed in turn below.

Synthesising other abnormalities

So far we have only considered a method for synthesising the appearance of malignant breast masses in mammograms. We have not considered synthesising other mammographic signs of breast cancer or any benign abnormalities present in the breast. As discussed in section 2.2.4, there are numerous other abnormalities that may be present in mammograms – including both signs of cancer and benign findings. We focused our attention on malignant masses because these are one of the hardest abnormalities to convincingly simulate and one the most frequent mammographic signs of breast cancer.

However, in learning to read mammograms, trainees must be familiar with the appearance of all abnormalities. In addition, obtaining real sets of mammograms containing other abnormalities may be as hard as obtaining sets of malignant masses[16],

---

[16] Indeed for less common signs of disease such as diffuse asymmetry it may be even harder to obtain sufficient training sets

and therefore there is an analogous motivation for synthesising the appearance of other mammographic abnormalities.

As observed in section 3.2, methods for generating realistic circumscribed masses and microcalcifications have already been presented in the literature [70, 112]. Thus we believe it would be most interesting to extend our work by attempting to synthesise breast asymmetry and architectural distortion. To the best of our knowledge, no methods for synthesising these abnormalities have been reported in the literature. Conceptually, the problem of synthesising architectural distortion is similar to the one that we addressed in chapter 9, where we tried to model how a mass distorted structures in breast tissue in regions from which we had removed the central mass.

Moreover, unlike mass lesions or microcalcifications, there are no existing methods for consistently detecting asymmetries or architectural distortions in mammograms. Thus in addition to synthesis, a model that could be used to detect architectural distortion or breast asymmetry would be of great clinical value. This is discussed further in section 12.3.

Synthesising masses in four-view mammogram sets

In synthesising the appearance of malignant masses, we have considered the appearance of a single mass within a single mammogram. However during screening, mammograms are usually read in sets comprising the CC and MLO views of both breasts. Therefore we may want to:

i) Consider generating a mass in both CC and MLO views of a breast such that the appearance and location of the mass between the two views is consistent with

real data

ii) Consider generating multiple masses within the same breast

iii) Consider generating masses in both left and right breasts in a single set

To generate the appearance of a mass in both CC and MLO views we need to how the appearance of a mass in one view constrains the appearance of the mass in the other. In terms of constructing statistical models of appearance, this requires learning the joint probability of the vectors that parameterise a mass in the two views. Similarly we would need to learn a joint probability distribution for the location of masses in both CC and MLO views. By learning the joint distributions of both appearance and location, we are doubling the dimensions considered in the respective models. As a result we would require a larger training set of masses.

To synthesise multiple masses in a single breast we need learn whether there are any correlations in appearance and location between masses in the same breast in real data. Because we have only 10 cases in our data where there is more than one mass in a single breast, we do not yet have enough data to judge whether such correlations exist. Likewise, to synthesise one or more masses in both breasts of four-view mammogram case we would need to establish whether there are correlations in appearance between masses in different breasts of the same women. Again we note that we do not have enough data with which to make this judgement. However, given training data in which we did have sufficient cases with multiple masses in one or both breasts, then modelling any correlations in the data should be a simple extension of the methods used to model joint appearance in both CC and MLO views.

3-D Image Modalities

We conclude this section by considering the appearance of breast masses in imaging modalities other than mammograms. In section 2.2.1, we observed that x-ray mammography is still the only widely used modality for asymptomatic screening for breast cancer. Therefore the vast majority images used in the detection of breast cancer are mammograms.

However, in recent years there have been developments in 3-D image modalities such as MRI and digital breast tomosynthesis (DBT) to the extent that they may be used in place of mammography to detect breast cancer (particularly when imaging women whose breasts are too densely composed to obtain meaningful information in a mammogram). Therefore it may be clinically useful to attempt to synthesise signs of disease in 3-D modalities.

Moreover, having a method for synthesising disease within a 3-D image of the breast has useful implications for synthesising disease concurrently in both CC and MLO views. For example, given a mass simulated within a 3-D breast shape, by considering the projection of the breast into the two mammographic views, we can constrain the location and main appearance characteristics of the mass in both mammograms. We would then need only to consider synthesising the high-frequency details of the mammographic masses to account for the difference in spatial resolution between the 2-D and 3-D images.

We recall that in section 3.2, we reviewed methods in which a 3-D model was used to synthesise abnormalities in 2-D mammograms [17, 84]. We rejected these methods because the 3-D models generated were not learnt from real data. However, with a

suitable set of training data in a 3-D modality, then the models of appearance developed in this thesis could be extended in to 3-D. For example, as noted in section 4.2.3, similar statistical models have already been used in the Active Appearance Model framework to segment structures in 3-D medical image data [12, 83].

Finally we note that recent research by Hipwell et al. [57] has attempted to model the projected appearance of a mammogram given an MRI volume of a breast. Combined with a model of breast masses in 3-D, this could help create a complete set of methods to synthesise disease in both MR images and mammograms.

12.3 Detecting abnormal distortions in mammograms

Commercial computer-aided detection systems focus on detecting malignant masses and microcalcifications. They have proved successful at this task. However, it is less clear if such systems can detect abnormalities such breast asymmetry and architectural distortion in which no focal mass is present [13]. A study by Burrell reported that approximately a third of cancers that are missed during screening initially appeared in mammograms as undetected distortions [20].

From a research perspective, we reviewed methods by Karssemeijer [63], Parr et al. [88], and Rangayyan et al. [96] that attempted to detect and classify patterns of abnormal structure in mammograms (see section 3.3.1). Both Karssemeijer [63] and Parr et al. [88] reported a detection sensitivity of approximately 80% whilst generating, on average, less than 1 false positive per image. However both these studies concentrated on patterns of structure associated with spiculated masses. The more recent study by Rangayyan et al. [96] attempted to detect more general signs of architectural distortion, including examples where no focal mass was present. However, to obtain a

372

sensitivity of 84%, 7-8 false positives were generated per image. This suggests there is a need for developing a sensitive yet specific method to detect architectural distortion.

In this thesis, the mass background images we have generated have a close relationship to regions of distorted tissue in which no focal mass is present. Although we have focused on synthesis, the models we are attempting to build of mass background appearance may be equally applicable to analysing regions of architectural distortion.

Moreover, not only do we think classifying and detecting abnormal distortion in breast tissue is possible, it may be an easier task than the synthesis problem we have been trying to solve. To produce photo-realistic mass region appearance by reconstructing synthesised DT-CWT coefficients, we need to recreate all the coefficients that contribute to appearance in a region. As we saw in section 8.6.6, attempting to model all the coefficients within a single model proved unsuccessful. However, for detection we need only learn variation in the subset of coefficients that discriminate between classes of region appearance. If we can learn this subset of coefficients, then we can build a more compact representation of local texture that is easier to model.

In constructing a compact representation of texture, we aim to produce a richer labelling of local structure than a binary split between linear structures and background. Specifically, we aim to identify structure that is salient when assessing signs of abnormal distortion in breast tissue. We believe this is possible because of the structural information provided by DT-CWT coefficients (and particularly measures of local phase and orientation).

If this was achieved, we could then build models of orientation patterns as, for example,

in the work by Parr et al., but using only salient structures. With the addition of a suitable search algorithm, such a model could be used to detect regions of architectural distortion in mammograms. We believe that such a scheme would have improved specificity compared to the methods that consider all types of linear structure when classifying patterns of orientation.

12.4 Key contributions of our work

Having summarised our project on synthesising malignant breast masses in mammograms, and considered the ways in which our work could be improved and extended, we now detail the contributions we believe our work has made.

12.4.1   Lesion synthesis

For any work within the field of mammographic imaging research, the ultimate goal is to improve the way in which breast cancer is detected and/or diagnosed. As discussed in sections 2.2.7 and 2.2.8, we identified the task of synthesising malignant breast masses within mammogram as having significant benefit towards many other projects that have a direct impact on this overall goal.

Following a review of previous work, we concluded there had yet to be a method for synthesising malignant breast masses in mammograms that met the criteria we demanded.

In summary, previous methods:

-   controlled synthetic appearance using parameters subjectively adjusted by human observers without attempting to learn the appearance of real masses [17, 54, 99, 112]

- lacked a probabilistic interpretation of appearance parameters thus preventing unique instantiations of synthetic masses being generated [17, 54, 99, 112]

- failed to account for the interaction between a mass and surrounding background tissue [17, 23, 54, 99, 101, 102, 112]

- lacked a method for automatically locating a synthetic mass and thus required manual intervention to position each mass in a mammogram [23, 54, 99, 101, 102, 112]

Therefore the principal contribution of work included in this thesis is the development of a method to generate malignant mammographic masses that overcomes all the limitations outlined above.

Of the previous methods reviewed, we identified one, described by Caulkin [23], as having the most potential as a basic framework from which to synthesise masses. However, having applied the method described by Caulkin to our set of training data, we identified several areas in which significant development was required. These developments are described below.

- The method for separating each mass in the training data from the local background tissue was improved in two significant ways: firstly, by re-ordering the steps in which the separation was computed, we ensured high-frequency structures belonging in the local breast tissue were not subtracted along with the mass (see section 6.4). Secondly, we employed the method for estimating background tissue intensities in an iterative algorithm to negate the effect mass intensities may have had on the initial estimation. We also conducted a new experiment to select the best set of parameters

375

to control the separation algorithm. As a result of these changes we significantly improved how mass texture was modelled, and provided a set of image data in which the interaction between mass spicules and local breast tissue could be analysed.

- To reduce the rotational variation between similar mass shapes, we applied the MDL optimisation algorithm [34] to our data prior to building a model of mass shape. As described in section 7.4.3, this produced a significant improvement in the ability of our shape model to generalise to unseen mass shapes.

- To improve how the individual models of shape, texture and scale were combined, we optimised the weighting applied to each model parameter (section 7.6.4). Using the method for weighting parameters described by Caulkin on our data, shape and texture were suppressed to the extent that variation in the combined model was comprised almost entirely of changes in mass size. As discussed in section 7.6.1, we believe the weighting method originally applied by Caulkin was theoretically flawed and by chance happened to produce a satisfactory weighting given the relative metrics of shape and texture in Caulkin's data. Given a different set of data with metrics on a different scale (for example, the mammograms we used had double the spatial resolution of those used in Caulkin's work, and thus distances in pixels are a factor of two larger in our data), the weighting proved unsuitable. The optimisation method we proposed in section 7.6.3 can be applied to any dataset of masses (regardless of a change in metric) to compute a suitable sets of weights. Thus the method we have described can be used in any future models.

In addition to the changes outlined above, we have developed a method for automatically selecting target locations at which to sample synthetic masses within

376

normal mammograms. A method for constructing a statistical model of mass location in MLO mammograms had been proposed by Caulkin [23], however the steps required to make such a model viable for sampling new locations had not been developed. Caulkin's model of mass location relied on manual segmentations of the breast a pectoral muscle; was designed only for MLO mammograms; and allowed unrealistic locations in the immediate vicinity of the skin-air boundary in a mammogram to be selected. To overcome these limitations we:

- Developed an automatic breast segmentation algorithm to extract the smooth skin-air boundary from each mammogram in our training data as described in section 10.2

- Implemented Karssemeijer's algorithm [62] to fit a straight line to the pectoral muscle edge in each MLO mammogram. This, coupled with the segmentation of the skin-air boundary allowed us to automatically extract the breast shape from each MLO mammogram in the data, and thus allowed us to implement to model of location proposed by Caulkin

- Defined a set of correspondences to describe breast shape in CC mammograms, and thus developed an equivalent location model for CC mammograms on automatic segmentations of the breast as described in section 10.3.2

- Developed a sampling method in which the model distributions were masked at the skin-air boundary of the mean breast shapes, with the radius of the mask dependent on the size of synthetic mass to be generated, as described in section 10.3.3. This prevented masses from being unrealistically located at or on the skin-air boundary in target normal mammograms, and had the additional effect of recording a correlation

between mass appearance and mass location.

By developing the steps outlined above we were able to automate a complete method for generating a synthetic mass in a normal mammogram. Thus of the methods to synthesise malignant masses that have been quantitatively evaluated [23, 101, 102], ours is the only one that generates masses without requiring user input. As stated previously in section 2.2.7, we believe this is crucial in harnessing the full potential of using synthetic data.

The most novel contribution of our mass method for generating synthetic masses is the attempt we made to model the interaction between a mass and the surrounding breast tissue. As acknowledged in previous attempts at synthesising malignant breast masses [23, 101, 102], this is an issue that has not previously been addressed. Indeed to the best of our knowledge our method is the first in the synthesis literature to properly consider the problem. We believe this problem must be considered for the full characteristics of real malignant masses to be recreated. Therefore whilst the solution we adopted to synthesise the distortion of breast tissue in mass regions produced unrealistic artefacts in some regions, we believe the way in which we framed the problem provides a starting point from which future methods can embark. Further, in section 8.7, we described a possible solution to modelling the distortion caused to breast tissue by the presence of a mass and showed promising resulting in the form of some initial synthesised textures.

Therefore despite not achieving our target of generating malignant breast masses that are all indistinguishable from real masses, we believe the work presented in this thesis is to date the most comprehensive treatment of synthesising disease signs in mammograms

and makes a significant contribution to the literature in this area of research.

We also note that within our complete method for generating synthetic masses, separate modules can be applied individually in other synthesis methods. For example, the method we have described to construct a model of mass location requires only a set of training mammograms in which lesion location is known, whilst sampling a new location requires only knowledge of the size of synthetic mass being generated. Thus this method could be used as part of any algorithm to simulate lesions in whole mammograms. A reliable method for locating synthetic lesions in whole mammograms has yet to be published, and so this work may be of direct benefit to researchers currently working with synthetic lesions.

12.4.2   Modelling mammographic tissue

With regard to the wider field of image analysis in mammography, we note that there have been very few attempts at modelling the appearance of general distortions of breast tissue in mammograms that may be indicative of disease. Whilst computer-aided detection algorithms have been used to detect mammographic masses in a clinical setting for several years [74], such systems focus on detecting masses and microcalcifications. However, it is estimated that approximately a third of cancers that are missed during screening initially appeared in mammograms as undetected distortions [20], highlighting the clinical importance of detecting such abnormalities.

In this thesis we have concentrated on how breast tissue is affected by the presence of a central mass. However, as discussed in section 12.3, work aimed at modelling this form of tissue distortion in mammograms is equally applicable towards distortions of breast tissue in which a mass is not present or not visible. Therefore, although the

379

experimental work in chapter 8 was originally developed for synthesis, it can be seen as initial work towards an analysis problem that has a much more direct benefit to detecting breast cancer in a clinical setting.

From a technical point of view, modelling and synthesising the appearance of mammographic tissue is extremely challenging. The complex projection of structures in a breast forms an image that cannot adequately be described by local texture models. However, the structures in any given region do not form a consistent set of correspondences and are often ill-defined; globally constraining appearance is also problematic. To obtain a richer description of mammographic texture than that provided by image grey levels alone, we decomposed mammogram regions using the dual-tree wavelet transform (DT-CWT) [65].

As described in sections 4.3.3 and 8.3.2, the properties of the DT-CWT make it highly applicable to analysing texture and structure in mammograms and we believe it will be a powerful tool in further research. However, as yet there is no published work in which the dual-tree complex wavelet (DT-CWT) has been applied to mammograms. Therefore the work described throughout chapter 8 comprises a novel contribution to mammography research. As discussed in section 8.6.6, we acknowledge that we have not yet made full use of the properties offered by the DT-CWT (for example, we have not used its ability to create rotation invariant feature vectors or its ability to model intrinsically 2-D structures such as junctions in addition to 1-D structures such as lines and edges). However, because of the lack of previous research in this area, our work forms the first steps towards realising the potential of the DT-CWT as a tool in mammography research.

More generally, the methods for synthesising mass region appearance presented in sections 8.6 and 8.7 are novel texture synthesis algorithms. In particular, if the algorithm combining a global structure model with local texture models described in section 8.7 can be refined, this could be applied to any image class with similar properties to mammograms. Such image classes cannot be adequately synthesised by the algorithms currently published in the literature and so our algorithm is of benefit to the general field of texture synthesis.

12.4.3   Additional contributions

In addition to the main contributions we have outlined above, there are several smaller contributions that have been made during our project. These are listed below:

- We have developed standalone software with which radiologists can annotate mammographic masses. As described in section 5.2 and appendix B, we have designed and implemented a software interface for annotating mammograms. The software allowed the user to highlight regions of a mammogram, delineate the boundaries of any mass in the region and annotated spicules belonging to the mass.

- We have developed standalone software to run observer studies in which participants visually assess mammogram regions. As described in section 11.4.2, we designed and implemented the software used by mammography readers an observer study. The software provided instructions for the users and allowed them to complete the study unaided. It allowed users to assign ratings for each mass in the study and provide feedback on masses identified as synthetic. The basic framework of the software can easily be adapted to perform similar observer studies and thus may be beneficial testing other forms of synthetic data.

- We have developed a segmentation algorithm to delineate the border of breasts in both CC and MLO mammograms. As discussed in section 3.3.2, breast segmentation algorithms have previously been published. However the algorithm we developed was designed to meet the specific requirements of the task in which it was used (to obtain breast shape for mass location modelling). Within our algorithm, we note that our formulation of the traditional active contour algorithm, in which image edge energy was maximised in a separate co-ordinate frame to that in which contour energy was minimised could be an efficient way of refining the breast border returned by existing segmentation algorithms

- In sections 8.4.2 and 8.4.3, we described two further transforms of the DT-CWT: the inter-level product (ILP) [7] and the inter-coefficient product (ICP) [8]. These transforms were developed by Anderson et al. and provide a valuable way of extracting structural information in an image from the phase of DT-CWT coefficients. We suggested technical changes to how both the ICP and ILP should be computed. Whilst these modifications do not alter the concept of the transforms, they allow the coefficients computed by the transforms to vary more consistently with rotation. We suggest that any researcher applying both the ICP and ILP as a tool in image analysis considers the changes we recommend.

12.5 Final conclusions

In this thesis we have presented a method for generating malignant breast masses in normal mammograms. Our method is the first that allows masses to be synthesised directly in a whole mammogram without requiring a user to manually adjust any appearance parameters. Given the relative ease with which whole normal mammograms

can be obtained, this allows large sets of malignant masses to be generated automatically, thus realising the full potential of using synthetic data.

Whilst a proportion of the masses we have generated can be identified as synthetic by mammography experts, we have identified ways in which our synthesis method can be improved to overcome this limitation. If these changes are successfully implemented we envisage a system in which synthetic masses may be used in place of real masses in a variety of applications. In particular, we propose that software in which a potentially infinite set of synthetic masses could be generated with which to train readers in mammography would be of great benefit.

As part of our synthesis method we have attempted to model the way in which breast tissue is distorted by the presence of a mass. This is a subject of great clinical importance, but one that as yet has little published research. The work in this thesis includes initial experiments to fit statistical models describing distortions in breast tissue and we intend to continue work in this exciting area of mammography research.

Appendix A: Computational Techniques

A.1 Principal component analysis

Principal Component Analysis (PCA) is a well established and commonly used statistical technique that can be used to parameterise high dimensional data in a compact way. PCA is fundamental to the application of the appearance models described in section 4.2 and applied throughout this thesis. A brief review of the technique is given below.

Consider we have data consisting of N examples, each of which is an n-dimensional vector. That is,

$$X = \{\mathbf{x}_i \mid i = 1, ..., N\} \qquad \mathbf{x}_i = (x_{i1}, ..., x_{in})^T$$

First we calculate the covariance matrix C of the sample

$$C = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$$

We then perform an eigenvalue decomposition $C = PDP^T$, such that $P = (\phi_1 ... \phi_n)$ is an $n \times n$ matrix of the eigenvectors of C and D is a diagonal matrix of the associated eigenvalues λ_j.

If $N < n$ there are only $m = N - 1$ eigenvectors with non-zero associated eigenvalues, and so we define P_X as the $n \times m$ matrix consisting of the eigenvectors $\phi_1, ..., \phi_m$ with associated non-zero eigenvalues $\lambda_1, ... \lambda_m$, sorted such that $\lambda_j > \lambda_{j+1}$

If we now define the m-dimensional vector $b_{\mathbf{x}_i}$, as $b_{\mathbf{x}_i} = P_X^T (\mathbf{x}_i - \overline{\mathbf{x}})$ then we can write

each of our sample points as

$$\mathbf{x}_i = \overline{\mathbf{x}} + P b_{\mathbf{x}_i}$$

That is, each data point is described by its displacement from the mean along a set of orthogonal principal components (modes) defined by the eigenvectors. The variance of the data along each of the modes is given by the associated eigenvalues, and the total variance by the sum of the eigenvalues. Since many of the eigenvalues will be very small, we can discard the associated modes by using only the first t rather than all m eigenvectors. That is, we define $P_X' = (\phi_1 ... \phi_t)$ and calculate $b'_{\mathbf{x}_i} = P_X'^T (\mathbf{x}_i - \overline{\mathbf{x}})$, so that each data point is approximated as:

$$\mathbf{x}_i \approx \overline{\mathbf{x}} + P'_X b'_{\mathbf{x}_i}$$

Thus our data is now represented in the smaller t-dimensional space as opposed to the original n-dimensional space. The total variance encapsulated in this model is $\sum_{i=1}^{t} \lambda_i$, and so the value of t can be chosen to ensure an arbitrary level of accuracy is maintained.

A.2: Thin-plate spline interpolation

The thin-plate spline is a commonly used tool for interpolating surfaces over scattered data. We assume we have set of data $\{(x_i, y_i, z_i) \ i = 1,...,n\}$ where (x_i, y_i) are coordinates on an infinite thin plate of metal, and z_i is the vertical displacement of the plate at that point. The thin-plate spline returns a function $f(x, y)$ that defines the height of the plate for any (x, y) pair such that the overall bending energy of the plate is minimised.

Formally, we aim to find the function $f(x, y)$ that minimises

$$I_f = \iint \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dx dy$$

The solution can be expressed as a superposition of fundamental solutions of the biharmonic equation $\Delta^2 U = 0$ given by $U(x, y) = (x^2 + y^2) \log(x^2 + y^2)$

First we define the following matrices:

$$K = \begin{pmatrix} 0 & U(r_{12}) & \dots & U(r_{1n}) \\ U(r_{21}) & 0 & \dots & U(r_{2n}) \\ \dots & \dots & \dots & \dots \\ U(r_{n1}) & U(r_{n2}) & \dots & 0 \end{pmatrix} \qquad r_{ij} = (x_i - x_j, y_i - y_j)$$

$$P = \begin{pmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \dots & \dots & \dots \\ 1 & x_n & y_n \end{pmatrix} \qquad L = \begin{pmatrix} K & P \\ P^T & O \end{pmatrix}$$

and the vector of known displacements

$$V_z = \begin{pmatrix} z_1 & z_2 & \dots & z_n & 0 & 0 & 0 \end{pmatrix}^T$$

From these we can calculate a set of weighting coefficients

$$W_z = L^{-1} V_z = \begin{pmatrix} w_1 & w_2 & \dots & w_n & a_1 & a_x & a_y \end{pmatrix}^T$$

We can now explicitly write

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U(|(x_i, y_i) - (x, y)|)$$

The method as shown above leads itself naturally to interpolating grey levels in a 2-D by thinking of the image as a surface with vertical displacements defined by the known grey levels of a set of source pixels. The method can also be extended to be used a warping tool, performing non-linear deformations as is described in the following section.

A.2.1   Warping using thin-plate spline interpolation

Thin-plate spline interpolation can be used to warp one image (the source) to another (the target), based on a set of landmark points in the two images.

Let $\{(x_i, y_i)\ i = 1, ..., n\}$ be the coordinates of the landmark points in the source image. Similarly, let $\{(x_i', y_i')\ i = 1, ..., n\}$ be the coordinates of the landmark points in the target image.

To compute a function that maps every point in the source image to a point in the target image we calculate two thin-plate splines, defining the displacement of the source points to the target points in the x and y directions.

Thus using the method described in the previous section, we calculate the first spline using input data

$$\{(x_i, y_i, z_i = x_i' - x_i)\ i = 1, ..., n\}$$

to compute an interpolating function $f_x(x, y)$. Similarly, we compute $f_y(x, y)$ using input data

$$\{(x_i, y_i, z_i = y_i' - y_i)\ i = 1, ..., n\}$$

The resulting (f_x, f_y) defines a non-linear mapping of any point in the source image to the target image

A.3 Chord length curvature

In this section we describe a measure of curvature for a 2-D curve. We assume the curve has been discretised to be defined by control points $\{(x_i, y_i)\}$ such that the points are spread equidistantly along the curve.

At each point (x_i, y_i), we consider the triangle formed by the triplet $\{(x_{i-1}, y_{i-1}), (x_i, y_i), (x_{i+1}, y_{i+1})\}$. If we label this triangle ABC such that $B = (x_i, y_i)$, and AC forms the base of the triangle, then the we compute the curvature c_i at (x_i, y_i) as the square of height of the triangle.

Appendix B: Annotation Software

We created software to allow radiologists to annotate the salient features of malignant spiculated masses in mammograms. The software was designed using the graphical user interface tools in Matlab®. Its key features are described below:

− The user can inspect a whole mammogram, and select a region of interest containing an abnormality. The user can zoom and pan within the full mammogram and mark the location of the nipple

− Having selected a region of interest, the region is displayed simultaneously in three windows using: the original grey-scale colour map; a contrast enhanced grey-scale colour map; and a contrast enhanced colour representation. A further option of displaying the region with lines enhanced using the linear operator [36] was developed, but after consultation with radiologists at the Nightingale Breast Centre, Manchester, this was deemed unnecessary. The user can enlarge the region of interest windows to any size, and can zoom and pan within the regions.

− Having selected a region of interest, the user can outline the border of a mass and/or mark any spicules in the region. Annotation may be performed in any of the three region of interest windows and appears simultaneously in all three. Any annotation can be saved, reloaded and modified at a later date.

A screen shot of the software is depicted in Figure B.1. The figure displays: the tools panel used to operate the software; the window in which the full mammogram is displayed - the nipple has been marked (red star) and region of interest containing a mass selected (blue rectangle); the enhanced grey-scale and colour representation of the

region of interest in which the mass border has been annotated.
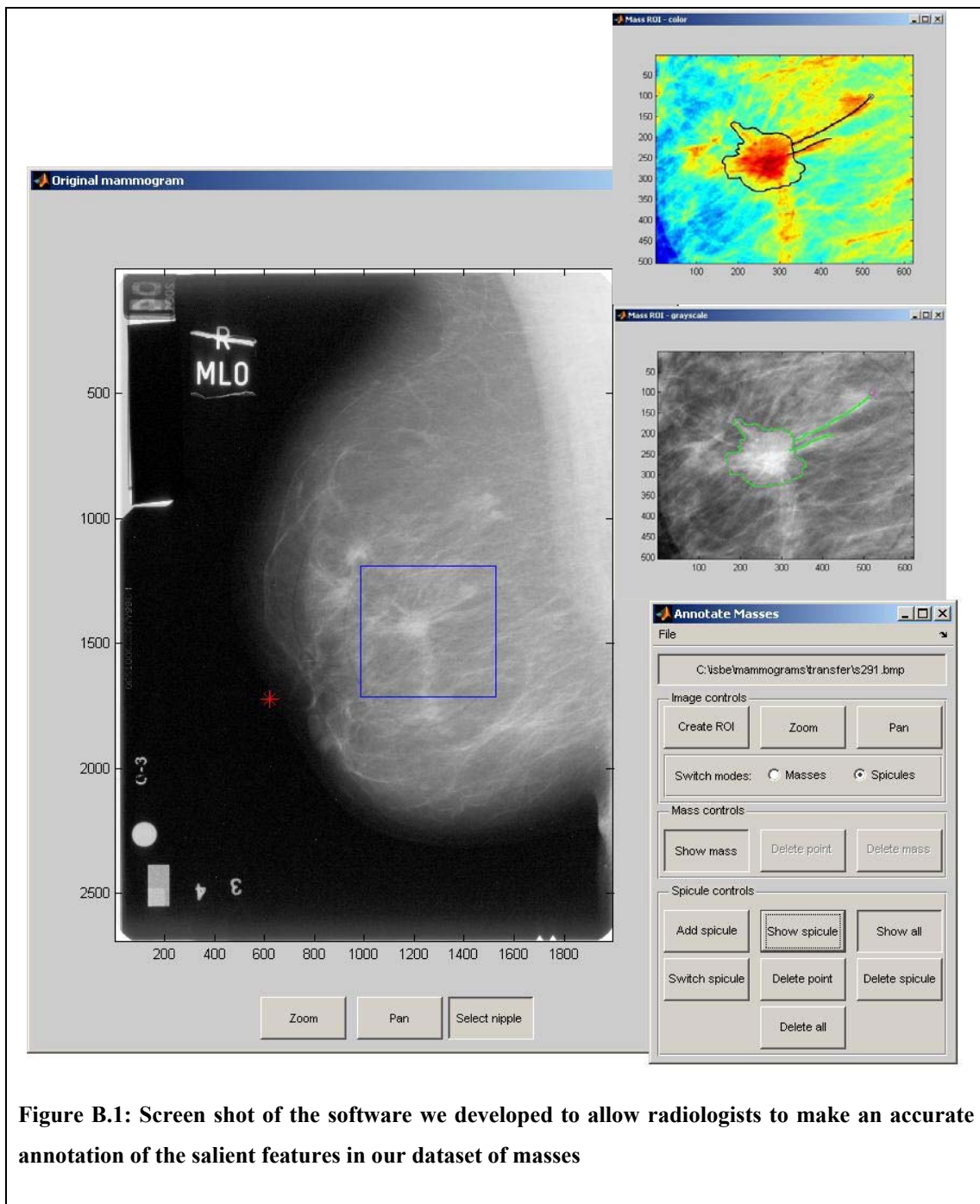
Figure B.1: Screen shot of the software we developed to allow radiologists to make an accurate annotation of the salient features in our dataset of masses

Bibliography

1. Breast cancer incidence and mortality figures for England and Wales, Office for National Statistics. 2005

2. Screening for Breast Cancer in England: Past and Future, in *NHSCSP Publication No 61*. 2006

3. Cancer Reform Strategy, Department of Health, 2007

4. An Audit of Screen Detected Breast Cancers for the Year of Screening April 2007 to March 2008, NHSBSP & Association of Breast Surgery, UK. 2009

5. Althuis, M.D., Dozier, J.M., Anderson, W.F., et al., Global trends in breast cancer incidence and mortality 1973-1997. *Int. J. Epidemiol.* 34(2): p. 405-412, 2005.

6. Anderson, R., Phase-Based Object Matching using Complex Wavelets, a PhD thesis submitted to Cambridge University, 2007

7. Anderson, R., Kingsbury, N., and Fauqueur, J. Coarse-level object recognition using interlevel products of complex wavelets. in *IEEE International Conference on Image Processing*. 2005

8. Anderson, R., Kingsbury, N., and Fauqueur, J., Determining Multiscale Image Feature Angles from Complex Wavelet Phases. *Lecture notes in computer science*. 3656: p. 490, 2005.

9. Ashikhmin, M., Synthesizing natural textures, in *Proceedings of the 2001 symposium on Interactive 3D graphics*. 2001

10. Ayres, F.J. and Rangayyan, R.M., Characterization of architectural distortion in mammograms. *IEEE Eng Med Biol Mag*. 24(1): p. 59-67, 2005.

11. Azevedo, R. and Lajoie, S.P., The Cognitive Basis for the Design of a Mammography Interpretation Tutor. *International Journal of Artificial Intelligence in Education*. 9(1): p. 32–44, 1998.

12. Babalola, K., Cootes, T., Twining, C., et al., 3D Brain Segmentation Using Active Appearance Models and Local Regressors, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. p. 401-408 2008.

13. Baker, J.A., Rosen, E.L., Lo, J.Y., et al., Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. *Am. J. Roentgenol.* 181(4): p. 1083-1088, 2003.

14. Balleyguier, C., Ayadi, S., Van Nguyen, K., et al., BIRADS classification in mammography. *European journal of radiology*. 61(2): p. 192-194, 2007.

15. Bassett, L.W. and Gold, R.H., The evolution of mammography. *AJR, American journal of roentgenology*. 150: p. 493, 1988.

16. Bick, U., Giger, M.L., Schmidt, R.A., et al., Automated segmentation of digitized mammograms. *Acad Radiol*. 2(1): p. 1-9, 1995.

17. Bliznakova, K., Bliznakov, Z., Bravou, V., et al., A three-dimensional breast software phantom for mammography simulation. *Phys Med Biol*. 48(22): p. 3699-3719, 2003.

18. Bookstein, F.L., Principal Warps: Thin-Plate Splines and the Decomposition of Deformatics. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 11(6): p. 567 -585, 1989.

19. Boyd, N.F., Byng, J.W., Jong, R.A., et al., Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results From the Canadian National Breast Screening Study. *J. Natl. Cancer Inst.* 87(9): p. 670-675, 1995.

20. Burrell, H.C., Evans, A.J., Wilson, A.R., et al., False-negative breast screening assessment: what lessons can we learn? *Clin Radiol*. 56(5): p. 385-8, 2001.

21. Byng, J.W., Boyd, N.F., Fishell, E., et al., Automated analysis of mammographic densities. *Physics in Medicine and Biology*(5): p. 909, 1996.

22. Canny, J., A Computational Approach to Edge Detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. PAMI-8(6): p. 679-698, 1986.

23. Caulkin, S., Generating Synthetic Abnormalities in Digital Mammograms Using Statistical Models, a PhD thesis submitted to University of Manchester, UK, 2001

24. Cerneaz, N. and Brady, M., Finding curvilinear structures in mammograms, in *Computer Vision, Virtual Reality and Robotics in Medicine*. p. 372-382 1995.

25. Chandrasekhar, R. and Attikiouzel, Y. Automatic Breast Border Segmentation by Background Modeling and Subtraction. in *5th International Workshop on Digital Mammography*. Toronto, Canada: Medical Physics Publishing. 2000

26. Claridge, E. and Richter, J. Characterisation of mammographic lesions. in *2nd International Workshop on Digital Mammography*. York, England: Elsevier. 1994

27. Cootes, T.F., Edwards, G.J., and Taylor, C.J. Active Appearance Models. in *European Conference on Computer Vision 1998*. Springer. 1998

28. Cootes, T.F. and Taylor, C.J., A mixture model for representing shape variation.

Image and Vision Computing. 17(8): p. 567-573, 1999.

29. Cootes, T.F., Taylor, C.J., Cooper, D.H., et al. Training Models of Shape from Sets of Examples. in *3rd British Machine Vision Conference*. 1992

30. Cootes, T.F., Taylor, C.J., Cooper, D.H., et al., Active Shape Models - Their Training and Application. *Computer Vision and Image Understanding*. 61(1): p. 38-59, 1995.

31. Cootes, T.F., Twining, C.J., Petrovic, V., et al., Groupwise Construction of Appearance Models using Piece-wise Affine Deformations. *Proc. British Machine Vision Conference*. 2: p. 879-888, 2005.

32. Cowley, H.C. and Gale, A.G., Breast cancer screening: comparison of radiologists' performance in a self-assessment scheme and in actual breast screening. *Proceedings of SPIE*. 3663: p. 157, 2003.

33. Crowley, R.S. and Medvedeva, O., An intelligent tutoring system for visual classification problem solving. *Artif Intell Med*. 36(1): p. 85-117, 2006.

34. Davies, R.H., Twining, C.J., Cootes, T.F., et al., A minimum description length approach to statistical shape modeling. *IEEE Trans Med Imaging*. 21(5): p. 525-37, 2002.

35. Diffey, J., Hufton, A., and Astley, S., A New Step-Wedge for the Volumetric Measurement of Mammographic Density, in *Digital Mammography*. p. 1-9 2006.

36. Dixon, R.N. and Taylor, C.J., Automated asbestos fibre counting. *Institute of Physics Conference Series*. 44: p. 178-185, 1979.

37. Duda, R.O. and Hart, P.E., Use of Hough transformation to detect lines and curves in pictures. *Communications of the ACM*. 15(1): p. 11, 1972.

38. Edwards, G.J., Lanitis, A., Taylor, C.J., et al., Statistical models of face images - - improving specificity. *Image and Vision Computing*. 16(3): p. 203-211, 1998.

39. Efros, A.A. and Freeman, W.T. Image Quilting for Texture Synthesis and Transfer. in *ACM SIGGRAPH*. Los Angeles, CA. 2001

40. Efros, A.A. and Leung, T.K. Texture synthesis by non-parametric sampling. in *7th IEEE International Conference on Computer Vision*. 1999

41. Elliot, C.R., Williams, K.A., and Woolf, B.P., An intelligent learning environment for advanced cardiac life support. *Proc AMIA Annu Fall Symp*: p. 7-11, 1996.

42. Felsberg, M. and Sommer, G., The monogenic signal. *Signal Processing, IEEE Transactions on*. 49(12): p. 3136-3144, 2001.

43.   Ferrari, R.J., Rangayyan, R.M., Borges, R.A., et al., Segmentation of the fibro-glandular disc in mammograms using Gaussian mixture modelling. *Med Biol Eng Comput*. 42(3): p. 378-87, 2004.

44.   Ferrari, R.J., Rangayyan, R.M., Desautels, J.E., et al., Identification of the breast boundary in mammograms using active contour models. *Med Biol Eng Comput*. 42(2): p. 201-8, 2004.

45.   Ferrari, R.J., Rangayyan, R.M., Desautels, J.E., et al. Segmentation of Mammograms: Identification of Boundary, Pectoral Muscle, and Fibroglandular Disc. in *5th International Workshop on Digital Mammography*. Toronto, Canada: Medical Physics Publishing. 2000

46.   Forrest, P., Breast cancer screening: report to the Health Ministers of England, Wales, Scotland & Northern Ireland, HMSO 1986.

47.   Freeman, W.T. and Adelson, E.H., The Design and Use of Steerable Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 13: p. 891-906, 1991.

48.   Freer, T.W. and Ulissey, M.J., Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center1. *Radiology*. 220(3): p. 781-786, 2001.

49.   Gilbert, F.J., Astley, S.M., Gillan, M.G.C., et al., Single Reading with Computer-Aided Detection for Screening Mammography. *N Engl J Med*. 359(16): p. 1675-1684, 2008.

50.   Goodall, C., Procrustes Methods in the Statistical Analysis of Shape. *J. R. Statis. Soc. B*. 53(2): p. 285-339, 1991.

51.   Hall, P., Sheather, S.J., Jones, M.C., et al., On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*. 78(2): p. 263, 1991.

52.   Hayes, D.F., Atlas of Breast Cancer, Mosby Europe Limited 1993.

53.   Heath, M., Bowyer, K., Kopans, D., et al. The Digital Database for Screening Mammography. in *5th International Workshop on Digital Mammography*. Medical Physics Publishing, Madison, WI, USA. 2000

54.   Highnam, R., Brady, M., and English, R. Simulating Disease in Mammography. in *5th International Workshop on Digital Mammography*. Toronto, Canada: Medical Physics Publishing. 2000

55.   Highnam, R., Pan, X., Warren, R., et al., Breast composition measurements using retrospective standard mammogram form (SMF). *Physics in Medicine and Biology*(11): p. 2695, 2006.

56. Highnam, R.P., Brady, J.M., and Shepstone, B.J., Mammographic image analysis. *European Journal of Radiology*. 24(1): p. 20-32, 1997.

57. Hipwell, J., Tanner, C., Crum, W., et al., X-Ray Mammogram Registration: A Novel Validation Method, in *Digital Mammography*. p. 197-204 2006.

58. Hologic. http://www.hologic.com/breast-imaging/. 2009 [cited 1st January 2010].

59. Huang, Y., Liu, Q., and Metaxas, D.N. A component based deformable model for generalized face alignment. in *IEEE Int'l Conf. on Comp. Vis.* 2007

60. Huo, Z., Giger, M.L., Vyborny, C.J., et al., Breast Cancer: Effectiveness of Computer-aided Diagnosis - Observer Study with Independent Database of Mammograms. *Radiology*. 224(2): p. 560-568, 2002.

61. Kallergi, M., Gavrielides, M.A., He, L., et al., Simulation model of mammographic calcifications based on the American College of Radiology Breast Imaging Reporting and Data System, or BIRADS. *Academic Radiology*. 5(10): p. 670-9, 1998.

62. Karssemeijer, N., Automated classification of parenchymal patterns in mammograms. *Phys Med Biol*. 43(2): p. 365-78, 1998.

63. Karssemeijer, N. and te Brake, G.M., Detection of stellate distortions in mammograms. *Medical Imaging, IEEE Transactions on*. 15(5): p. 611-619, 1996.

64. Kass, M., Witkin, A., and Terzopoulos, D., Snakes: Active contour models. *International Journal of Computer Vision*. 1(4): p. 321-331, 1988.

65. Kingsbury, N., Complex wavelets for shift invariant analysis and filtering of signals. *Applied and computational harmonic analysis*. 10(3): p. 234, 2001.

66. Kingsbury, N., Rotation-invariant local feature matching with complex wavelets, in *European Conference on Signal Processing (EUSIPCO)*. 2006

67. Kingsbury, N.G., Design of Q-shift complex wavelets for image processing using frequency domain energy minimization. *Proceedings of the IEEE*. 1: p. 1013, 2003.

68. Kovesi, P., Image features from phase congruency. *Videre: Journal of Computer Vision Research*. 1(3): p. 1, 1999.

69. Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., et al., Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Trans Med Imaging*. 23(9): p. 1129-40, 2004.

70. Lado, M.J., Tahoces, P.G., Souto, M., et al., Real and simulated clustered

microcalcifications in digital mammograms. ROC study of observer performance. *Med Phys*. 24(9): p. 1385-94, 1997.

71.   Lanitis, A., Taylor, C.J., and Cootes, T.F., An automatic face identification system using flexible appearance models. *Image and Vision Computing*. 13(5): p. 393-402, 1995.

72.   Lefebvre, F., Benali, H., Gilles, R., et al., A simulation model of clustered breast microcalcifications. *Med Phys*. 21(12): p. 1865-74, 1994.

73.   Leung, T. and Malik, J., Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. *International journal of computer vision*. 43(1): p. 29, 2001.

74.   Lindfors, K.K., McGahan, M.C., Rosenquist, C.J., et al., Computer-aided Detection of Breast Cancer: A Cost-effectiveness Study1. *Radiology*. 239(3): p. 710-717, 2006.

75.   Lloyd, S., Least Squares Quantization in PCM. *IEEE Trans Information Theory*. 28: p. 129-137, 1982.

76.   MacQueen, J.B. Some Methods for Classification and Analysis of Multivariate Observations. in *5th Berkeley Symposium on Mathematical Studies and Probability*. 1967

77.   Mardia, K., Hughes, G., and Taylor, C.C., A Multivariate von Mises Distribution with Applications to Bioinformatics. *Can. J. of Statist*. 36: p. 99–109, 2008.

78.   McLoughlin, K.J., Bones, P.J., and Kovesi, P. Connective tissue representation for detection of microcalcifications in digital mammograms. in *Medical Imaging 2002: Image Processing*. San Diego, CA, USA: SPIE. 2002

79.   Mellor, M. and Brady, M., Non-rigid Multimodal Image Registration Using Local Phase, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*. p. 789-796 2004.

80.   Mellor, M. and Brady, M., Phase mutual information as a similarity measure for registration. *Medical Image Analysis*. 9(4): p. 330-343, 2005.

81.   Méndez, A.J., Tahoces, P.G., Lado, M.J., et al., Automatic detection of breast border and nipple in digital mammograms. *Computer Methods and Programs in Biomedicine*. 49(3): p. 253-262, 1996.

82.   Miller, P.I., Automated Detection of Mammographic Asymmetry, a PhD thesis submitted to University of Manchester, 1994

83.   Mitchell, S.C., Bosch, J.G., F., B.P., et al., 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE transactions on*

medical imaging. 21(9): p. 1167, 2002.

84. Nappi, J., Dean, P.B., Nevalainen, O., et al., Algorithmic 3D simulation of breast calcifications for digital mammography. *Comput Methods Programs Biomed*. 66(1): p. 115-24, 2001.

85. Ojala, T., Nappi, J., and Nevalainen, O., Accurate segmentation of the breast region from digitized mammograms. *Comput Med Imaging Graph*. 25(1): p. 47-59, 2001.

86. Pan, X.-B. The use of multi-scale monogenic signal on structure orientation identification and segmentation. in *8th International Workshop on Digital Mammography*. Manchester, UK. 2006

87. Parr, T.C., Astley, S., Taylor, C.J., et al. Model based classification of linear structures in digital mammograms. in *Digital Mammography*. Elsevier. 1996

88. Parr, T.C., Taylor, C.J., Astley, S., et al., Statistical modeling of oriented line patterns in mammograms. *Proceedings of SPIE--the international society for optical engineering*. 3034: p. 44, 1997.

89. Pawluczyk, O., Augustine, B.J., Yaffe, M.J., et al., A volumetric method for estimation of breast density on digitized screen-film mammograms. *Medical Physics*. 30(3): p. 352-364, 2003.

90. Pennec, X., Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of mathematical imaging and vision*. 25(1): p. 127, 2006.

91. Petroudi, S. and Brady, M., Breast Density Segmentation Using Texture, in *Digital Mammography*. p. 609-615 2006.

92. Petroudi, S., Kadir, T., and Brady, M. Automatic classification of mammographic parenchymal patterns: a statistical approach. in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. 2003

93. Petrovic, V., Cootes, T.F., Twining, C.J., et al., Automated Analysis of Deformable Structure in Groups of Images. *Proc. British Machine Vision Conference*. 2: p. 1060-1069, 2007.

94. Pisano, E.D., Gatsonis, C., Hendrick, E., et al., Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *N Engl J Med*. 353(17): p. 1773-1783, 2005.

95. Portilla, J. and Simoncelli, E.P., A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International journal of computer vision*. 40(1): p. 49, 2000.

96. Rangayyan, R. and Ayres, F., Gabor filters and phase portraits for the detection of architectural distortion in mammograms. *Medical and Biological Engineering and Computing*. 44(10): p. 883-894, 2006.

97. Romdhani, S., Gong, S., and Psarrou, A. A Multi-View Nonlinear Active Shape Model Using Kernel PCA. in *British Machine Vision Conference*. 1999

98. Rose, C.J., Statistical Models of Mammographic Texture and Appearance, a PhD thesis submitted to University of Manchester, UK, 2005

99. Ruschin, M., Tingberg, A., Bath, M., et al., Using simple mathematical functions to simulate pathological structures--input for digital mammography clinical trial. *Radiat Prot Dosimetry*. 114(1-3): p. 424-31, 2005.

100. Saha, P.K., Udupa, J.K., Conant, E.F., et al., Breast tissue density quantification via digitized mammograms. *IEEE Trans Med Imaging*. 20(8): p. 792-803, 2001.

101. Saunders, R.R., Samei, E.E., Baker, J.J., et al., Simulation of mammographic lesions. *Academic radiology*. 13(7): p. 860-70, 2006.

102. Saunders, R.S. and Samei, E., Characterization of Breast Masses for Simulation Purposes. *Medical Imaging 2004: Proceedings of SPIE*. 5372: p. 242-250, 2004.

103. Schenk, V. and Brady, M. Finding CLS Using Multiresolution Orientated Local Energy Feature Detection. in *6th International Workshop on Digital Mammography*. Bremen, Germany: Springer. 2002

104. Schittkowski, K., NLPQL: A fortran subroutine solving constrained nonlinear programming problems. *Annals of operations research*. 5(1): p. 485, 1986.

105. Selesnick, I.W., The dual-tree complex wavelet transform. *IEEE signal processing magazine*. 22(6): p. 123, 2005.

106. Shapiro, S., Coleman, E.A., Broeders, M., et al., Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. *Int. J. Epidemiol*. 27(5): p. 735-742, 1998.

107. Sharples, M., Jeffery, N.P., du Boulay, B., et al., Structured computer-based training in the interpretation of neuroradiological images. *Int J Med Inform*. 60(3): p. 263-80, 2000.

108. Sickles, E.A., Mammographic features of 300 consecutive nonpalpable breast cancers. *Am. J. Roentgenol*. 146(4): p. 661-663, 1986.

109. Simoncelli, E.P. and Freeman, W.T. The steerable pyramid: a flexible architecture for multi-scale derivative computation. in *Image Processing, 1995. Proceedings., International Conference on*. 1995

110. Skaane, P. and Skjennald, A., Screen-Film Mammography versus Full-Field

Digital Mammography with Soft-Copy Reading: Randomized Trial in a Population-based Screening Programâ€"The Oslo II Study1. *Radiology*. 232(1): p. 197-204, 2004.

111. Skaane, P., Skjennald, A., Young, K., et al., Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. *Acta Radiol*. 46(7): p. 679-89, 2005.

112. Skiadopoulos, S., Costaridou, L., Kalogeropoulou, C.P., et al., Simulating the mammographic appearance of circumscribed lesions. *Eur Radiol*. 13(5): p. 1137-47, 2003.

113. Smith, J., Prediction of the risk of breast cancer using computer vision techniques, a PhD thesis submitted to University of Manchester, UK, 1998

114. Stein, E.M. and Weiss., G., Introduction to Fourier analysis on Euclidean spaces, Princeton University Press, 1972.

115. Tabar, L. and Dean, P.B., Teaching Atlas of Mammography, Thieme Publishing Group 2000.

116. Taylor, P. and Potts, H.W.W., Computer aids and human second reading as interventions in screening mammography: Two systematic reviews to compare effects on cancer detection and recall rate. *European journal of cancer*. 44(6): p. 798, 2008.

117. Unser, M., Aldroubi, A., and Eden, M., Fast B-spline transforms for continuous image representation and interpolation. *IEEE transactions on pattern analysis and machine intelligence*. 13(3): p. 277, 1991.

118. van Engeland, S., Snoeren, P., Hendriks, J., et al., A comparison of methods for mammogram registration. *IEEE Trans Med Imaging*. 22(11): p. 1436-44, 2003.

119. Varma, M. and Zisserman, A. Classifying Images in Materials: Achieving Viewpoint and Illumination Independence. in *European Conference on Computer Vision*. Copenhagen, Denmark. 2002

120. Vyborny, C.J., Giger, M.L., and Nishikawa, R.M., Computer-Aided Detection and Diagnosis of Breast Cancer. *Radiologic Clinics of North America*. 38(4): p. 725-740, 2000.

121. Wai, L.C., Mellor, M., and Brady, M. A Multi-resolution CLS Detection Algorithm for Mammographic Image Analysis. in *MICCAI 2004*. Springer-Verlag. 2004

122. Wei, L.-Y. and Levoy, M., Fast texture synthesis using tree-structured vector quantization, in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000

123. Wei, L.-Y. and Levoy, M., Order-Independent Texture Synthesis. *Stanford Computer Science TR-2002-01*, 2002.

124. Wietzke, L. and Sommer, G., The Conformal Monogenic Signal, in *Pattern Recognition*. p. 527-536 2008.

125. Winsberg, F., Elkin, M., Macy, J., et al., Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*. 89(2): p. 211-215, 1967.

126. Wirth, M.A. and Stapinski, A., Segmentation of the breast region in mammograms using active contours. *Proceedings of SPIE*. 5150: p. 1995, 2003.

127. Wu, Q. and Yu, Y., Feature matching and deformation for texture synthesis. *ACM transactions on graphics*. 23(3): p. 364, 2004.

128. Zhou, C., Chan, H.P., Petrick, N., et al., Computerized image analysis: estimation of breast density on mammograms. *Med Phys*. 28(6): p. 1056-69, 2001.

129. Zwiggelaar, R., Astley, S.M., Boggis, C.R., et al., Linear structures in mammographic images: detection and classification. *IEEE Trans Med Imaging*. 23(9): p. 1077-86, 2004.

130. Zwiggelaar, R., Parr, T.C., Schumm, J.E., et al., Model-based detection of spiculated lesions in mammograms. *Med Image Anal*. 3(1): p. 39-62, 1999.