Model-based Analysis of Mammograms

Nicholas J. Cerneaz

A thesis submitted to the Department of Engineering Science, University of Oxford, in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Department of Engineering Science
University of Oxford

Michaelmas Term
1994

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
30 MAR 1995
THESIS

Model-based Analysis of Mammograms

Nicholas J. Cerneaz

A thesis submitted for the
degree of Doctor of Philosophy
at the University of Oxford

Robotics Research Group
Department of Engineering Science

St. John's College
Michaelmas Term 1994

Abstract

Metastasised breast cancer kills. There is no known cure, there are no known preventative measures, there are no drugs available with proven capacity to abate its effects. *Early* identification and excision of a malignancy prior to metastasis is the only method currently available for reducing the mortality due to breast disease.

Automated analysis of mammograms has been proposed as a tool to aid radiologists detect breast disease earlier and with greater efficiency and success. This thesis addresses some of the major difficulties associated with the automated analysis of mammograms, in particular the difficulties caused by the high-frequency, relatively insignificant curvi-linear structures (CLS) comprising the blood vessels, milk-ducts and fibrous tissues. Previous attempts at automation have been overlooked these structures and the resultant complexity of that oversight has been handled inappropriately.

We develop a model-based analysis of the CLS features, from the very anatomy of the breast, through mammography and digitisation to the image intensities. The model immediately dictates an algorithm for extracting a high-level feature description of the CLS features. This high-level feature description allows a systematic treatment of these image features prior to searching for instances of breast disease.

We demonstrate a procedure for implementing such prior treatment by 'removing' the CLS features from the images. Furthermore, we develop a model of the expected appearance of mammographic densities in the CLS-removed image, which leads directly to an algorithm for their identification. Unfortunately the model also extracts many regions of the image that are not significant mammographic densities, and this therefore requires a subsequent segmentation stage. Unlike previous attempts which apply neural networks to this task, and therefore incorporate inherent insignificance as a consequence of insufficient data availability describing the significant mammographic densities, we illustrate the application of a new statistical method (novelty analysis) for achieving a statistically significant segmentation of the mammographic densities from the plethora of candidates identified at the previous stage.

We demonstrate the ability of the CLS feature description to identify instances of radial-scar in mammograms, and note the suitability of the CLS and density descriptions for assessment of bilateral and temporal asymmetry. Some additional potential applications of these feature descriptions in arenas other than mammogram analysis are also noted.

Acknowledgements

The work to be found on the following pages represents a small fraction of the paths I have travelled during my time in Oxford. Life in this city is well described by an adaptation of a quote of Louis Adamic, who noted "My grandfather always said that living is like licking honey off a thorn." Indeed life in Oxford is often like this, and I have been extremely fortunate to have many friends capable of dipping the thorn back into the honey pot.

From the very beginning, Michael Brady predicted that the thorn would soon fade, leaving only the honey, variously described as the rosy taste of British bitter, the long warm summers and the charm and ambiance of Oxford itself. Having sat through his trial-by-ordeal, I can confirm that indeed the honey is sweet, but his details are slightly skewed. More on this later. As my supervisor, Mike has carefully prodded and guided me out of many a dark patch, salvaging my self-esteem on more than one occasion. Our discussions, both technical and otherwise, have always been a source of inspiration, and it is no understatement to say that Mike's involvement and enthusiasm has raised both the quality of the work reported here and my opinion of Oxford.

My understanding of mammography, the English (as a race) and football, would be substantially inferior were it not for Ralph Highnam, to whom I am greatly indebted. At the Breast Care Unit of the Churchill Hospital, Basil Shepstone and Ruth English have always given freely of their radiological expertise.

Thanks go to Lionel Tarassenko, under whose guidance the work on novelty analysis was conducted. Lionel, like Mike, takes a keen interest in good cricket, and the resulting debates and tussles were made that much more rewarding by the fine result of last summer's Ashes Tour! Indeed, England is a nice place in summer, with long nights, imported lager and an Ashes victory . . .

I have been fortunate that this work was undertaken in an environment conducive to excellence. In addition to the discussions with Ralph, Mike, and Lionel, those with Siew-Li Kok, Nicola Ferrier, Bill Triggs, Larry Shapiro, David Sinclair and Rupert Curwen have also been a tremendous help. I also wish to thank the Rhodes Trust for their financial support.

To my friends who have made life in Oxford such a treat and thankfully done their best to hinder the progress of this thesis, especially Jamie, Sarah, Rosy, Julie, Kath, Davo, Kato, Cath, Danny, Myles and Rob, I look forward to many more good years of friendship, deep reds and fine food.

There is a special bunch of people who deserve a gold medal for their efforts to drag me up to the swimming pool and around the countryside on my bicycle in a (futile?) attempt to maintain some fitness, let alone some sanity. To Rosy, Cath, Nicola, Rob, Davo, Danny and Kato, and above all Kath for her cavalier wagers, thank you.

Further away, though never far, Mum, Dad, Anna, Jane and Peter have consistently supported my every endeavour, and to them I offer my deepest thanks.

Nick Cerneaz
October 1994

Contents

1 Introduction **1**
- 1.1 Incidence of Breast Cancer . . . 1
- 1.2 Automating mammogram analysis . . . 3
- 1.3 Model-based mammogram image analysis . . . 6
- 1.4 Thesis Outline . . . 10

2 Breast Anatomy, Oncology and Disease Detection **12**
- 2.1 Breast Anatomy . . . 12
  - 2.1.1 The Lobes of the Breast . . . 14
  - 2.1.2 Connective and Supportive Stroma . . . 14
  - 2.1.3 Lymphatic System . . . 14
- 2.2 Breast Oncology . . . 15
  - 2.2.1 Malignant abnormalities . . . 15
  - 2.2.2 Benign abnormalities . . . 16
  - 2.2.3 Calcifications . . . 18
- 2.3 Techniques for detecting breast abnormalities . . . 19
  - 2.3.1 Clinical indicators . . . 19
  - 2.3.2 Imaging techniques . . . 22
  - 2.3.3 Summary . . . 28
- 2.4 The Mammographic Process . . . 29
- 2.5 Mammographic signs of breast disease . . . 30
- 2.6 Summary . . . 33

3 Automated mammogram interpretation **34**
- 3.1 Calcifications . . . 36
- 3.2 Mammographic densities . . . 39
- 3.3 Evaluation of parenchymal pattern . . . 43
- 3.4 Noise . . . 43
- 3.5 Segmentation . . . 44
- 3.6 Bilateral registration and asymmetry . . . 45
- 3.7 Intensity normalisation . . . 46
- 3.8 General image enhancement . . . 47
- 3.9 Other techniques . . . 50
- 3.10 Summary . . . 50

4 Finding the curvi-linear structures (CLS) **52**
- 4.1 Introduction . . . 52
  - 4.1.1 Anatomical attributes of the CLS features . . . 52
  - 4.1.2 Mammographic attributes of the CLS features . . . 53
  - 4.1.3 Image Processing Constraints . . . 55

4.1.4 Summary . . . 58
4.2 CLS Detection — Related Work . . . 58
4.3 The CLS detector algorithm . . . 70
4.4 Finding the CLS pixels — Low-level description . . . 71
4.4.1 Modelling the expected CLS profile . . . 71
4.4.2 From continuous to discrete image surfaces . . . 79
4.4.3 Is the elliptical CLS model adequate? . . . 82
4.4.4 Poly-energetic beams and scattering . . . 84
4.4.5 Film-density or Transmitted-light Image Representations? . . . 87
4.4.6 Spatial extent of the CLS features . . . 89
4.4.7 CLS Modelling Conclusions . . . 92
4.4.8 Finding the image surface curvature . . . 93
4.4.9 Merging the directed-difference information . . . 96
4.4.10 Forming the set **N** of all CLS pixels . . . 98
4.5 Region Thinning . . . 99
4.6 Linking Cantons into Bones . . . 101
4.6.1 Introduction . . . 101
4.6.2 The algorithm . . . 102
4.6.3 Deleting known insignificant cantons/bones . . . 106
4.7 Finding the CLS 'flesh' . . . 106
4.8 Results . . . 107
4.9 Stability Analysis . . . 110
4.9.1 Second difference response threshold T . . . 112
4.9.2 Noise . . . 115
4.10 Summary . . . 116

5 Applications of the CLS feature description **121**
5.1 Simulation of an image without CLS features . . . 121
5.2 Radial scar . . . 130
5.3 Differential Compression Mammography . . . 131
5.4 Asymmetry and temporal changes . . . 132
5.5 Other Applications . . . 133
5.5.1 Synthetic Aperture Radar (SAR) images . . . 133
5.5.2 General ridge and ravine finding . . . 138
5.6 Summary . . . 141

6 Finding mammographic densities **142**
6.1 Introduction . . . 142
6.2 Identifying blobs in images . . . 145
6.3 Proposing potential densities . . . 149
6.3.1 Blob saliency metrics . . . 150
6.3.2 Finding the image surface gradient . . . 153
6.3.3 The Blob-detector algorithm . . . 158
6.3.4 Results . . . 160
6.4 Identifying significant densities . . . 160
6.4.1 Previous work . . . 164
6.4.2 Novelty detection . . . 166
6.4.3 Implementation and results . . . 175
6.5 Conclusions . . . 177

7 Conclusions and Future Work **182**
7.1 Conclusions . . . 182
7.2 Future Work . . . 184

A Region thinning **189**
A.1 Thinning algorithm requirements 189
A.2 Related work in region thinning 193
A.3 Haralick & Shapiro's thinning algorithm 194
A.3.1 Discussion 201
A.4 An extension to Haralick & Shapiro's thinning algorithm 204
A.4.1 Introduction 204
A.4.2 The details 205
A.5 Limitations of the Extended Haralick & Shapiro algorithm 210
A.6 Region Thinning Conclusions 211

Bibliography **213**

1

Introduction

1.1 Incidence of Breast Cancer

It is estimated that in the United States 182 000 women and 1000 men will be diagnosed as new cases of breast cancer during 1994, whilst 46 000 women and 300 men will die of breast cancer [16]. This rate of mortality (22.4 per 100 000 population) places breast cancer second only to lung cancer as a cause of cancer related deaths in women. It is the leading cause of cancer related death in women aged 15–54. In England and Wales the situation is worse, where the mortality rate is the highest in the world at 28.7 deaths per 100 000 population [16]. During 1994, 15 000 English and Welsh women will lose their lives to breast cancer. One in eight women will contract the disease during their life [63].

Breast cancer is a heterogeneous collection of diseases which behave in an unpredictable manner with extreme clinical variability. The disease can develop within many of the tissues of the breast, initially doing so in an in situ phase. At some point in its development it will metastasise and shed cancer cells that spread, generally via the bloodstream, to other parts of the body, such as the bone, liver, lungs or brain. The breast is not a life-sustaining organ and therefore it is not the cancer of the breast itself that kills a patient, rather it is the secondary disease developing at distant sites that contribute to the patient's death, such as bone marrow failure, liver failure, or respiratory failure.

There is much uncertainty surrounding the risk factors leading to the development of breast cancer, however it is universally agreed that we know not of, nor can we therefore implement, a preventative strategy. Consequently the management of breast cancer is undertaken by treating the known cases of the disease. In order to effectively treat cases of breast cancer with the aim of reducing the mortality rate, it is not only necessary that the instances of the disease be identified, but they must be identified at a stage of disease development for which

the available medical procedures are effective, i.e., at the stages for which the prognosis is good.

The medical procedures available today for treating the disease do not extend to tackling the problem once the cancer has metastasised, and in fact the only real option is surgical intervention to remove the in situ cancerous growth. It is therefore imperative that cancers be identified prior to metastasis. Further to this it is acknowledged that the earlier that breast cancer is diagnosed, the better the prognosis.

So protean are the manifestations of breast cancer and often so subtle are its indications that the early detection of all instances of breast cancer is very problematic. The majority of patients diagnosed with breast cancer are asymptomatic, a reflection of this diversity and subtlety, and it is these factors that ultimately negate any attempt to develop a preventative approach to the management of this disease. To illustrate this point Henry Leis, Jr. offers the following [132]:

> "Perhaps the best way to summarise all the currently known risk factors for breast cancer is to view them in terms of the ultimate search for the woman at highest risk. Such a search would culminate in the finding of a 58-year-old, wide body-type, hypertensive, diabetic, hypothyroid, Caucasian, Jewish convert nun, taking reserpine for hypertension and estrogens for severe climacteric symptoms, living in a cold climate in the western hemisphere, whose mother and sister had bilateral premenopausal breast cancer, who has a wet type of ear wax, a low estriol titer and subnormal androgen excretion levels, who had previous endometrial cancer and cancer in one breast, who nursed from her mother who had B viral particles in her milk, whose menarche was at the age of 9, whose remaining breast reveals precancerous mastopathy on the random biopsy, a DY parenchymal pattern on X-ray, and an abnormal thermogram, who received multiple fluroscopies for tuberculosis therapy when she was 19 years of age, who had severe hepatitis and now has liver dysfunction, and who lives on a high-fat, high-beef, low-fish diet, deficient in both vitamin C and B complex, and who drinks an excessive amount of coffee and dyes her hair."

Strax [201] however, states that:

> "a statistical study by the ACS [American Cancer Society] teaches us that less than 25% of cancers develop in women with any of these [risk] factors. The vast majority, up to 75% of breast cancers, occur in women with none of the known risk factors. The sad truth is that we have to consider all women over the age of 30 at risk for this disease."

Since every woman is a candidate for breast disease, and there is no way to identify a symptomatic population, and the incidental clinical indicators of breast disease only appear at a stage of the disease for which the prognosis is very poor, the management of breast disease must be based on active and dedicated assessment of each and every candidate for the disease. There are a number of options available to the clinician to aid the detection of early breast cancer: mammography, ultrasound, palpation, magnification mammography, fine needle aspiration biopsy and open surgical biopsy. On the basis of many studies (HIP [187], Swedish two-county [208], Utrecht [41], Nijmegen [216], etc.) evidence for the effectiveness and

superiority of mammography as the primary tool for identifying early stage breast cancer was gathered. With this evidence the medical communities of many nations were able to secure the vital government backing necessary to establish national screening programs based on mammography.

In the United Kingdom, the 1986 Forrest Report [69] to the UK Health Ministers carried the recommendation that all women between the ages of 50 and 64 years be screened for breast cancer. The report further recommended that the screening programs be based on "high quality single medio-lateral oblique view mammography." These recommendations were accepted by the government and the National Health Service Breast Screening Program (NHSBSP) was subsequently established during the following years. In a recent review of the NHSBSP [50] it was estimated that "by the year 2000, the screening program is expected to prevent about 25% of deaths from breast cancer in the population of women invited for screening. ... On average each of the women in whom death from breast cancer is prevented will live about 20 years more. Thus by the year 2000 the screening program is expected to result in about 25 000 extra years of life gained annually in the United Kingdom."

With the benefits of screening established, we must now consider the drawbacks. Although mammography has superior performance over all other techniques in the screening environment (i.e., it has the best balance between sensitivity and specificity), the diagnosis of breast cancer is, for many reasons, still difficult. These reasons include: the ambiguity of some signs, subtle differences in X-ray attenuation can be crucial, important signs are often hidden behind other dense tissues by the projective nature of the image, the images can sometimes be of a poor quality. These difficulties mean that some 8% of cancers are not identified during screening [146] and 70–80% of open surgical biopsies are benign [189].

One of the biggest difficulties associated with a 'screening' philosophy is that huge numbers of mammograms must be viewed and interpreted by trained professionals. In the United Kingdom alone, three million mammograms are acquired annually, each requiring assessment. The burden this places on the clinicians who must view and assess the films is enormous, and clearly the potential for error is both large and very undesirable.

1.2 Automating mammogram analysis

From the very early stages of mammography it was acknowledged that automation of some or all of the tasks involved in reading mammograms could be an enormous benefit. In 1967, Winsberg et al. [222] wrote: "Because of the problems inherent in the routine viewing of large numbers of examinations of presumably asymptomatic patients we have proposed the automation of reading of the radiographs ...."

There have been many proposed applications of automated technologies to aid the analysis of mammograms, yet they may be grouped into two main categories: those that complement the diagnostic clinician, and those that seek to emulate their function (including both the direct emulation of their actions/procedures, or by identifying new procedures and extracting other

information from mammograms not normally available to the clinician). In recent times it has been the more complex task of emulating the function of a diagnostic clinician, i.e., interpreting a mammogram, that has been the subject gaining most attention. The ultimate long term goal of this work is to achieve fully automated diagnosis, however there are many technical and ethical considerations to address before this goal might be realised. More realistically achievable goals include:

Prompting The clinician is prompted to look at certain regions of the mammogram which have been deemed suspicious by the automated analysis.

Second opinion Mammograms are viewed by two clinicians in order to improve the quality of the assessment of any given film. An automated analysis system might take the place of one of the clinicians.

It is important to realise that the success of any attempt to automate the analysis of mammograms is inherently linked to the technology chosen as the framework for the analysis, the analysis itself, and the reaction of the radiological/patient communities to that technology/implementation. The most obvious framework for automating the analysis of mammograms with today's technologies is that of the computer. There is, of course, nothing to say that this technology will ultimately provide the best implementation of any analysis algorithms, however due to the versatility of computer programs and the ease of changing and testing new ideas, the computer provides a framework that is sufficiently flexible and powerful to quickly assess those new ideas and proposals. Consequently, the work reported throughout the literature, and indeed in this thesis, has been implemented within the framework of computerised analysis as a collection of computer programs.

Using computers as the framework for automated analysis carries with it important constraints. The popularisation of the computer since the introduction of the personal computer in the early 1980's has precipitated an impression amongst some people that although it is fallible, it does have the potential to accomplish remarkable feats, feats beyond, not simply the manual processing capabilities of a human, but also beyond the logical, reasoning capacities of a human. Clearly this is not the case, but it is a point not fully appreciated by many people unfamiliar with computers. The importance of this point is two-fold: (i) the support of the medical/radiological community for an automated analysis system must *not* be based upon an unfounded belief in the capabilities of computers, but rather in a thorough understanding of the algorithms, *including* an understanding of their limitations, and (ii) the algorithms must be logically related to the underlying task that they seek to accomplish, not merely an intermediate step that is *assumed* to relate to the underlying task.

Regardless of the framework chosen for implementing an algorithm, it is imperative that the algorithm itself be accurate and defensible under scrutiny. A flawless implementation of a poor algorithm will naturally give poor results, as will a flawed implementation of a very good algorithm. The difficulty with this last scenario is that in the context of a computer program, tiny implementation flaws can develop into both catastrophic errors, and also into insipid

errors that may go unnoticed for some time, perhaps indefinitely. Examples of such errors are well documented in nearly every field of endeavour, and include the following examples within medical applications [68]:

- In 1992 a hospital in Stoke-on-Trent discovered an unnecessary software 'correction factor' inadvertently added to the radiation therapy control system. During the preceding 10 years 989 patients receiving treatment for bladder, pelvis, lung and throat cancer had received radiation doses 30% *below* their prescribed value. Although officials denied that the error had had a "deleterious effect", only 447 of the 989 patients treated were still alive in February 1992.
- During 1985 and 1986 a subtle software error in the computerised Therac 25 X-ray machines at the East Texas Cancer Center in Tyler, Texas, and the Kennestone Regional Oncology Center in Galveston, led to the *overdosing* of patients undergoing radiation treatment. The error caused some tissues to receive between 17–25 krads. Research has shown that doses as low as 1 krad delivered to the whole body can be fatal. The final impact of the error was one fatality, several bad burns and partial paralysis in others.
- On 26 October 1992, London's computerised ambulance dispatch system went "live" without manual backup—despite repeated problems in a local trial earlier in the year. After the system received 2900 emergency calls (600 more than a 'normal' day), it sent several ambulance to the same request, diverting them from other emergencies. Following a 36-hour breakdown, health service unions claimed that there were 10 to 20 cases where people had possibly died because of delays in ambulances reaching them. At least one non-fatal case had to wait 11 hours before an ambulance arrived.
- It doesn't have to be a large and complex system to break down due to unforeseen factors or errors: in 1980 a man undergoing microwave arthritis therapy was killed when the therapy reprogrammed his pacemaker.

The point of this discussion is not to denigrate automated philosophies based on computer systems, but rather to acknowledge that even well implemented algorithms can produce errors that are very difficult to explain to the end user. There is no doubt that the computerised ambulance dispatch system implemented in London contained many highly sophisticated algorithms that probably worked flawlessly. An appreciation of precisely why that system collapsed might only be possible for the computer programmers who implemented the system, and the answers might not seem very convincing to the patients who waited for an ambulance. Bad publicity is bad publicity.

In light of these and other similar incidents, clinicians are justifiably wary of the superlatives used to describe any computerised system. It is therefore a fundamental requirement of any automated system that its operation be fully explainable in terms understood by the clinician, and that the system be seen to operate as *expected* by those clinicians. There is only one way to achieve these requirements, and that is to build a model of the data and to implement

algorithms based on that model. If the model is couched in terms of the information in a mammogram, and the algorithms of analysis exploit the attributes of the features within that information, the very attributes and features that the clinician recognises and understands, then so too will the clinician appreciate the algorithms and their results. These are the tenets of the 'model-based' approach to analysis, and it is this approach that is adopted in the present work.

The importance of gaining the trust and understanding of the clinician cannot be stressed too highly, particularly when considering the legal and ethical dilemmas they face. Additionally, the need for accuracy is imperative when considering the scale of the task at hand: a single error in 10 000 cases (99.99% success rate) would affect 300 women annually in the UK alone.

It is within this context that we aim to automate the analysis of mammograms by using a computer to implement algorithms developed from a model of the processes forming the raw data. The model must begin right from the anatomy of the breast, and proceed through the mammography process to the image intensities in the computer memory in order to fully capture the attributes of the anatomical features as they appear in the image[1].

1.3 Model-based mammogram image analysis

Previous attempts to automate the analysis of mammograms have approached the problem by emulating the procedures and actions of the diagnostic radiologist. Although philosophically equivalent, the digital implementation of those algorithms none-the-less reflects the specific constraints of the digital environment, and thus may appear relatively obtuse, cumbersome and laboured, despite the general equivalence of the philosophy founding the approaches. Of course this is merely a reflection of the difficulty a computer vision system has in emulating the very complex tasks accomplished effortlessly by the human visual system, and this gives rise to the *apparent* disparity in complexity between human and automated assessment. Coupled with the diversity and subtlety of the mammographic indicators of breast disease, the lot of an automated analysis system is indeed a complex one. It is little wonder then that the task has been kept as simple as possible by utilising the wealth of knowledge and experience of the radiological community to dictate the goals and direct the procedures of the automated analysis. In general the mammographic features most indicative of disease are microcalcifications, densities and asymmetry (both bilateral and temporal), and like the radiologist who seeks to identify instances of these features in a mammogram, so too does the approach adopted by the automated techniques.

A consequence of attempting to directly emulate a radiologist by searching for calcifications and masses etc., is that sometimes the techniques overlook the very complex tasks that radiologists undertake and accomplish with ease in the course of their analysis. The consequences

[1]Throughout this thesis we will use the word *image* to refer to the digital representation of the mammogram film stored in the computer memory, and retain the word *mammogram* as a reference to the piece of plastic film normally termed a mammogram. When the word mammogram is appended by the word image, as in *mammogram image*, that shall be taken to refer to the image, not the mammogram itself.

of such oversights (or the deliberate attempt to simplify the problem by omitting some tasks) is generally to introduce some additionally difficulty that must be accommodated or corrected at a later stage. Given the formulation of the analysis, these difficulties are either accommodated/corrected by an ad hoc procedure or left unaddressed altogether. When implemented, the extra complexity of the correction is normally accepted under the assumption that the computerised approach will be much more complex anyway (as previously mentioned), and this is just an example of such complexity. Normally there is substantial difficulty justifying the details of the correction since they are specific to the precise situation which results from a number of processing stages previously undertaken, and thus the correction generally bears little relation to the original data in the image, let alone the anatomy of the breast giving rise to the present difficulties. In those instances where these difficulties have been specifically addressed, they have often required a complex and substantial effort. Generally the details of the solution are couched in the framework of the particular problem are therefore application specific.

The classic example of this situation involves the blood-vessels, milk-ducts and fibrous tissues of the breast. Mammographically these structures appear as nominally 1-dimensional, locally-linear features of slightly lower film density than the surrounding tissues, and are therefore known as the *curvi-linear structures* (CLS) of the breast. When imaged, these structures retain their 1-dimensional, linear attributes and appear with image intensities slightly higher than the surrounding tissues. These tissues are not themselves indicators of breast disease (although breast disease often initiates within the CLS tissues), and they are therefore generally not the subject of investigation by a radiologist (the notable exceptions to this generalisation are the fibrous spicules associated with spiculated masses). Consequently the CLS structures are not systematically searched for by a radiologist, although when identifying other features such as calcifications and masses, these features are easily identified and accounted for by the radiologist. The situation is not quite as simple when undertaking automated analysis of mammogram images. The formulation of algorithms for identifying the microcalcifications and masses etc., normally rely on identifying a few attributes of those features that are consistent with their definition and reliably accompany any instance of such features, for instance small area, high intensity patches indicate microcalcifications. Unfortunately these attributes are not solely the preserve of microcalcifications, and thus the identification of these attributes in an image identifies more features than simply the microcalcifications. An identical argument applies to densities, and the textural analysis associated with the general assessment of parenchymal tissues. In all these instances the CLS features so easily accommodated by the radiologist are not so easily separated from the list of features identified during the automated analysis, and the effects they contribute to the automated search for breast disease not so readily accommodated/corrected. Consider the following examples from the literature which illustrate these types of difficulties. Karssemeijer [109] describes a stochastic technique for detecting calcifications in mammograms by segmenting the pixels of an image into the classes of 'calcification' and 'background'. The introduction of a third class to explicitly accommodate

the "thin bright patches of connective tissue" was required as these features could not be adequately described by the original two classes. Valatx et al. [215], although acknowledging the difficulty their microcalcification detection scheme has with "parenchymal structures", choose to ignore the difficulty. Lau and Bischof [131] introduce a 'directionality' metric to account for the "strong responses ... generated in regions with blood vessels or ... glandular tissues" in their search for densities. Clearly there are instances in the literature where the CLS features have posed a difficulty. Some authors have considered the difficulty explicitly, others have developed relatively ad hoc solutions, whilst others have left the problem simply unaddressed.

In this thesis we begin by developing a systematic treatment of the CLS structures, which can then be used to aid the analysis of an image in search of breast abnormalities. We develop a model of the CLS features based on the physical characteristics of the very breast structures giving rise to their existence. The model, beginning with breast anatomy, follows the mammography process through film acquisition and then digitisation to the image intensities stored in the computer memory. With a direct understanding of the image attributes describing the CLS features we can directly identify their presence, extracting a high-level feature description for a given image. As a preview of the results from the algorithms developed in the present work consider figure 1.1(b) which illustrates the pixels of the CLS features identified in the original image of figure 1.1(a).

The CLS feature description allows a range of subsequent image analysis tasks to be undertaken. We describe a technique for 'removing' the CLS features from an image. The resultant CLS-removed image (figure 1.1(c)) has removed the high-frequency information associated with the CLS features, yet retained the high-frequency information associated with other processes, for instance the edges of densities. This *selective* low-pass filtering produces an image retaining the characteristics of the densities and general parenchymal structure of an image whilst removing the effects of the CLS features. We assert that such an image can be used to identify the densities in an image, and also to identify the other parenchymal structure for a subsequent assessment of bilateral or temporal asymmetry.

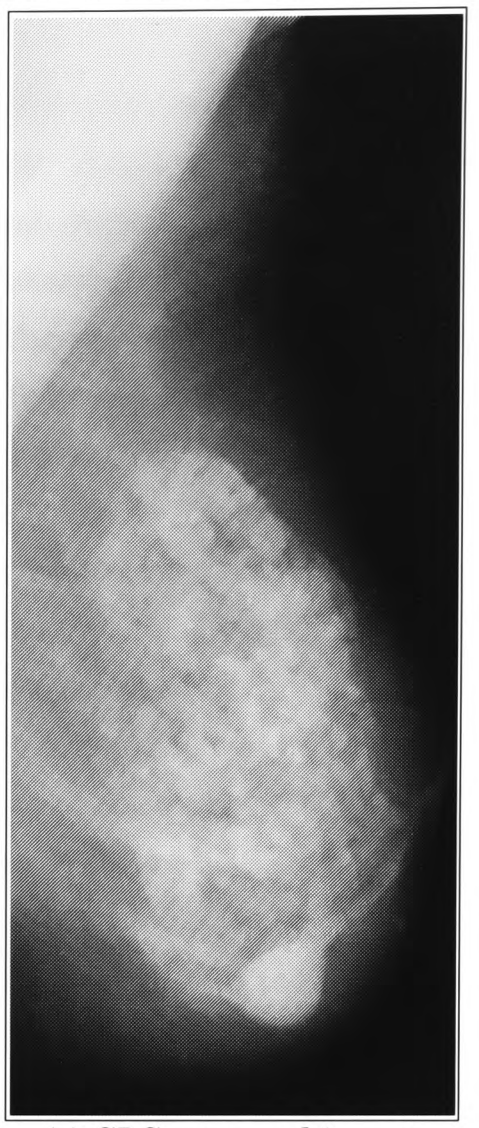
Further to this assertion, we develop a procedure for extracting a description of the locally salient regions of higher image intensity (the 'blobs') from the CLS-removed image. A model of a mammographic density is developed as a region of higher image intensity resulting from the increased attenuation of the X-ray beam as it traverses the extra material density of the appropriate anatomical features. The blob detector developed is therefore used to identify the regions of higher image intensity in the CLS-removed image. Figure 1.1(d) illustrates the bounding contours of the blobs identified in the image of figure 1.1(c).

The salient blobs include both the densities (which may or may not indicate disease processes) and many other regions of no significance (in terms of breast disease) and we therefore need to segment the description into the two classes of significant densities and insignificant blobs[2]. As previously mentioned, the indications of breast cancer are extremely diverse and of-

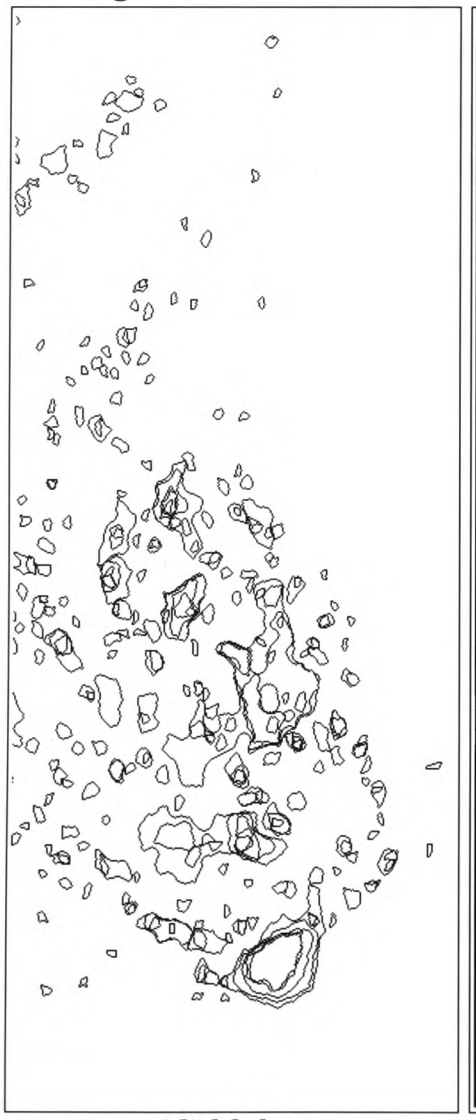
[2]Note that throughout this thesis, the term *blob* refers to all the blobs identified in an image whilst the terms *density* or *significant densities* refer to the real instances of mammographic densities (including both anatomical densities—

(a) original image

(b) CLS features



(c) CLS-removed image



(d) blobs

(e) blob novelty.

Figure 1.1: Preview of results presented in this thesis.

ten very subtle. This diversity and subtlety is also reflected in the indicators of densities alone, making the identification of the significant densities problematic. Traditional approaches have used neural networks to effect the segmentation, however these implementations have, without exception, been extremely limited by the lack of sufficient examples of significant densities to train the networks. At best the statistical significance of the applications reported in the literature is questionable, and at worse simply unacceptable. To overcome these difficulties we employ a new statistical technique, termed *novelty analysis*. A description of the insignificant densities is learnt from the enormous amount of data we have available, and all the blobs are tested for *novelty* against this description of *normality*. In the preliminary work presented in this thesis we have demonstrated the suitability of this technique for identifying the significant densities from the list of blobs identified in each image. Figure 1.1(e) illustrates the relative novelty of the blobs shown in figure 1.1(d), with white depicting highest novelty and dark grey through black indicating normal (note that the range of the normal class extends beyond black due to the formulation of the novelty analysis—see the caption to figure 6.17 on page 178 for further explanation).

The work presented in this thesis is couched within the framework of a model-based approach. By developing models of the CLS features and the densities, we proceed directly to algorithms for identifying these features. Consequently the algorithms are explainable in terms of the breast anatomy, fulfilling a key requirement of automated analysis systems as previously mentioned. The high-level feature descriptions support a range of subsequent processing, in both mammogram image analysis and more generally for any application which can sustain the assumptions of the models (some examples are noted at appropriate times during the following chapters).

1.4 Thesis Outline

The layout of the following pages essentially follows the presentation of ideas in the preceding paragraphs, although to place all the subsequent material in context **Chapter 2** reviews the anatomy and oncology of the breast, concluding with a description of the mammography process. This then describes the acquisition of the raw material for this work (the mammogram film) and the features indicating breast disease that ultimately form the subject of the analysis we seek to automate.

Chapter 3 provides a review of the literature describing attempts to automate the analysis of mammograms. **Chapter 4** develops a model of the CLS features and their expected attributes in an image. Identification of those attributes allows the extraction of a high-level feature description from an algorithm also presented and developed in that chapter.

Using the CLS feature description so obtained, **Chapter 5** proposes a technique for remov-

benign and malignant—and the unfortunate overlapping of parenchymal structures that appear as densities and cannot be distinguished from them on the basis of a single mammogram), whilst the term *insignificant blobs* refers to those locally salient blobs identified that are not considered to be mammographically significant.

ing the CLS-features from an image to give the CLS-removed image. This technique is further illustrated by its application to high contrast images taken from another arena of computer vision, illustrating both the CLS-removal procedure in greater clarity and an alternative application of the CLS feature description. Techniques for identifying instances of radial-scar and improving the analysis of differential compression mammography are also proposed.

Chapter 6 develops a model of the expected attributes of mammographic densities in the CLS-removed image. The model leads directly to an algorithm for identifying the locally salient blobs in an image. Finally, the remainder of this chapter describes the novelty analysis techniques used to discriminate between the mammographically significant densities and the insignificant blobs identified by the blob detector.

2

Breast Anatomy, Oncology and Disease Detection

This chapter begins by describing the anatomy of the breast and then its oncology. This introduction provides the basis for a review of the imaging modalities available for the detection of breast abnormalities, noting the opinion of the literature that mammography is the only viable technique for detecting *early* stage, clinically occult lesions within a screening setting. Thus, the mammographic signs of abnormality are then described in more detail, providing the foundations upon which the automated detection/diagnosis procedures developed in the subsequent chapters are based.

2.1 Breast Anatomy

Until the onset of puberty, the structure of the male and female breast is identical. During the fifth week of interuterine life two thickened bands of ectoderm appear on the ventral surface of the developing embryo. The bands, which run along the *nipple line* from the clavicle to the groin, soon regress leaving only two small segments corresponding to the future breasts. During the sixth month of interuterine life, rudimentary lactiferous ducts develop within the ectoderm bands.

At birth the child has well developed, although small *nipples*, between 15 and 20 lactiferous ducts, and a complete *areola*. The breast tissue remains within the margins of the areola, and the lactiferous ducts each possess a dilated *ampulla* just prior to termination in a separate orifice at the nipple. Each duct leads to a separate segment or *lobe* of the breast from which it will draw milk later in life. These lobes are arranged radially around the breast extending out from the nipple.

Blood supply for each lobe is derived from the *intercostal*, *internal* and *external mammary*,

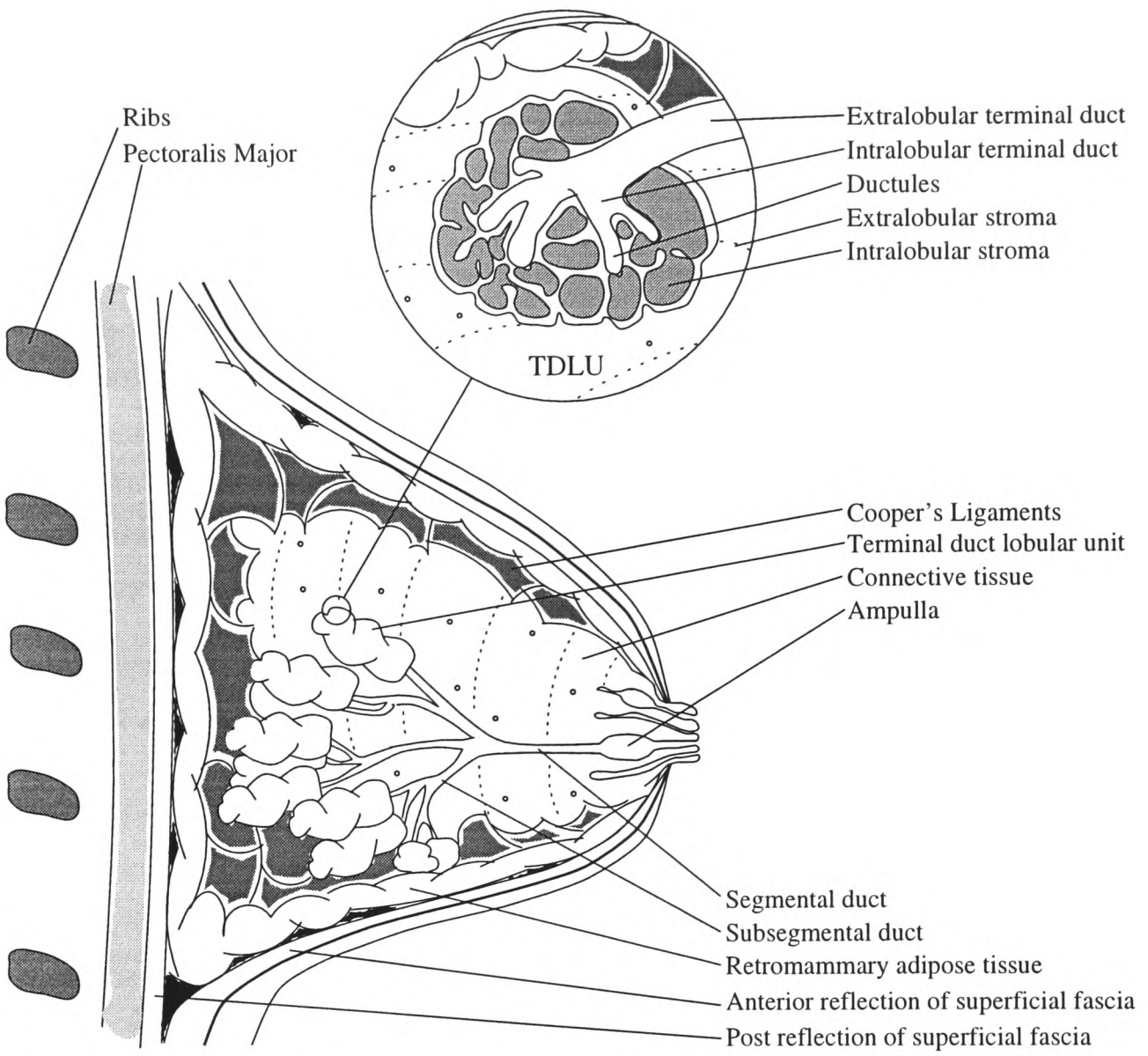


Figure 2.1: Cross-sectional structure of the breast.

and *subscapular* arteries in varying proportions. There are both deep and superficial venous networks, and although they are essentially bilaterally symmetric, dramatic variations between individuals are common. A fully developed lymphatic system exists, and its structure mimics that of the segmented nature of the breast. Motor and sensory nerves enter the lobes from the posterior surface. It is in this state that the breast remains dormant until shortly before puberty.

Development of the female breast into its final functional form progresses in two steps: initially under the influence of ovarian hormone secretion during puberty; and then under the hormonal influences associated with pregnancy.

Following the pubertal influences, the breast is larger due to an increase in both the non-functional supportive tissues (stroma and adipose) and a proliferation of the lactiferous duct system. Figure 2.1 shows a cross-section of the developed female breast in its rest state (i.e., not pregnant).

2.1.1 The Lobes of the Breast

Each main duct branches within the lobe into a series of progressively smaller *subsegmental ducts* until finally becoming the *extralobular terminal duct* upon reaching the *lobule*. The lobule, which is approximately 1–2*mm* in size, contains the milk-producing elements of the breast. Once inside the lobule, the duct divides into the *intralobular ducts* and ends at the 10–100 or so *ductules* within a lobule. It is currently believed [6, p123] that acini are only truly formed within the ductule during pregnancy, come to full maturation during lactation, and disappear at its completion. The physiologically functional unit from the extralobular terminal ducts to the terminal ductules is termed the *terminal duct lobular unit* (TDLU), and it is considered important in the search for breast cancer since it is from the epithelium of the TDLU that many carcinoma originate [221].

2.1.2 Connective and Supportive Stroma

The breast is enveloped by the *pectoria fascia* and the *superficial fascia* of the skin. A series of fibrous membranes, known as *Cooper's Ligaments*[1] extend from the deep pectoria fascia, through the breast and onto that of the skin. These membranes incompletely encapsulate the individual lobes of the breast and provide both the structural framework to support the lobes (and the breast generally), and an anchor to locate the breast anterior to the pectoralis major muscle.

As a result of multiple pregnancies and lactation, or increasing obesity and age, Cooper's ligaments can stretch, resulting in the evolution of the breast from its virginal firmness to pendulous old age.

2.1.3 Lymphatic System

The lymphatic system mimics the segmented structure of the ducts, located alongside them. Below the areola the lymph system is a highly anastomosing network linking the many segments of the breast. Lymph is drained into the deep submammary or a superficial subareolar plexus. From here it radiates laterally to the *axillary* nodes (pectoral, central and apical), upward to the *infraclavicular* and *supraclavicular* nodes, medially to the contralateral breast and through the *intercostal spaces* to the anterior *intercostal* nodes, and inferiorly through the abdominal wall to the *mediastinal* nodes via the *extraperitoneal* tissue. The majority of lymph, however, travels via the axillary nodes [30, 83].

Although the bloodstream is considered the primary route for metastasis it is believed that lymph node activity is highly indicative of such a condition [30].

[1] After Sir Astley Cooper who first identified them [42].

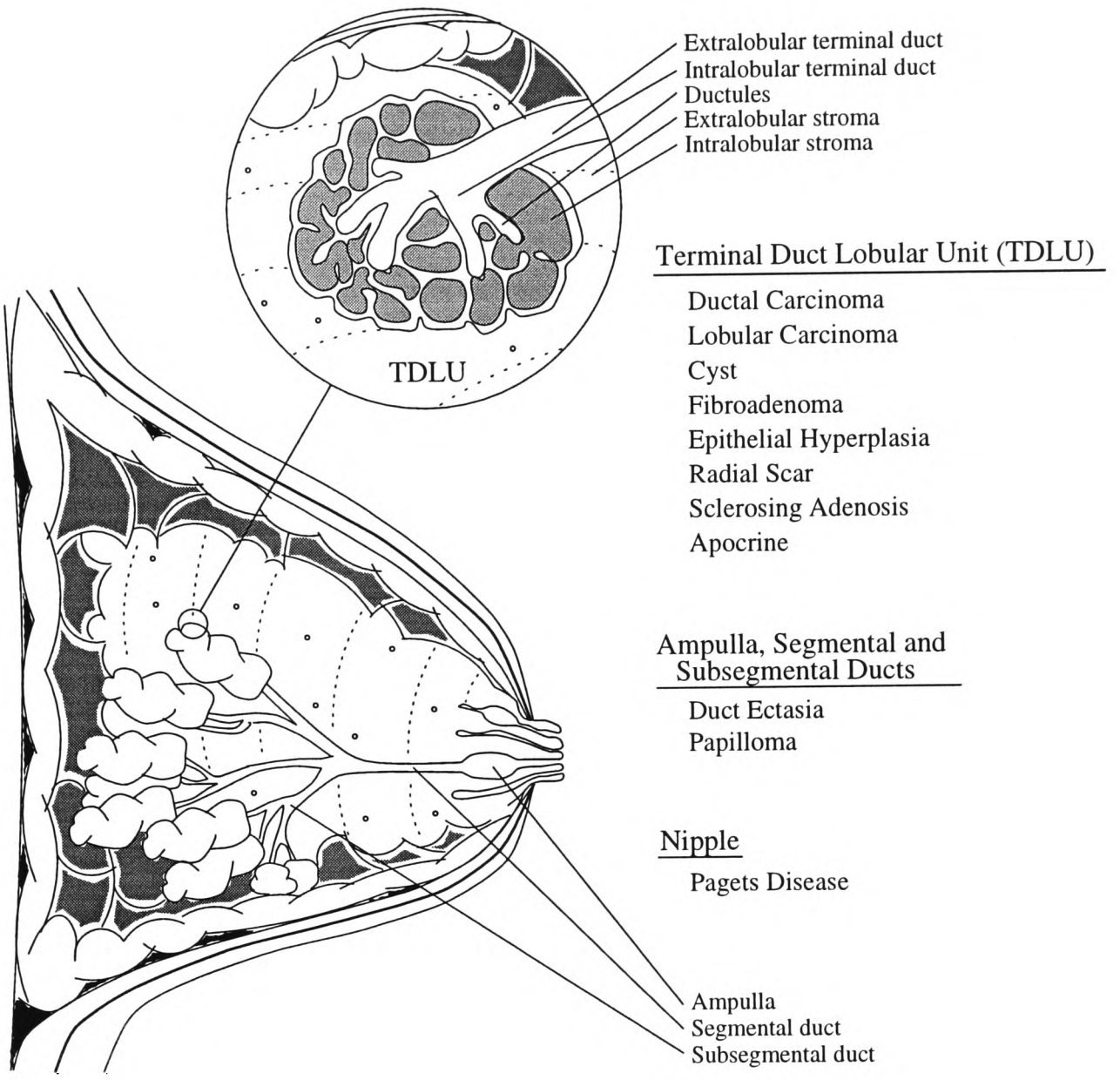


Figure 2.2: Various breast diseases and their most common site of origin.

2.2 Breast Oncology

Abnormalities of the breast broadly fall into two types—benign and malignant. Figure 2.2 displays the various breast diseases and their most common locations.

2.2.1 Malignant abnormalities

It is believed that the majority of breast cancers develop as carcinomas in the TDLU [6, 30, 221]. The process is believed to begin with epithelial hyperplasia (sometimes called epitheliosis or papillomatosis) where the epithelial cells increase in number. This can then lead to atypical hyperplasia where the cells in addition to increasing in number, also change in a way that is not normal for them. At this stage the process is believed to be naturally reversible [6]. The next step is carcinoma *in situ* where the carcinoma has fully developed, but is still confined to the duct. This is considered to be in an irreversible condition, and at some time it will develop into infiltrating or invasive carcinoma, in which the cancer cells break out of the ductal walls and invade surrounding tissues.

There are differing opinions regarding the classification of the many disease processes (e.g. [6, 30, 221]) however we choose to present them according to the classification of Wellings [30, 221], wherein breast carcinomas fall into two broad categories: *in situ* and invasive carcinoma.

In situ carcinoma

The cells involved in *in situ* carcinoma have already developed into a malignant process, however the mechanical structure of the confining membrane has not yet been compromised, and accordingly there is no evidence of stromal invasion. It is at this stage that full biopsy of the *in situ* carcinoma can radically improve the prognosis of a patient. On the other hand, if allowed to remain in the breast the contained carcinoma will eventually metastasise, at which point the invasive carcinoma is much harder to successfully treat. Commonly *in situ* carcinomas arise in two cases: those of *ductal* and *lobular* carcinomas *in situ*.

Ductal carcinoma *in situ* (DCIS) can involve a variable number of ducts and ductal structures, may be solid, and often involve central necrosis (commonly called *comedo* carcinoma), but are always contained within the epithelium of the ductal walls, and thus within the ducts. Occasionally the carcinoma cells of a deep DCIS can migrate up the epidermis towards the nipple pushing the nipple epithelium ahead of it, resulting in an eczematous rash involving the nipple and areola. This condition, known as Paget's disease, is an indicator of DCIS.

Lobular carcinoma *in situ* (LCIS) describes an epithelial proliferation within the lobule acini which completely fills the cavity of the acini. The acini swell to accommodate the growth, effectively reducing the intralobular stroma, however the membranes remain intact, containing the growth.

Invasive carcinoma

DCIS and LCIS will eventually develop into invasive carcinomas by compromising their containing membranes. Upon this development the carcinoma cells are free to invade the surrounding stroma, and via the bloodstream and the lymph-network, are carried to sites further afield. It is the metastasis to sites such as the lungs, bones, liver, etc., and the subsequent secondary cancers and then failure of these organs, that eventually results in the death of a woman.

Invasive carcinomas commonly present as either ductal (more than 50% of cases), infiltrating lobular (about 10%), tubular, medullary, mucoid or papillary ($< 1\%$).

2.2.2 Benign abnormalities

There are many varieties of benign abnormalities which can be manifest in the breast. A summary of the most common conditions is given below.

Fibrocystic change

These lesions may be evident physically and/or mammographically, however some are only evident histopathologically as part of a disease process and are discovered as an incidental finding upon biopsy. These changes can produce masses, calcifications, or prominent ductal patterns.

Adenosis is the enlargement and/or development of new lobular units. It may be considered 'normal' change, as with the 'adenosis of pregnancy', or pathologic when involved with other processes. It is not considered premalignant.

Sclerosing adenosis is adenosis with sclerosing of the intralobular stroma and can present calcifications or masses. It is not considered premalignant.

Fibrosis is the formation of fibrous tissue stemming from the connective and supportive stroma. It is a benign condition.

Cysts can occur in the TDLU when the extralobular terminal duct becomes blocked and the normal ductal secretions begin to accumulate. Their presence alone does not increase the risk of subsequent carcinoma. However, although very rare, an intracystic carcinoma or papilloma can arise from the epithelium of the wall of a cyst.

Apocrine metaplasia describes the change of epithelial cells such that they exhibit the characteristics of apocrine sweat glands. It does not indicate breast cancer.

Radial scar or infiltrating epitheliosis is a lesion with a central fibrous core with radiating arms made up of benign epithelial growth and sclerosis. It can mimic cancer both mammographically and histologically. It is unknown if this lesion signals premalignancy (see the 'radial scar' entry in the next section).

Epithelial hyperplasia is the excessive growth of epithelial cells above their normal level. In extreme cases this can be considered premalignant.

Duct ectasia is a benign process of widened ducts, containing thickened material, often accompanied by inflammation.

Lesions Presenting as Benign Masses

There are many conditions which occur as benign masses and include:

Lipoma is a fatty tumour with no epithelial component. Mammographically it is benign-appearing, radiolucent, well-circumscribed and encapsulated. It can feel soft and easily movable, but occasionally can physically resemble a carcinoma.

Hamartoma is an island of glandular tissue separated from the normal ductal structures. Mammographically it appears as a well-encapsulated smoothly outlined mass.

Feature	Benign	Malignant
Shape	Round, ring-like	Varying shapes
Density	Same density	Varying densities
Distribution	Scattered; benign Ca^{++} can also be clustered	Clustered
Definition	Well-defined borders	Poorly defined borders
Unilateral or bilateral	If the same type of calcifications occur in both breasts, they are more likely benign	Unilateral
Surrounding Tissue	If the calcium is seen within a benign-appearing mass or if the calcium appears normal, this is a benign indicator	If there is architectural or parenchymal distortion associated with the calcium, malignancy must be considered
Increasing in number from prior mammogram	Can be benign	Not an indicator alone, but when considered with other characteristics can indicate malignancy
Size	Can be large or small	Most often small

Figure 2.3: Mammographic characteristics of calcifications, reproduced from [6]

Fibroadenoma is a benign overgrowth of the fibrous tissue of the lobule. It contains epithelial tissue, responds to cyclical hormonal changes and can characteristically calcify. Mammographically it appears an oval well-circumscribed mass.

Papilloma is an epithelial growth attached to the wall of larger ducts with a connective tissue stalk. Most are too small to be seen in a mammogram (except by ductography).

Radial scar present as spiculated features with radiolucent centres. Mammographically they display many similarities with carcinoma. Despite their usual large size they are rarely palpable. There is considerable concern regarding the classification of radial scar as a benign process, and consequently many radiologists recommend biopsy as a matter of course.

Haematoma (Contusion) can occur following injury or surgery resulting in internal bleeding within the breast. The accumulation of blood can appear mammographically as a mass density. These lesions are round, well-circumscribed and more dense than the surrounding glandular structures. During recovery, a haematoma can appear dense and spiculated and can often calcify, imitating carcinoma.

2.2.3 Calcifications

Calcifications can occur in both benign and malignant breast disease, or as a totally normal occurrence. They result from secretions within the ductal structures that have become thickened and dried, or as a result of necrotic processes. It is estimated that microcalcification is the only sign of breast disease in 17–42% of cases [176] and that 30–50% of breast cancers have microcalcifications [13]. Figure 2.3 tabulates some of the mammographic signs of both benign and malignant calcifications.

Certain types of calcifications are almost always benign and these are:

Popcorn-type calcifications are large, thick, dense popcorn-shaped calcifications resulting from involuting fibroadenomas and occasionally other benign processes.

Rim calcifications occur along the border of a benign mass.

Milk of calcium calcifications occur in microcysts containing radio-opaque particles mixed with the fluid. They often appear faint, ill-defined and smudgy.

Arterial calcifications result from arterial atherosclerosis and are seen within an easily identifiable blood vessel.

Skin calcifications are typically round in shape and have radiolucent centres.

Although calcifications are such an important indicator of breast disease, there is also much confusion due to the vague demarcation between benign and malignant calcifications (see figure 2.3). When a radiologist assess a mammogram the problem naturally subdivides into the identification and then classification of calcifications.

2.3 Techniques for detecting breast abnormalities

There have been many techniques proposed and investigated for the detection of breast cancers. Kopans [125] gives a review of many techniques and concludes that:

> "X-ray mammography is the only imaging method currently available with any proven efficacy for screening to detect early-stage, clinically occult breast cancer. Sonography has a limited role in the differentiation of cystic from solid masses and as a guide for aspiration and preoperative localisation of selected breast lesions. Computed tomography has a more limited role to determine the spatial orientation of a lesion detected only in the lateral mammographic position. All other imaging methods should be considered experimental."

Although Kopans paper is now a decade old, the respective imaging technologies have not developed sufficiently to warrant a review of this assessment, except perhaps the enthusiasm for computer tomography would, in the light of technological developments, be redirected to one of the magnetic resonance imaging techniques (see below). Consequently all over the world, both the breast screening programmes currently in place, and those being introduced, are founded upon mammography, and it is on this basis that the focus of the present research is restricted to mammography.

Clearly, detection of abnormalitites is a necessary but not sufficient condition for the diagnosis of breast disease, and as previously mentioned, it it the *early* detection of disease that is of paramount importance for reducing mortality. In the next section we review the techniques available for the detection of breast disease, beginning with the clinical indicators available by simple visual inspection and palpation. This is then followed by a review of the major imaging technologies proposed or in use today to enable the detection of early-stage disease.

2.3.1 Clinical indicators

The clinical indicators of breast disease are as varied and complex as the manifestations of the disease itself. Diagnosis between the possibilities is made more difficult by the subtle (and

sometimes non-existent) variations in those indicators. Some of the signs and their possible interpretations are:

Lump in the Breast A palpable lump in the breast is a classic sign of breast disease. A tender, red and 'hot' lump is generally an abscess, whilst a hard and ill-defined one may indicate cancer or a benign cyst. A well-defined soft lump is usually associated with benign disease.

The location of the lump is important, particularly if it occurs in the armpit, in which case it is usually an axillary lymph node, hardening in response to the collection of lymph being drained from an invasive cancer somewhere in the breast.

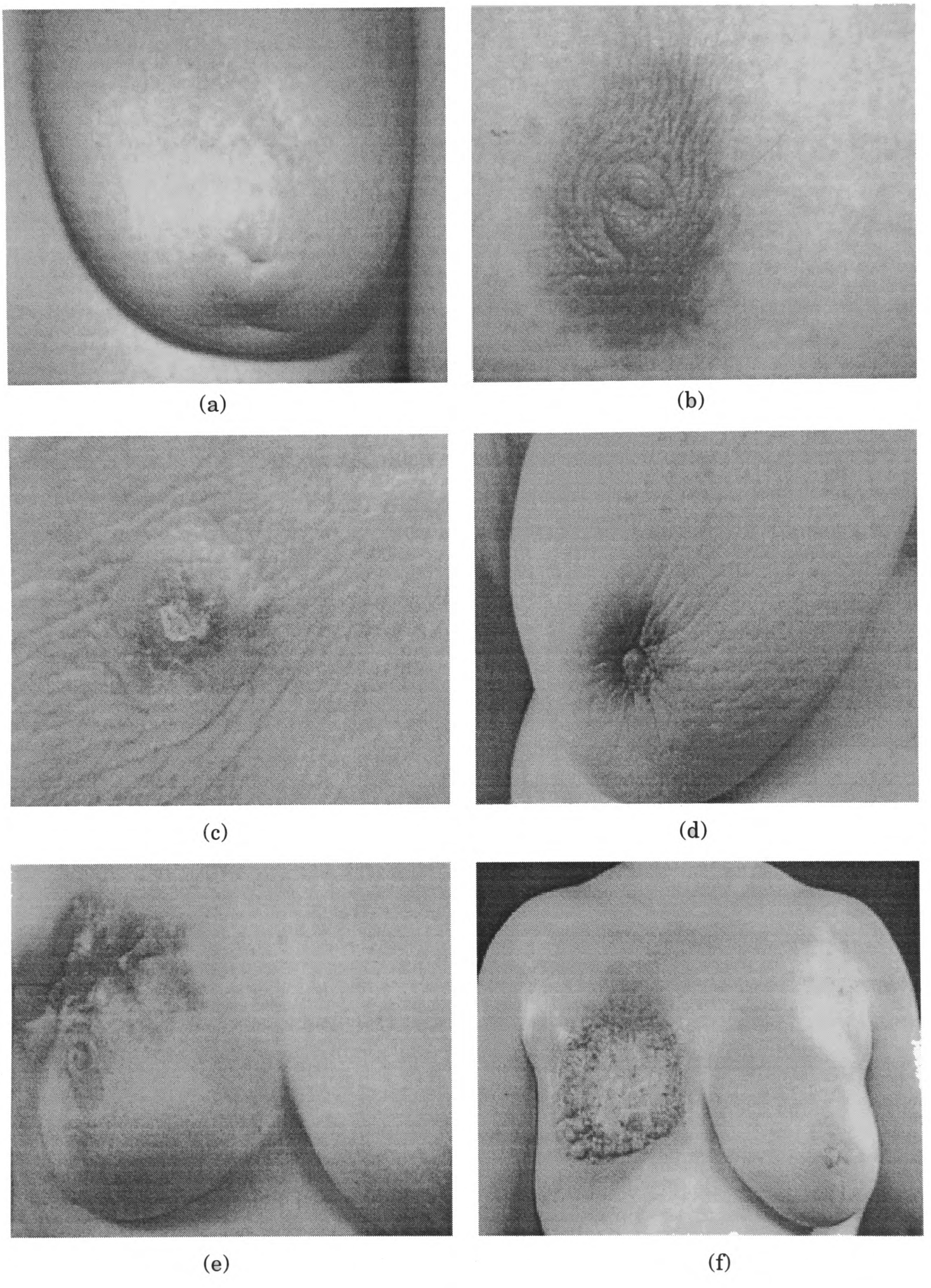
The presence of a palpable lump can be a very grave sign indeed. A $10mm$ tumour has between 10^8 and 10^9 cells. By the time tumours have grown to $20mm$ in diameter they have often metastasised, and for these patients the median survival time is 2 years. The average size palpable tumour detected by a women by palpation is between 20–$30mm$, by which time 50–60% have axillary lymph node activity, giving a 5-year survival rate of 60%. If the tumour is caught at an early stage, that is before metastasis, the equivalent survival rate is 95%. Clearly then the *early* detection of cancer is of paramount importance. Therefore the detection of breast cancer from the perspective of national health must involve methods other than simple palpation, although obviously at the personal patient level, palpation must still be considered.

Nipple and Skin disorders Visual examination of the breast can reveal many indicators of breast disease. Skin tether and nipple indrawing (figure 2.4(a)) are generally associated with underlying carcinomas. The tether results from the strain induced in the stroma by the growth of an underlying lump, whilst a carcinoma beneath the nipple may draw both the nipple and the areola inwards. Sometimes an inflammation of the ducts beneath the nipple can cause them to shorten and pull the nipple in, however this is distinguishable from the malignant case as only the centre of the nipple itself is drawn in (figure 2.4(b)).

Carcinoma of the ducts underlying the nipple—*Paget's disease* of the nipple—causes the nipple to become eczematous with a red, raw appearance (figure 2.4(c)). Widespread infiltration of cancer throughout the breast causes a buildup of excess fluid in the overlying skin, causing it to thicken and take the appearance of the skin of an orange (figure 2.4(d)). This is known as *Peau d'orange*.

A woman may complain of a fluid discharge from the nipple. Discharge from multiple ducts in both breasts is an indication of simple oversecretion and does not indicate disease. Single duct discharge in the absence of a palpable lump can be due to very small *in situ* carcinomas (in approximately 5% of these cases [30]), however it is more commonly caused by duct ectasia or duct papillomas.

Skin Ulceration Visually the most horrific, and often mephitic indication of breast cancer is an ulceration of the skin overlying a malignant growth (figure 2.4(e–f)). This is a very advanced cancer and the prognosis is almost certainly terminal.



(a) (b) (c) (d) (e) (f)

Figure 2.4: Visual indicators of breast disease: (a) Skin tether and nipple retraction, (b) Nipple retraction due to benign duct shortening, (c) Paget's disease of the nipple, (d) *Peau d'orange*, (e) and (f) Skin ulceration due to advanced carcinoma.

Inflammation of the breast is usually in response to an acute bacterial infection within the duct system and is generally confined to the late stages of pregnancy and early stages of lactation. An abscess can form, however this is almost always benign.

In general these techniques of simple examination can indicate the presence of an abnormal process, however they certainly cannot exclude the possibility of one.

2.3.2 Imaging techniques

Mammography is the process of taking an image of the breast by passing a diverging beam of ionising X-rays through the breast and recording the transmitted X-ray photons. The record is taken by allowing the X-ray photons to strike an intensifying screen which absorbs the X-rays and radiates visible light photons which are then recorded on a photographic film. A schematic arrangement of this equipment is shown in figure 2.5. There has been much debate in the literature concerning the harmful risks associated with the ionising radiation used in mammography. Some authors have reported a net increase in the incidence of breast cancer as a consequence of the introduction of screening programs based on mammography [10, 155]. Feig [62] however assess this increase to be at best theoretical, concluding that modern mammography induces 1 excess breast cancer case per year per 2 million women examined. Other factors fuelling the debate include the reported decrease in mortality from breast cancer and its related diseases with the introduction of the same screening programs [155, 207], and more significantly, a net increase in the survival rate [148, 153] (a truer measure of mammography's impact). Mammography has proven very successful in the detection of early, and particularly non-palpable lesions [4, 91, 191], and thus it appears, that although the technique may actually induce cancer in some patients, its ability to detect at an *early* stage both the so induced cancers, and more significantly, the naturally occurring cancers, means that they can be treated before metastasis, resulting in the reduction in mortality observed [155].

It is the ability of mammography to reduce mortality that has driven its implementation as the primary screening technique, and consequently it is this technique which forms the basis of the present work. We defer a more detailed description of the technique until section 2.4, following the review of alternative techniques for detecting breast cancers given below.

Mono-energetic mammography Beccherle et al. [11] have recently reported the ability to image phantoms and mastectomy samples to the same or better image quality than conventional mammography using 50% less radiation by using a mono-energetic synchrotron radiation beam at 18keV. They reason that the reduction in radiation dose necessary for imaging may be due to both an optimisation of the beam energy with respect to the sample thickness and a reduction in the Compton scattering contribution. This technology may one day provide an important development for mammography, although currently it is in its infancy due to

the general unavailability of mono-energetic X-ray production techniques (as a consequence of their extreme cost, synchrotrons are only available in the largest teaching hospitals). Significantly however, the analysis of mammograms acquired using mono-energetic beams can be currently undertaken using computer simulations of mono-energetic beams calculated from the normal mammograms acquired using poly-energetic beams. Highnam [97] describes such a technique, reporting that the expected higher contrast and greater feature detail is realised in practice [98]. Additionally this technique naturally accommodates dynamic adjustment of the (simulated) exposure, allowing an optimal choice of beam energy to be made.

Xeromammography is very similar to mammography given that they are both X-ray imaging techniques. In this arrangement however, the X-ray beam emerging from the breast tissue strikes a charged selenium plate, forming a latent image by photoconductive discharge. The plate is then dusted with a charged cloud of toner and a contact image formed on paper. Either negative or positive images can be produced depending on the respective charges of the toner cloud. Due to the distortion of the selenium plate's electric field in the regions above the boundaries of features in the image, the deposition of toner particles is biased towards the feature, depleting the toner in the region immediately beyond the boundary (and therefore depositing it within the region of the feature). This results in a significant 'edge-sharpening' effect that may or may not be considered a desirable feature, particularly when assessing lesion contours for malignancy as this effect can bias the assessment of the lesion's boundary roughness (an indicator of malignancy).

This technique has found diminishing application with the comparatively massive advances made in the film/screen technologies used in mammography.

(Ultra)sonography forms images of the internal structure of the breast by measuring and mapping the sonic echo amplitude when the breast is exposed to high-frequency sound. The technique utilises the variations in sonic reflection/absorption of the different tissues to distinguish between them. In general, this procedure is not useful for screening since it is incapable of detecting microcalcifications and other early signs of abnormality [9, 193, 194]. It does however provide an indispensable adjunct to mammography for distinguishing between cystic and solid masses, both palpable and mammographically visible masses [9, 30, 163].

A dozen years ago, Burns et al. [25] used continuous wave Doppler ultrasound techniques to show that although there is a significant inter-patient variability, the majority of patients with benign disease were found to have no significant asymmetry of Doppler-detected vascularity. In contrast, patients with cancer, were found to have an increased number of vessels in the diseased breast. More recent work in continuous wave Doppler-sonography suggests a high correlation between abnormal signals and a poor prognosis [183].

Recent technological advances have given rise to colour Doppler ultrasound, giving improved detection of small vessels with flow rates as low as $2mm/s$ [94]. However due to the huge

variation in results and reported performances from this technique [3, 54, 55, 139], its initially suggested application for the discrimination of malignant from benign tissues (by more accurately measuring tumour vascularity) has not been realised.

The general conclusion of the literature is that sonography of the breast (in all its forms) is a very important and useful adjunct to mammography, however it can in no way replace mammography as a first line-of-defence screening tool.

Digital luminescence radiography, DLR or Computed Radiography (CR) has been applied to most projection radiography applications. A projective image is formed by X-ray irradiation of the compressed breast as in mammography, however the image is recorded by X-ray activation of a europium doped barium halogenide crystal film on a polyester base film carrier. The exposed CR plate can store such an image for up to 24 hours. The recovery of the information is achieved by laser scanning the crystals, which luminesce with an intensity proportional to the original X-ray intensity. The visible radiation is then imaged (CCD or photomultiplier tube) and digitised to recover a useable digital image.

The major advantages of the system over standard mammography concern the wide dynamic film-density range, the improved linearity of the 'film'-gradient, wider exposure ranges and direct digital image acquisition. These properties allow CR to improve the image contrast due to slight density differences in soft tissue, small fibrous strands and low-contrast calcifications. However the low spatial resolution of CR ($\approx$5–10 line-pairs/mm) has precluded its general introduction into breast analysis (cf. mammography at $\approx$18–20 line-pairs/mm).

Prior et al. [169] describe a magnification procedure, DIMA^{TM}, utilising a micro-focus X-ray tube coupled with specialised electron beam focussing technology to produce a magnification CR system with very little distortion, The system is capable of resolving 12.4 line-pairs/mm at 4 times magnification, and at least 22.4 line-pairs/mm at 8 times magnification. The geometry of the equipment allows for the significant reduction in scattered radiation exposing the CR plate, allowing for a reduction in overall radiation dose to the patient. Although this is still a research modality there is considerable interest in the technique due to the anticipated prospects of direct digital mammography.

Magnetic resonance imaging, MRI comes in two flavours: plain (or conventional) MRI, and contrast-enhanced MRI. As planar tomographic tools, the data returned is a true three-dimensional representation of the tissue, formed without mechanical distortion, but involving relatively large time investment for image acquisition.

MRI acquires an image by measuring the (typically) proton density in a sample. This value is inferred from a measurement of the energy absorbed from stimulating pulsed RF radiation at the resonant precession frequency of the proton nuclei about an imposed magnetic flux axis. By introducing a flux gradient in the pulsed RF stimulation the temporal axis of the absorption signals can return the spatial location of the response, and so give rise to the

two-dimensional image for the tissue-plane being investigated. By repeating the procedure at a number of parallel planes, the images of each slice can be assembled into the three-dimensional 'image' of the sample. Fortunately, the image contrast is not only a function of the variation in proton density for different tissue types, but also the very response of a proton from one tissue type is slightly different to that from another type, allowing greater differentiation between the tissue types. This gives an image with improved image contrast across the various tissues of the breast, and thus can be used as a tool to aid the detection of disease.

Plain MRI can show excellent contrast between densities and surrounding fatty tissues [2, 46, 70], however trials continue to show the difficulty of differentiating carcinoma from many benign breast changes [92, 179, 234]. Heywang-Köbrunner [94] concludes that apart from the diagnosis of fibrous fibroadenomas there is no widely accepted role for plain MR imaging of the breast.

Improved image contrast has been achieved by contrast-enhanced MR imaging utilising the selective uptake of contrast agents such as gadolinium diethylene triamine pentaacetic acid (Gd-DTPA). Imaging can be undertaken as either a steady state uptake image, or by assessing the *rate* of contrast agent uptake. Heywang et al. [93] have shown that assessing the dynamic uptake of Gd-DTPA improves the differentiation between benign and malignant lesions in the breast. A sequence of FLASH (fast low-angle shot) or GRASS (gradient acquisition in steady state) images are captured to plot the course of the contrast-agent uptake by the tissues. As a consequence, probably [94], of their increased vascularity and vascular permeability, malignancies tend to have a greater uptake of the contrast agent and are therefore selectively enhanced in the images.

In comparative studies against mammography the literature concludes, without exception, that at this time MR imaging, both plain and contrast-enhanced, does not have the capacity to adequately delineate tissue types for detecting breast cancer. Monsees [151] goes further to say, that along with transillumination and thermography, MRI cannot "be considered a useful adjunct to mammography because they do not dictate the next step in patient care". However, in cases where a mammographic mass has proved negative on biopsy or fine-needle aspiration, Lewis-Jones [134] asserts that contrast-enhanced MRI does have a role to play in the assessment of suspected recurrent carcinoma. Further to this, Hickman et al. [95] have used dynamic Gd-DTPA enhanced MRI to assess indeterminate masses. The technique shows good differentiation between normal tissues associated with the architectural distortion and scar tissues resulting from previous surgical intervention that can mimic malignancy on mammography, and the abnormal tissues associated with disease. In addition to giving good results in the classification of known lesions subsequently proved on pathology, the technique proved capable of detecting a $4mm$ lesion which was not visible on mammography or static Gd-DTPA MRI. Although these results are extremely encouraging, the work is still experimental.

Despite the encouraging results obtained in specific instances from the various forms of MRI, there is still much development required to achieve sufficient sensitivity, specificity and robustness for their use in general service. Thus we must consider that within the context of breast analysis these techniques are still research modalities, whilst acknowledging that they can, in certain instances, provide a radiologist with useful supplementary information.

Magnetic resonance spectroscopy, MRS is a spectroscopic modality based on the MR principle outline above. Typically, researchers have analysed ^{31}P spectra from breast tissues since this gives an indication of a cell's energy which is stored in the form of high-energy phosphate bonds, either as adenosine triphosphate (ATP) or creatine phosphate. In general, high-energy phosphates are not present in fat cells, and some studies have shown that increases in high-energy phosphates, changes of some phosphate ratios and the presence of phosphomonoesters can be correlated with malignancy. Thus the technique tries to detect localised areas with high concentrations of phosphates to pinpoint malignancies. The technique is generally hampered however, by the fact that the signal strength from MR of the ^{31}P nucleus is very weak at only $\approx$7% of that from the ^{1}H nucleus. Consequently some researchers have attempted to use high resolution ^{1}H MRS to analyse the water and lipid components of breast tissue [154, 18] in the search for malignancy. Although many encouraging results have been obtained, these techniques have not yet proven themselves, and at this time cannot be considered as more than research modalities.

Transillumination has received a great deal of attention over the years, with each new wave of enthusiasm initiated by some technological breakthrough. To date however, these techniques have not proven their efficacy for detecting abnormalities let alone drawing distinctions between the various classes.

In its simplest and most clinically deployed form, a hand-held wand is passed over the breast surface, emitting alternate pulses of visible red and infrared photons. As the light passes through the tissues it is reflected, scattered and absorbed, and the light returning to the wand is captured to form a dynamic image in real time. It is the interaction of the two wavelengths in the tissues, the intensity of the absorption, the shape of the absorptive areas, variations in vascularity and chromatic absorption variations that provide diagnostic information. Although this is a non-invasive, non-ionising technique with a sensitivity equalling that of mammography, it is time intensive and produces a high false-positive result [30, p55–56]. Sickles [190] notes that this procedure is unable to detect small early breast cancer.

An alternative imaging modality forms an image from the photons that traverse the entire breast section (in much the same sense as conventional mammography forms an image from X-rays that have traversed the breast section). The diagnostic information is provided by the relative interaction of the photons with the breast tissue. The technique is hampered by the extremely low photon counts emerging from the breast — a very small fraction of 1% of the incident light energy emerges from the breast [125]. Additionally, the majority of the

light that does emerge is mostly scattered radiation [192], overwhelming the ability of the system to resolve spatial detail. Sickles [192] noted in 1985 that with the advent of laser radiation, the emergence of "other imaging techniques to neutralise the deleterious effects of scatter may substantially improve the quality of transillumination images". In separate work Alfano and colleagues [45, 219] and Berg et al. [12, 5] employ pulsed laser radiation ($8ps$ pulse width) for illumination, and synchronised optical Kerr-gate shutters to select only the first few photons from each laser pulse emerging from the breast for counting. By rejecting all other photons this avoids including the scattered photons which have taken longer to arrive at the detector due to their increased path-length. Initially the technique could resolve features approaching $0.2mm$ when imaged through $3.5mm$ of breast tissue. More recently Das et al. [45] have been able to detect a $2.5mm$ thick strip of fat within $40mm$ thick tissue to millimetre resolution.

In further research Liu et al. [138] have used $100fs$ pulses to give sub-millimetre resolution from tissue samples $6.5mm$ thick. Wang et al. [218] have used a pico-second double-stage Kerr-gate system that results in 3-orders-of-magnitude improvement in SNR and three-fold improvement in shutter speed over single Kerr-gate. In still further developments Wang et al. [217] have implemented a Kerr-Fourier imaging system to resolve features as small as $250\mu m$ in $55mm$ thick breast-(optically)-like material at a signal level of $\approx 10^{-10}$ of the illumination intensity.

Although this advanced work is still experimental, and a very long way from becoming a clinically indispensable tool, there is every reason to believe that further advances will be made that may, one day, enable the demonstration of its efficacy for the detection of breast abnormalities. For the present however, clinical applications of transillumination remain very limited.

Computed tomography Early research into the use of CT imaging for breast cancer reported favourably its ability to detect cancers that are occult to other methods [37, 36]. This modality was further supported at the time by its ability to resolve the three-dimensional location of the lesion within the breast [126, 125]. This enthusiasm was spurred on by the fact that for most locations in the body, CT scanning is able to resolve density differences much smaller than those demonstrable by conventional X-ray techniques. Unfortunately it turns out that this is not the case for the breast [192]. Furthermore, neither density differences nor lesion size have proven to be useful criterion for indicating a benign state. Controversy surrounds the claims [212] and counter-claims [144] of CT's ability to evaluate lymph node activity, and the matter may never be resolved. Although contrast-enhanced CT provides more information than conventional CT, it is secondary in usefulness to that provided by contrast-enhanced MRI. Other disadvantages of CT scanning include a much higher radiation dose than mammography [36], the necessary radiation of the entire thorax, poor patient tolerance of the radiographic contrast material, the lengthy image acquisition time, and the heavy financial burden of the examination. It comes as little surprise then

that this technique is used very infrequently today for breast imaging, and certainly not as a screening tool.

Thermography measures the surface temperature of the breast and attempts to infer information about the malignancy or otherwise of the tissue beneath, based upon the premise that a cancer produces heat and the conduction and convection of that heat to the surface will produce a detectable temperature change on the surface. Temperature measurement and imaging can be implemented in a variety of forms including infrared, liquid-crystal, computer assisted and microwave thermography. The technique is regarded as ineffective in detecting impalpable breast cancers either as a screening test or in a symptomatic population [30, 151, 200]. Osman and Afify [160, 161] have developed a finite element model of the breast and modelled its thermodynamic characteristics, achieving good agreement with experimental results taken from surface thermograms. Since the interpretation of this information is problematic, the current opinion in the literature [94, 125, 151, 200] is that thermography does not have a role in the detection and diagnosis of breast cancer.

Biomagnetism is indeed one of the newest technologies to have been suggested for detecting breast cancer. Superconducting quantum interference devices (SQUIDs) are used to detect the small magnetic fields in the body. Nordenström [158] has established that breast tumours cause a distortion in the magnetic fields that can be detected by these means. Initial studies have revealed the applicability of these techniques to the right breast, however due to the significant magnetic distortion from the heart, it is uncertain if the left breast will be accessible by these means.

2.3.3 Summary

As was mentioned in chapter 1 it is the *early* detection of cancer that is of utmost importance. Although there are many technologies for breast cancer detection currently under investigation, none currently have the capacity to detect early-stage clinically occult carcinoma with an efficacy anywhere near that of mammography. Consequently the first line of defence in the war against female mortality due to breast cancer is, the world over, based on mammography.

Since mammography is the only technique with proven efficacy for detecting breast cancer, it is within this modality that the scope of the present work is rooted. By applying the tools and techniques of image analysis and artificial intelligence to mammogram images, the current work (developed in subsequent chapters) aims to further the capabilities of this modality.

Before moving on to describe the automated techniques developed for the analysis of mammogram images, both in the present work and in the literature in general, it is useful to consider the starting point for this analysis, that is the mammogram itself, and how it is formed. From a thorough understanding of the image forming process it is possible to develop strategies for the automated analysis of such images based upon hard facts rather than ad hoc intuition.

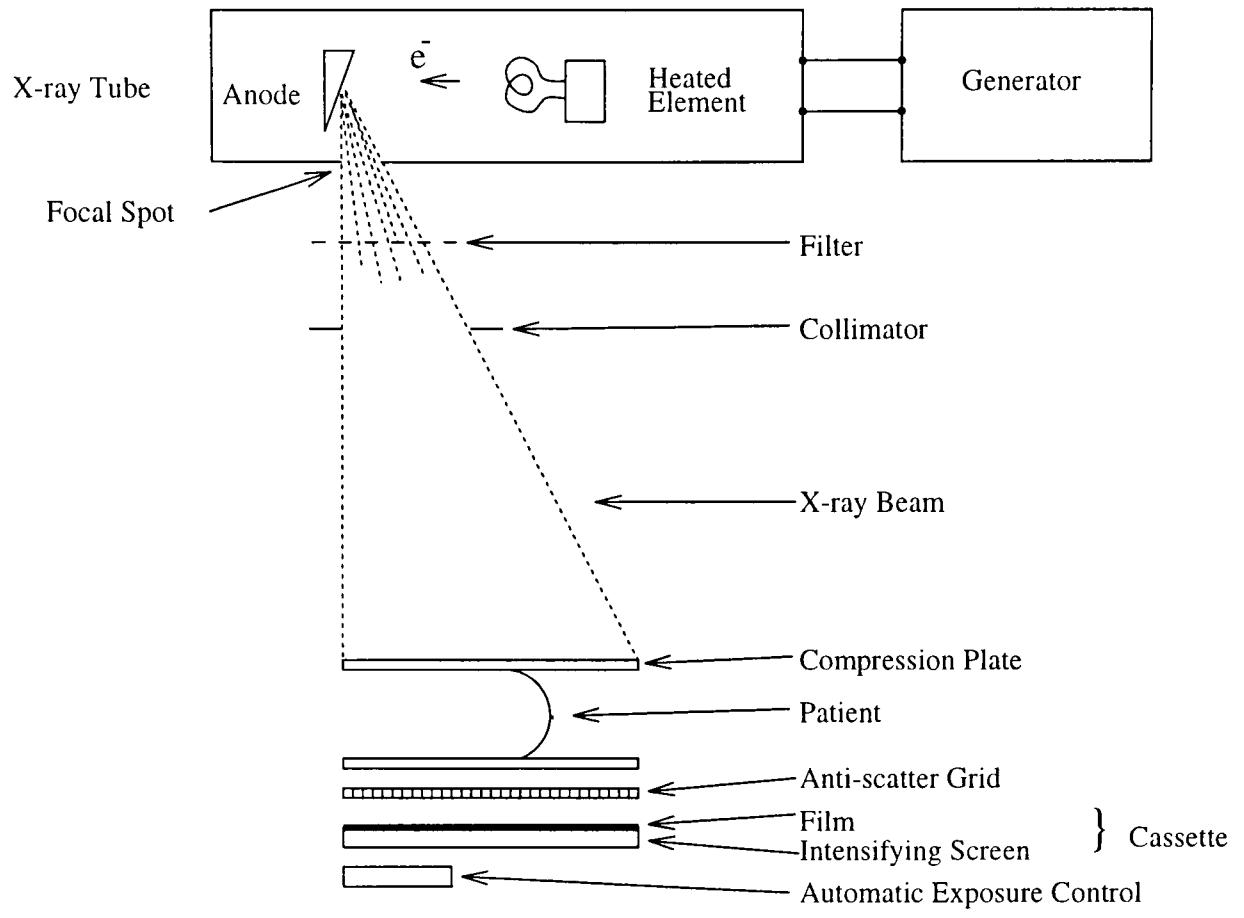


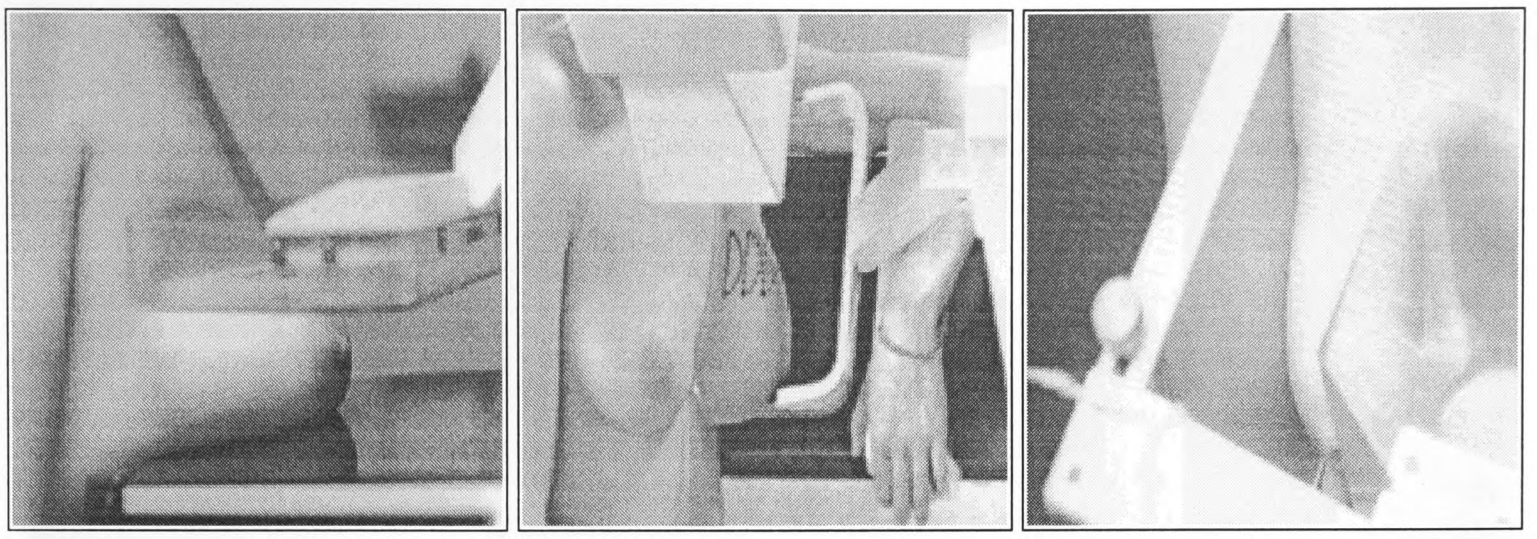
Figure 2.5: Schematic representation of mammographic system.

2.4 The Mammographic Process

A mammogram is a photographic film of an X-ray image of the breast, and mammography is the process by which the film was produced. Figure 2.5 gives a schematic representation of the components of a typical film-screen mammographic system. A filtered, collimated, diverging beam of low energy X-rays from a molybdenum anode is directed towards a compressed breast. As the beam traverses the tissue of the breast it is attenuated due to absorption and scattering. After exiting the breast, the beam passes through an anti-scatter grid to minimise scattered photons, and then through the photographic film and onto an intensifying screen which absorbs the X-ray photons and radiates visible light photons which expose the photographic film. An Automatic Exposure Controller (AEC) placed beneath the film-screen cassette terminates the exposure when a set amount of radiation has been received. In this fashion the attenuation characteristics of individual breasts (due to variations in compressed thickness and radiographic density) can be accommodated giving neither under- nor over-exposed films, but rather keeping the exposure within the unsaturated linear range of the film.

There are three usual views for mammograms: cranio-caudal (figure 2.6(a)), 45° mediolateral oblique (figure 2.6(b)), and the mediolateral view (figure 2.6(c)). A typical pair of mediolateral oblique images (the standard UK screening view) is shown in figure 2.7.

A large number of variables in this process can degrade the quality of the mammogram, and breast imaging clinics have found the need to institute codes of practice to ensure the quality of the images obtained. Figure 2.8 presents some of the variables which influence the



(a) (b) (c)

Figure 2.6: Typical views of mammography: a patient positioned for (a) cranio-caudal (b) 45° mediolateral oblique (c) mediolateral mammogram.

final appearance of a mammogram and thus make quantitative analysis of a mammogram difficult. In particular note the factors effecting image contrast and intensity, and noise in the film. These processes will prove to be important when attempting to automate the analysis of mammogram films to be discussed in subsequent chapters.

The process itself is quite straightforward and is based upon the same principles as all medical X-ray imaging. It is the interpretation of the mammogram that provides the challenge.

2.5 Mammographic signs of breast disease

Mammographically, the signs of abnormal pathology fall basically into three broad types: (i) calcifications, (ii) masses, and (iii) dispersed global processes of the parenchyma, such as asymmetry and architectural distortion. Other secondary signs of disease are dilated ducts, dilated veins and skin thickening.

As noted earlier, calcifications occur in the breast as a result of secretions within the ductal structures that have become thickened and dried, or as a result of necrotic processes. Due to the high attenuation of the X-ray beam by these inorganic calcium salts, they appear as small regions of very high radiographic density, and consequently high image intensity. Important characteristics of calcifications with regard to malignancy include their shape, density, distribution, definition, size, surrounding tissue, unilateral or bilateral appearance and the temporal changes of these properties. A radiologist will consider all of these factors in conjunction with the clinical and historical information available to make a decision.

Mammographic masses appear with a variety of characteristics indicative of the underlying disease process. In general they are of higher radiographic density than the surrounding glandular or fatty tissues (except in the case of lipoma, a benign process). The tissue composition of the mass is important since regions that appear as glandular or fibrous tissues can be either malignant or benign, whilst fatty tissues (e.g., lipoma) are always benign. Masses

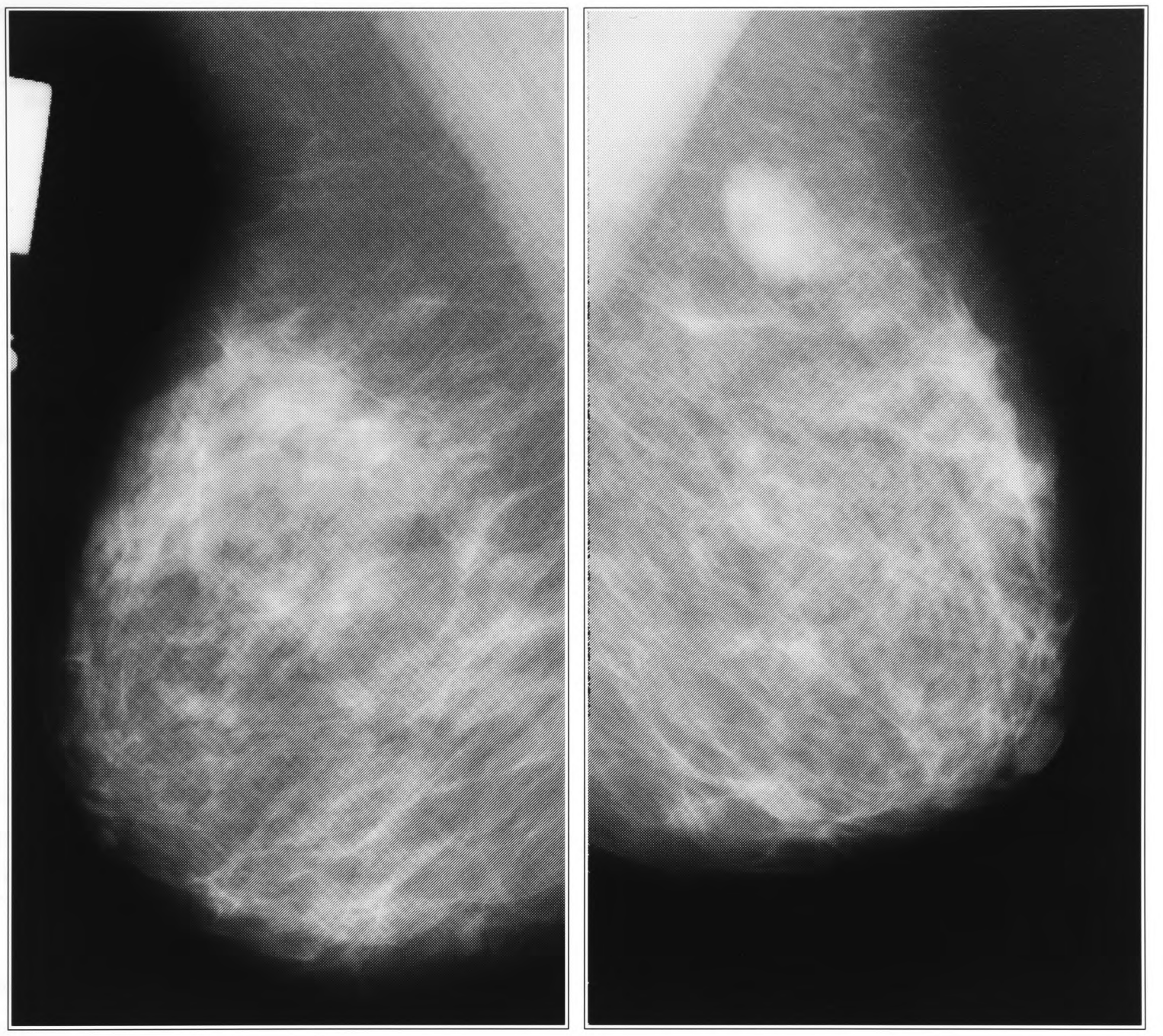


Figure 2.7: A typical mediolateral oblique mammogram pair. This breast pair is not 'typical' in the sense that there is a large difference in bilateral breast size, or that the left breast (right image) has a large cyst, but rather the images are typical *mediolateral oblique* views, with the pectoral muscle visible at the top centre of the figure, and the breasts taking the overall appearance as shown in these images.

with spiculated borders (rough borders with many linear structures extending radially from the mass into the surrounding tissues) are strongly implicated in malignancy, whilst smooth circumscribed masses normally indicate a benign process. Alternatively, the appearance of linear structures which can be traced right through a mass, that is they enter at one point and can be followed until they emerge at some other point, are not indicative of malignancy, since this simply represents overlapping structures resulting from the projective nature of the mammography system. If a mass is truly spiculated, the structures will disappear towards the centre of the mass.

The left and right breasts form a mirror pair and in general the gross appearance of mammograms reflects this. Involution occurs bilaterally in the same proportions, and consequently

Radiographic Sharpness

Radiographic Contrast	Radiographic Blurring
Subject Contrast	
• Absorption differences in breasts: Thickness, density and atomic number	**Motion Blurring**
• Radiation quality: Target material, tube voltage, filtration	• Breast immobilization (compression), Exposure time
• Scattered radiation: Beam limitation, compression, air gap, grid	**Geometric Blurring**
Receptor Contrast	• Focal spot size, Focal spot–object distance, Object–image receptor distance
• Film Type	**Receptor Blurring**
• Processing: Chemistry, temperature, time, agitation	• Phosphor thickness, Light absorbing dyes and pigments, Phosphor particle size, Screen-film contact
• Photographic density	
• Fog: Storage, safe-light, light leaks	

Radiographic Noise

Radiographic Mottle	Artifacts
Quantum mottle	
• Film speed	**Handling**
• Film contrast	• Crimp marks, fingerprints, scratches, static
• Screen absorption	**Processing**
• Screen conversion efficiency	• Streaks, spots, scratches, dirt, stains
• Light diffusion	**Exposure fog**
• Radiation quality	
Structure mottle	
Receptor mottle	

Figure 2.8: Some of the many factors affecting mammographic image quality. Reproduced from [88].

breasts are expected to have bilaterally equivalent glandular distributions. In cases where one breast is larger than the other (e.g., figure 2.7) the glandular tissue is expected to be in proportionately equal distributions. Uneven distributions of glandular tissue, commonly termed *asymmetry*, is often an indicator of an abnormality, although sometimes it is due to uneven development or surgical alteration. Andolina et al. [6] note that "asymmetry is the radiologist's greatest aid in determining abnormalities both benign and malignant".

Ordinarily the ductal structures converge at the nipple in a series of radial structures centred at the nipple. Disturbance of this 'natural' order is termed *architectural distortion* and is another indicator of disease processes, both benign and malignant. Occasionally this distortion might be the only indication of a mass, although there are normal breast structures such as the blood vessels and Cooper's ligaments that can imitate a distorted duct pattern. This is a difficult sign to interpret accurately, however it is not one to dismiss lightly.

The careful reader would note the rather vague terms in which these last few paragraphs have been couched: 'normally' this, ... 'generally' that, and so on. The difficulty with interpreting mammograms is that there are no hard and fast rules concerning disease and malignancy, and the wealth of knowledge developed through years of experience by the diagnostic radiologist is indispensable for correctly assessing mammograms. Despite the vagueness of these mammographic signs, they are the only real indicators available for detecting early-stage can-

cers, and the real skill of mammogram interpretation is being able to identify these signs when they occur. Consequently it is not the intention of the projected, long-term aim of the current work to find all the answers and provide a fully automated system capable of interpreting mammograms, and therefore dispense with the need for a radiologist. Rather it is hoped, and will be shown, that the application of automated technologies can *aid* the radiologist by providing extra information upon which to base a decision.

2.6 Summary

In this chapter we have reviewed the anatomy of the breast and described the disease processes which afflict the organ. A review of the techniques available and under investigation for the detection of breast disease concludes that mammography is the only imaging modality capable of detecting early-stage clinically occult disease, and consequently it is this technique that is employed the world over in screening programs to combat female mortality due to breast cancer.

A description of the mammogram imaging system and the mammographic signs of breast disease has established the starting point for the analysis and interpretation of mammograms, both in the clinical setting, and the work to be presented on the pages that follow. Noting that the quality of mammographic images is affected by a large range of factors that cannot be completely controlled, it is therefore not possible to assess a mammogram on the basis of absolute image intensities (or film density values). The mammographic signs of breast disease can be classified into the three broad classes: calcifications, masses and dispersed global processes of the parenchyma. So protean are the signs and manifestations of breast disease, that it is imperative an attempt to automate the analysis and interpretation of mammogram images be built on a 'model-based' understanding of these processes. The work presented in this thesis aims for such a model-based identification of mammographically significant features.

3

Automated mammogram interpretation

Researchers began attempting to automate the analysis of mammograms before the very inception of modern mammography, and certainly a long time before the introduction of screening programs placed such heavy demands upon those who would interpret them.

Modern mammography began in France in the mid 1960s pioneered by Charles Gros, M.D. and the CGR Company who developed the first dedicated mammography unit. The equipment was centred around a new X-ray tube with a molybdenum target which had been borrowed from the quality control programs of the automotive tyre manufacturing industries. This new technology was commercially introduced into the United States in 1969, and for the first time, gave the physician high contrast images of the breast to consider.

However, in 1967, two years before the birth of modern-mammographic equipment, Winsberg et al. [222] digitised positive prints of bolus immersion mammograms[1] at 180 pixels/inch ($141\mu m$/pixel) and 5-bit intensity resolution using a converted 'radio-facsimile' scanner. Using a CDC-160A computer[2] they demonstrated the feasibility of this approach for the automatic detection of suspicious areas of the images. From these humble beginnings, each new advance in mammogram imaging, digital image acquisition, computing technologies, and algorithmic analysis has taken another step toward reliable and robust automated mammogram analysis. In this chapter we review the literature on this topic to establish the current state-of-the-art in

[1]This ancient technique involved immersing the breasts in either mineral oil or alcohol whilst the patient lay face down. Radiographs (as they were termed in those days) were then exposed trans-laterally to give an image in a film-screen cassette in much the same fashion as those used today. The images were *very* poor by modern standards.

[2]The CDC-160A, circa 1966, was based around a 12-bit processor with an instruction cycle of $6.4\mu s$ (that's 0.16MHz!), a single accumulator performing 1's complement arithmetic. RAM was limited to a maximum of 8K of magnetic core RAM (in two 4K banks). Without an assembler, programs were entered in octal through the register buttons, and without any mass storage, stored on paper tape. This was before the development of the ASCII character standard and thus the machine used its own 6-bit character encoding scheme. Truly humble beginnings and quite remarkable results considering today's technology and the continuing difficulties of analysing mammograms.

mammogram analysis, and so set the scene for the work developed in the subsequent chapters of this thesis.

The algorithms proposed for analysing mammograms have in general tried to do one of two things: either emulate the functions of the clinician by searching for the same signs that he/she does (see section 2.5), or try to overcome some of the difficulties that confound these attempts to emulate the clinician, such as image noise etc. Indeed this is hardly surprising since the traditional motivation for the investigation of computer-aided mammography methods has come from the medical community and has consequently focussed upon the issues that are familiar to, and deemed important by clinicians. In many ways the ability to display competence at the tasks a clinician undertakes is a necessary feature of any computer based system as this allows the community of clinicians to test such a system directly against their own knowledge and experience. Such interaction enables the clinician to develop trust in the automated system. Additionally, formulating an automated system based upon sound principles, enables a logical explanation of results produced, and is therefore important in developing the trust of the radiology community for those system capabilities that are not directly comparable to their experience. A system *perceived* (not necessarily correctly) to inadequately or inaccurately analyse a mammogram is of no use to the clinician or the community at large. Conversely an inaccurate system perceived to be correct, is downright dangerous and must be avoided.

The characterisation of human emulation for many of these algorithms can be carried further to consider both detection and then diagnosis of abnormality, and accordingly there are approaches reflecting this difference.

Evaluation of proposed algorithms has on the whole been undertaken by comparison of the algorithm's results against an expert radiologist's opinion. In a recent talk given at the SIWDM[3] [181], Dr. Robert Schmidt urged caution with regard to these assessments since *everybody* cannot claim an *expert's* opinion, for that negates the definition of an expert. None-the-less in light of the fact that an absolute 'gold-standard' does not exist[4], it is the opinion of the best available radiologist that must be used as the comparison for evaluation of automated techniques. There is another important consideration when evaluating an algorithm's robustness, and that is to ensure that the algorithm is exposed to a variety of cases that is statistically representative of the population. This point is particularly important in the current situation due to the vague and varied signs of mammographic disease (section 2.5), and it is therefore important to ensure that the algorithm is sufficiently exposed to this variety.

In light of the general mammographic signs of disease and these preliminary comments, consider the broad categorisation of the previous work under the following topics to facilitate its review:

- Calcifications,

[3]Second International Workshop on Digital Mammography, 10–12 July 1994, York UK.

[4]In fact the histology/pathology of the tissue is such a standard, however in the context of interpreting many mammograms and assessing the performance of computer algorithms, this information is not available, and can therefore be regarded as non-existent.

- Mammographic Densities,
- Evaluation of parenchymal patterns,
- Noise,
- Segmentation,
- Bilateral registration and asymmetry,
- Intensity normalisation,
- General image enhancement,
- Other techniques.

3.1 Calcifications

Calcifications are an early indicator of both malignant and benign disease, and in some instances are the only sign of malignancy. Consequently considerable effort has been expended in the search for techniques to automate their detection. The field naturally divides into the two subproblems of initially *detecting* calcifications, and then, due to the high false-positive rates from these detection algorithms, *classifying* the proposed calcifications.

Mammographically calcifications appear as small sharp patches of high intensity, and are therefore often very similar to noise. It is generally agreed however, that given sufficient resolution of the digitisation procedure, calcifications will span a handful of pixels, and be sufficiently spatially extensive to distinguish from noise. Naturally enough, the debate concerning the definition of 'sufficient resolution' has been intense with the desire for greater resolution in the hope of achieving greater accuracy conflicting with technological and practical limitations. Karssemeijer et al. [111] conclude that "$100\mu m$/pixel resolution does not prohibit high-quality diagnostic performance in digital mammography" since "spherical calcifications with diameters smaller that $130\mu m$ are not detectable with film-screen mammography". By contrast, Kimme-Smith et al. [123] state that an ability to detect $250\mu m$ diameter calcifications should be sufficient. This issue is far from resolved, however given that $50\mu m$/pixel resolution is commonly available today, it should be considered that in theory all mammographic calcifications can be represented by some signal in a digital representation. The question thus, is how to detect the signal?

In general, the approach taken has been to detect all manner of signals that possess the characteristics of calcifications and then to implement some higher processing to weed out the spurious signals due to noise and other artifacts. As a result many attempts to enhance the image are made prior to extracting the calcifications, either by highlighting calcifications or suppressing noise, or both.

To generally enhance an image for calcification detection Fox et al. [71] apply a matched filter for background removal, whilst Chan et al. have used used unsharp masking [35], a difference

image technique formed from signal suppressed images (by median, contrast-reversal and band-pass filters) subtracted from signal enhanced images (by matched and box-rim filters) [32, 34] and a filter supposedly emulating the human visual response system [31] (although why this is done is not clear, particularly as the results are then interpreted *by* the human visual system). Their results have been verified by evaluation of the results from processing 20 mammograms [33, 34].

Calcifications have been extracted using the property that they appear as regions of high local contrast. This has been done using local thresholds, edge detectors, local maximums and local roughness measures [1, 48, 49, 61, 131, 197, 220]. After initial detection and to ensure that the entire calcification is selected, Wee et al. [220] and Fox et al. [71] trace the edge looking for a closed periphery, whilst Chan et al. [33] and Fam et al. [61] use region growing.

Dhawan et al. [51, 53, 52] have specifically tuned a contrast enhancement algorithm for the detection of calcifications (see section 3.8 for an example of its implementation). Like many algorithms for general image enhancement, this approach is however, susceptible to the introduction of artifacts (see section 3.8), and since it is proposed that this enhancement form the basis of an *automated* analysis, this technique should be viewed with extreme caution.

Highnam [96] proposed the use of mathematical morphology as a possible technique for extraction of calcifications based upon shape criteria, although noted the difficulties of matching the structuring element shape to calcification shapes. In fact due to this diversity of calcification shapes which present mammographically, Levy-Vehel [133] dismissed classical morphology as it is difficult to define suitable operators. Fredfeldt [72] has however, on the basis of his morphology studies into detecting masses (see section 3.2), proposed that morphological detection of calcifications should be pursued.

Brettle et al. [21] propose matched filters for application in the frequency domain. The filters were developed from measuring the frequency response of simulated microcalcification phantoms imaged and then FFT transformed. Although their results have been encouraging they are limited by the current evaluation extending to 15 images, and only 8 with microcalcifications.

Valatx et al. [215] fit the image surface with a bicubic B-spline, defining the least squares error image which is then locally thresholded to identify microcalcifications. They note that this technique has difficulties with the responses from "parenchymal structures" which are also detected. Whilst acknowledging that these responses do not pose any difficulty if the images are used purely for prompting, the authors note that this area would need to be addressed if a fully automated analysis were required.

Parker et al. [164] propose a novel filter with an output given by the hysteresis limited tracking of an input 1-dimensional signal. Using each line of the image as a 1-dimensional signal in intensity, this filter finds local intensity changes with some sensitivity against noise. Variation of hysteresis alters the sensitivity to signal and noise. The final assessment of the output is undertaken by Kohonen feature maps and k-NN classifiers. The procedure was

evaluated over a combined training/test set of 350 microcalcifications.

Strickland and Hahn [202] and Laine et al. [129] use a wavelet decomposition to give a set of band-pass filtered images. Identification of the microcalcification response in the decomposed signal allows an amplification of the appropriate wavelet transform coefficients, such that upon reconstruction of the image the microcalcification features have been highlighted. Difficulties in detecting the calcification responses in the wavelet decomposed signals have, to date, limited the effectiveness of the technique.

The techniques for proposing calcifications reviewed above have typically been very successful in detecting all valid calcifications, however this has been at the expense of proposing many breast structures and artifacts which are not calcifications, such as blood vessels, sharp-edges, fibrous strands and some ductal patterns. It has therefore been necessary to classify the proposed calcifications. The majority of techniques have tried to match proposed calcifications with the expected size, shape and clustering characteristics of valid calcifications. Various decision making techniques have been used including upper/lower bound criteria [32, 34, 48, 49, 61, 159, 220], Bayesian decision techniques based on stochastic models of cluster patterns [112], and more recently neural classifiers [38, 164]. Without exception, the neural classifier implementations have trained the networks on small data sets and consequently the results have either been limited, or should be treated as preliminary.

Responses from noise have been rejected by cluster analysis as it is highly unlikely that noise will randomly accumulate as clusters. Nishikawa et al. [157] define a clustering procedure that nominates the heuristic "a cluster is a group of three or more signals in a localised area" (they use $3.2 \times 3.2mm$ areas) as being indicative of malignancy. They tested their procedure on 78 mammograms containing 41 clusters, attaining a true-positive classification rate of 85%, and a false-positive rate of 4.2–2.5/image.

Spiesberger [197] tried to apply cross-correlation between the calcifications proposed from two different projections of the same breast. Valid calcifications would be evident in both projections at the corresponding positions, however noise in the system would not. Although analytically sound, variations due to compression and the details of Spiesberger's implementation reduce its effectiveness [97].

In contrast to the methods above which segment and then classify proposed calcifications Karssemeijer [109] defines a stochastic model of both the calcifications and the imaging process that tries to integrate the segmentation and classification phases of calcification detection. Using Bayesian decision theory in a deterministic relaxation framework alongside careful calcification modelling has demonstrated this simple technique's ability to perform the complex recognition task. Karssemeijer found it necessary however, to introduce a third class in to the decision (on top of 'calcification' and 'background') to accommodate the "line-like structures" that would confound the two class solution.

Chitre et al. [38] use a global approach that avoids the need to pre-segment an image into proposed calcifications by analysing second order histogram statistics looking for correlations

with malignancy. Several neural network architectures were evaluated, and although early results have been promising, they have so far been trained on a data set of only 100 cases.

In general the robustness of these algorithms have been assessed by comparison with a radiologist's opinion, however some papers have analysed simulated data for which the correct answers are known a priori, although the validity of the simulation must then be assessed. In addition to the image/case sample space given for some of the techniques above:

- Davies and Dance [49, 48] used 25 images for feature characterisation and 50 images for testing, all containing calcification clusters;
- Chan et al. [35, 34] used 32 images, 12 containing 12 calcification clusters, 20 normal;
- Chan et al. [32, 31] superimposed 10 monte-carlo simulated calcification clusters into each of 6 normal images;
- Olson et al. [159] used 48 images containing 52 clusters and no normal images;
- Fam et al. use 50 [61] and 40 [60] images containing as many calcification clusters and no normal images;
- Wee et al. [220] use 51 calcification clusters, 28 benign and 23 malignant (shown on biopsy), and no normal images;
- Fox et al. [71] use 100 calcification clusters, 54 benign and 46 malignant (shown on biopsy), and no normal images;
- Levy-Vehel. [133] use 40 images, 26 benign and 14 malignant (shown on biopsy), and no normal images.

When compared against the three million mammograms acquired annually in the UK screening program alone, the infrequent appearance of abnormal cases, and the general variability of mammographic signs, there can be little argument as to the statistical insignificance of these examinations. Although these techniques offer many promising avenues to explore, clearly there is much work to be done before the community can declare the problem of automatically identifying calcifications solved.

3.2 Mammographic densities

As with calcifications, detection of masses generally proceeds by proposing all possible regions of suspicion which are then subjected to a series of tests to eliminate false alarms, followed by some higher level classification into benign or malignant masses. The characteristic appearance of mammographic densities that should be proposed as possible masses varies widely, from rounded, sharp edged and circumscribed benign masses through to classical malignant masses with small, bright centres and radial bright structures. Then there is the difficulty of differentiating projected densities that are nothing more than an unfortunate overlapping of ordinary glandular tissues that mammographically resemble masses. Often it is impossible

from the evidence available in a single mammogram to determine if a density is a real mass or if it is merely due to overlapping structures. In some such cases the radiologist would normally request another mammogram at an alternative projection, a facility not normally available to an automated approach. Thus in this situation it is normal for the automated approach to aim to report such overlapping structures as masses, and accordingly the detection of masses is generally a case of initially detecting such potential densities.

Brightness based assessment, utilising the observation that most masses appear as an area of increased brightness, has been a popular method [84, 122, 131, 222, 228]. This technique is however, limited by the relative concept of 'increased brightness' as often the mass is only as bright as surrounding tissues, such as the fibrous tissues of the parenchyma. This is overcome by including tests for other expected features. Occasionally, this approach is supplemented with a region-growing scheme to find the extents of the blob to be proposed [228].

Claridge and Richter [40] show that the spatial extent of simulated lesions extends beyond the zero-crossings of the image surface second derivative, a common technique for marking boundaries in the computer vision community [145]. Consequently they warn against using these methods for finding real lesions in mammograms.

Image intensity roughness is often used as an indicator [84, 122, 184, 222], since it is common for centres of malignant masses to be smooth, whilst benign masses are often rough. Lau and Bischof [131] include a combined brightness-roughness measure to overcome the ambiguity that sometimes results, using the principle that malignant tumours can be of high intensity with low roughness whilst fibroglandular tissues (a benign mass commonly confused with carcinoma) usually have medium-high brightness but always high roughness.

The observed rotational symmetry of masses has been exploited by using shape analysis. Fredfeldt [72] used thresholding to obtain an edge estimate of the tumour which was then subjected to a morphological closing operation, with a resulting open curve depicting malignancy. A purely circular region is deemed to be benign. Obviously the simplicity of this approach belies the complexity of the task at hand, and given that this procedure was tested over images of 30 breasts only, the encouraging results reported by the authors must be viewed as preliminary results.

Lai et al. [128] use a template matching algorithm in the search for circumscribed masses. Templates of various sizes tuned to the expected circular nature and uniform density of tumours are applied. Additional brightness contrast measures are also used. The successful operation of these methods generally relies upon a wise choice of operator or neighbourhood size which is matched to the size of the mass being searched for.

Giger and colleagues [73, 231, 232] have used bilateral subtraction (after image registration) exploiting the nominal bilateral symmetry of healthy parenchyma to label regions of significant difference as potential masses. They develop an elaborate mechanism to overcome the difficulties posed by the significant high-frequency textural components of the parenchyma, such as fibrous tissues and blood/milk vessels which otherwise return significant responses for sub-

traction techniques. Initially each image is thresholded at 10 different levels (at the intensity giving the low 5%, 10% ...50% of histogram area, values above threshold retain their value, those below set to zero) and 10 subtraction images are formed from the respective components. They assert that noise and normal anatomical background variations at a given pixel location will have a response in the 10 images on, relatively speaking, fewer occasions than real masses. Even overlooking the lack of justification for this last assertion, there is nothing to guarantee that the 'background' grey-levels for fatty tissue will be bilaterally comparable, nor that masses won't develop in exactly the same location bilaterally (in which case the subtraction will return a null response). The difficulty here is that this approach differences the grey-level images (for which the digital representations are wildly uncalibrated) and searches for features, rather than searching for features and *then* subtracting the feature-space descriptions looking for asymmetry. It is to this technique that we shall return in section 7.2.

Lau and Bischof [131] also use bilateral asymmetry for detecting masses however they initially normalise the images to account for imaging differences. Alongside the brightness-roughness metric mentioned earlier, they found the need to introduce a 'directionality' metric to account for the "strong responses ... generated in regions with blood vessels or ... glandular tissue".

Glatt et al. [75] use Weighted Majority Minimum Range Median filters and hexagonal image tesselations to find masses. No justification for the filter is given, nor does the explanation of its execution shed any light on why this might be used. It appears to be simply an ad hoc approach. In a comparative analysis of filters, Simpson and Bowyer [195] note the indifferent performance of this filter for noise suppression.

Karssemeijer [110] uses second order differential operators to determine the 'line' orientation of each pixel. The extended neighbourhood surrounding each location is searched for pixels that are part of line processes that intersect the immediately surrounding region under the assumption that these features form part of the spicules associated with a spiculated mass. In initial experiments this technique has shown promising results. Although not noted by Karssemeijer, since this technique does not depend upon the existence of a central mass, it could therefore also be used to identify radial scar processes.

The most important clinical sign for the classification of masses is the condition of the lesion boundary. It is the uncontrolled growth of a malignancy in all directions that gives a tumour its characteristic rough edge and spiculated radials. A sharp circular edge defines a benign lesion whilst anything other than a completely sharp edge for the entire perimeter of the lesion is believed to be a sign of malignancy. Techniques for automatic classification of densities are generally based upon assessing this attribute for each proposed mass.

Richter and Claridge [170, 171] try to derive a measure for the amount of blur on an edge by experimenting with simulated blur. Lai et al. [128] used lesion 'suspiciousness' variation as a function of their template size to measure the lesion blur as the template crossed the lesion border. Fredfeldt's [72] application of morphology to detect a closed boundary is an example of

this also.

Gradient measures across the boundary have been used [48, 49, 73, 84] to determine the 'smoothness' of the boundary, and hence the malignancy.

Measures of circularity using area to perimeter ratios have been used by Ackerman and Gose [1], and tumour circumference to the circumference of a circle of equivalent area by Fredfeldt [72].

Kegelmeyer et al. [118, 120] assess the orientations of edges in a large region, and some texture measures in a local region surrounding each pixel, and then use a binary decision tree (BDT) classifier to directly identify spiculated lesions. Although reporting good sensitivity (97%), their results should be interpreted in light of the fact that the decision tree has only been trained on datasets comprising 36 known spiculated lesions and 49 negative cases, and thus the ability to adequately represent the variation in the population space of lesions is limited, and thus these results should be considered as preliminary. It should also be noted that the testing of this decision tree was evaluated on the same datasets used for training.

Woods and Boywer [228] have used linear and quadratic neural classifiers to analyse the database used by Kegelmeyer et al. mentioned in the previous paragraph. Unlike the pixel based approach of that study, Woods and Boywer initially identify a candidate list of possible masses, and then use the classifier to identify the spiculated lesions using a number of metrics based on the collective pixels of the candidate region. Obviously the concerns regarding statistical significance expressed in the previous paragraph apply equally well here.

As previously mentioned, the robustness of these algorithms have in general been assessed by comparison with a radiologist's opinion. In addition to the image/case sample space given for some of the techniques above:

- Lau and Bischof [131] used 10 image pairs (20 images) containing 13 tumours, successfully identifying 12.
- Lai et al. [128] used 17 images containing 19 tumours, successfully identifying all 19, but also returning 29 false positives.
- Giger and colleagues used 46 image pairs (4-view images of 23 patients) containing 5 normal cases and 18 abnormal, successfully identify 95% of abnormalities, however at the expense of returning 3 false positives per image (i.e., 276 false positives) [231]. In another study [73] they used 36 image pairs, again with 95% true positive and 3 false-positives/image.
- Glatt et al. [75] used a "limited number of mammograms".
- Smith et al. [196] used 33 images containing 17 benign and 16 malignant masses, correctly segmenting them.
- Ackerman and Gose [1] examine 60 benign and 60 malignant masses.

Once again the statistical significance of these results are compromised by the small sample space of the data sets relative to the variability of the mammographic signs that these techniques are trying to characterise, and this must be borne in mind when considering their

potential. Naturally the acquisition of more data for training/analysis would shed further light on the suitability of the techniques.

3.3 Evaluation of parenchymal pattern

Wolfe [223, 224, 225, 226] associated four classes of parenchymal pattern with varying degrees of risk for developing malignancy. Following these publications, some attempts were made to develop systems capable of identifying and classifying these mammographically observed patterns. The main criteria of Wolfe's classification is the prominence (or otherwise) of ducts and the debate continues as to whether this classification is really of use in determining malignancy [17, 58, 57, 78, 107, 121, 147, 172, 227]. Consequently there has been only a few publications in this area to date.

Texture analysis for classification of parenchymal patterns has been applied by Magnin et al. [140] using measures of entropy, contrast and local homogeneity. They claim some correlation, but not Wolfe's four discrete classes in the assessment of 27 images.

Miller and Astley [148] use a morphological and texture energy approach to segment left and right breast images into significant regions, using them as the basis for comparison in the search for asymmetry. In recent work [149] they evaluate a number of metrics defining the global parenchymal tissue pattern in each breast of a left/right pair, including shape, brightness distribution and topology. These parameters are then used in neural network analysis to classify the mammogram pairs into normal and abnormal categories. Training and assessment was done on a leave-one-out basis for a data set of 52 mammogram pairs achieving 72% correct classification for a linear discriminant classifier.

Caldwell et al. [26, 27] determine the fractal dimension of the image surface, training their system on 29 mammogram images. They showed that the fractal classifier was able to identify many of the parenchymal patterns associated with a high risk of breast cancer, however it was unable to resolve the classification of parenchymal tissues described by Wolfe [225, 226]. They note this technique requires much development.

3.4 Noise

Accurate diagnosis of disease from a mammogram is naturally dependent upon the quality of the film, which in turn depends on many factors (section 2.4 and figure 2.8). Of particular importance is noise, which is classically modelled as a high frequency random distribution. As mentioned previously (section 3.1) calcifications can resemble noise processes and thus a great deal of work in tackling noise has been done in the context of calcification detection. By mobilising prior knowledge about the structure, shape, size and clustering of calcifications they can be differentiated from noise [32, 33, 34, 49, 48, 61, 60, 220, 108].

Many techniques have employed conventional noise suppression algorithms as a component of signal isolation, including median filtering [32, 33, 128], Gaussian and DoG filtering [170,

184]. In work related to chest radiographs DaPonte and Fox [44] use Gaussian and median filters to reduce noise, reporting that the "gaussian filters were much more computationally efficient than the median filters with approximately equal effectiveness in noise reduction". As these techniques indiscriminately smooth across all frequencies, often suppressing both signal and noise, many authors have avoided using them.

Chan et al. [32] claimed that noise is of a higher frequency than calcifications, and applied band-pass filters and derived a box-rim filter [34] to eliminate the noise.

In an attempt to extract the edge-blur surrounding a circumscribed mass, Richter and Claridge [170] noted that application of a spatial low-pass filter increases the blur of such an edge. Conversion to polar coordinates facilitated application of a one dimensional filter to the image tangentially, preserving the radial profile. This had the effect of reducing the noise, but retaining the directionally important high frequency information.

Lai et al. [128] use a selective median filter, since they claim ordinary median filtering does not preserve sufficient edge detail for their ultimate goal of detecting circumscribed lesions based on the edge strength of those features. This filter assigns a given pixel a new value that is the median of the pixel values in the neighbourhood of the pixel that are within some threshold difference from the current pixel's original value.

More recently, Highnam [97], and Karssemeijer and Erning [112] have applied the physics underlying the mammographic process to suppress the noise. Highnam develop a model of the mammographic process to account for some of the intrinsic variables (section 2.4), whilst Karssemeijer and Erning develop a grey-level transform to make the absolute noise level across CCD digitised images constant.

3.5 Segmentation

Segmentation of the image into a region of interest and the background is an issue tackled by many papers, for varying reasons, not least of which is the significant computational saving by applying the algorithms to the region of interest only. Other reasons include facilitating image registration and exclusion of false alarms arising from the dense tissue of the pectoral muscle.

Suckling et al. [203] use Kohonen self-organising neural networks retrained on data extracted from each mammogram resulting in a network for segmentation of the given mammogram into regions of pectoral muscle, fat and parenchymal tissue. The network is retrained for each and every mammogram to be segmented in a self-classifying scheme. Although initially promising, the technique has only been applied to 10 mammogram pairs.

Many algorithms for other higher-level processes require a segmentation step at some stage, and in general the requirements of this class of segmentation are not as stringent as those of the dedicated classifier of Suckling et al. since any errors can be corrected or accommodated in the subsequent processing. In this class of segmentation, the task has generally been achieved by some reasonably simple algorithms including simple thresholding [49, 48, 127, 131]. Kimme et al. [122] used a ridge following algorithm to segment the breast from the background

in xeromammograms by following the 'halo' which commonly, but unfortunately not always, accompanies the breast edge. Hand et al. [84] improved the robustness of this technique by maintaining candidate pixels and choosing the best when conflicts arose. Semmlow et al. [184] used a uniquely shaped spatial filter to search the central area for boundary features. It included a nipple detection filter and this point was used as the starting point for the boundary detection.

The difficulty in detecting the breast-edge boundary is intimately related to the digitisation process. Images recorded by high quality scanners generally have very well defined boundaries and consequently segmentation is trivial. However, those images recorded by CCD video rate cameras often have poorly defined boundaries and relatively 'active' backgrounds. With the possibilities of direct digital recording of the image (rather than film) it should be a trivial matter in the future.

3.6 Bilateral registration and asymmetry

Clinicians often look at images of the left and right breast together and assess the 'symmetry' of the breast pairs. Automated comparison of left and right images requires that they can be aligned. This however, is not a trivial problem as the breasts are often different sizes and shapes, and rarely are they positioned 'equivalently' or undergo equal compression when the mammogram is taken. In fact it is the compressible nature of the breast that makes successful alignment, or image registration so difficult. Once aligned, the images can be assessed for asymmetry. This is an extremely difficult task due to the vague definition of what constitutes normal symmetry in the context of breast disease, and this difficulty is reflected in the lack of papers reporting a degree of accomplishment of this task.

Traditional techniques for image registration have often assumed object rigidity and searched for corresponding control points in each image and then proceed by aligning the control points. The deformable nature of the breast means that even if control points could be defined (in any case it is unclear just what should be matched), their relative geometrical arrangement is almost certainly going to change between both left and right images and temporal exposures of the same breast. Another complication is that this change could form part of the 'asymmetry' that is being search for, thus a least squares fit of the control points is generally an undesirable approach.

The most rudimentary approach has been to align the images manually [73, 222] and provide a subsequent check for false alarms caused by misalignment [73].

The simplest automated approach has been to consider only the position differences and ignore the size and shape variations by simply inverting one image and align the breast/background border as best as possible. Hand et al. [84] initially aligned the nipple points and then used a least squares fit to align the boundaries.

Kimme et al. [122] assumed that the normal breast would have regions with corresponding matches in the other breast image. They triangulate the breasts using the nipple and image

edges as control points and then divide the breast into a large number of rectangular regions (up to 144). A range of statistical measures for each region are computed and compared across images to solve the correspondence problem, and thus define a dense warping map.

Image registration for purposes of asymmetry checks looking for 'extra' or missing features between breasts have employed the traditional control point alignment and image warping operations mentioned above. Zhou and Gordon [235] report that the control-point extraction is very difficult to automate (and was done manually) and the overall process computationally intensive. Lau and Bischof [131] simplify Zhou and Gordon's approach by extracting three control points only, the nipple and the two breast edge/chest wall corner points, and translate, rotate and scale one image to match the other. They report success in the detection of tumours, however warn that this method alone is not reliable enough for clinical application, due mainly to the high false-negatives reported. As described in section 3.2, Giger and colleagues [73, 231, 232] use a similar warping technique to align bilateral images and then use a subtraction technique to detect asymmetric densities.

Stamatakis et al. [198] block the images into 10×10 pixel blocks. Statistics of each block are measured and the bilateral correspondence sought, from which an affine transformation is evaluated and applied to the original images to effect the image alignment. The procedure was assessed on 10 pairs of images from which the authors claim good alignment, assessed visually.

The question of image registration must be defined within the context of the application for which it is intended. Geometrical warping of images may or may not be grey-level preserving and this must be considered in subsequent analysis. Highnam [97] notes that image registration between images of the same breast should be much simpler due to the obvious similarities of features, and similar imaging quality if the two images were taken by the same radiographer. Thus assessment of temporal change may prove to be more reliable than the assessment of bilateral asymmetry.

3.7 Intensity normalisation

The absolute film density values on mammographic films are affected by a large range of essentially uncontrollable variables (section 2.4) which make the direct comparison of film density between images impossible. This fact obviously has implications for the bilateral and temporal comparison techniques outlined earlier.

Attempts to overcome these limitations initially used thresholding from histogram analysis [1, 32, 84, 128, 220], pixel ratios [1, 61, 196] and image normalisation [122, 128].

Normalisation has been employed by Miller and Astley [148] to improve the stability of textural analysis of images. Conversely Magnin et al. [140] decided against normalisation due to the risk of eliminating valid intermammogram density variations arising from physical differences in the breast tissues.

In more recent work Highnam and colleagues [97, 100] have modelled the entire mammographic system from the X-ray beam right through to the digitisation process. The models of

beam hardening, scattered radiation, poly-energetic beams allow an image to be normalised for the imaging variables. This results in a 'level playing field' for subsequent analysis and interpretation.

3.8 General image enhancement

Given that calcifications present as very small sharp intensities, and masses are often clouded by surrounding tissues, it is an obvious step to try to 'enhance' the images prior to feature extraction. There are two distinct approaches possible for utilising enhanced images, each dictating quite different requirements of the enhancement algorithms. If an enhanced image is used solely to prompt a radiologist into looking at specific locations more thoroughly whilst ensuring that all interpretative and diagnostic decisions are based upon the original image, then the need for accuracy, precision and robustness of the algorithm is relatively weak. Naturally it is still desirable to be able to explain to the radiologist how the enhancement works and thus a formal model-based development of the algorithm is still desirable. If on the other hand however, it is proposed that the enhanced image should form the basis of interpretative or diagnostic decisions, either by human or automated analysis, then the requirements of the algorithm are significantly more stringent. If adopting the first technique of visual enhancement for prompting only, and furthermore utilising the freedom this approach allows, one must exercise extreme caution to ensure that the approach does not attempt to encroach on the second application of enhanced images mentioned above, that is, using the image as the basis for decision making.

Morrow et al. [152] propose an empirically derived contrast mapping function for visual enhancement of mammographic images. The contrast mapping function was chosen by a radiologist from a number of alternatives presented by the authors after the authors rejected the chest radiograph enhancement operators of digital unsharp masking filter used by Rogowska et al. [175] and Sezan et al. [185] and the Sobel operator used by DaPonte and Fox [44] on the grounds that "these procedures change the appearance of the image too radically to be applicable in mammograms". This statement, made in support of their own algorithm, highlights a difficulty with this approach, that is: who defines 'acceptable enhancement', and how does one control that parameter? Morrow et al. establish that the chosen contrast mapping function *appears* to enhance *the* mammogram images that it was applied to with greater aesthetic appeal than the other mappings they proposed. However, without a thorough understanding of the mapping function and its relation to the raw data, it is impossible to draw conclusions regarding the algorithm's optimality and areas for improvement, the results of applying the algorithm to a broader set of images, or its execution on different hardware. This last point is indeed of concern, since in the example of the Morrow et al. algorithm, the procedure does not include any parameters characterising the display hardware upon which the images are presented. Clearly these parameters are an integral component of the appearance of the images, and they have been included, in the loosest sense of the word, by selecting the prettiest

looking picture and thus implicitly specifying them in the mapping. Consequently, if the software was ported to hardware with different characteristics the results would be different, and certainly unpredictable. Thus even for an algorithm proposed merely as a prompting device it is important to realise that a totally ad hoc approach can lead to difficulties.

Rather than seeking to chastise Morrow et al., it is hoped that the use of their case as an example representative of many of the approaches in this field, will convey an understanding of the difficulties of these approaches due to the uncertainty and unpredictability of their results. With that in mind, we now review the available literature.

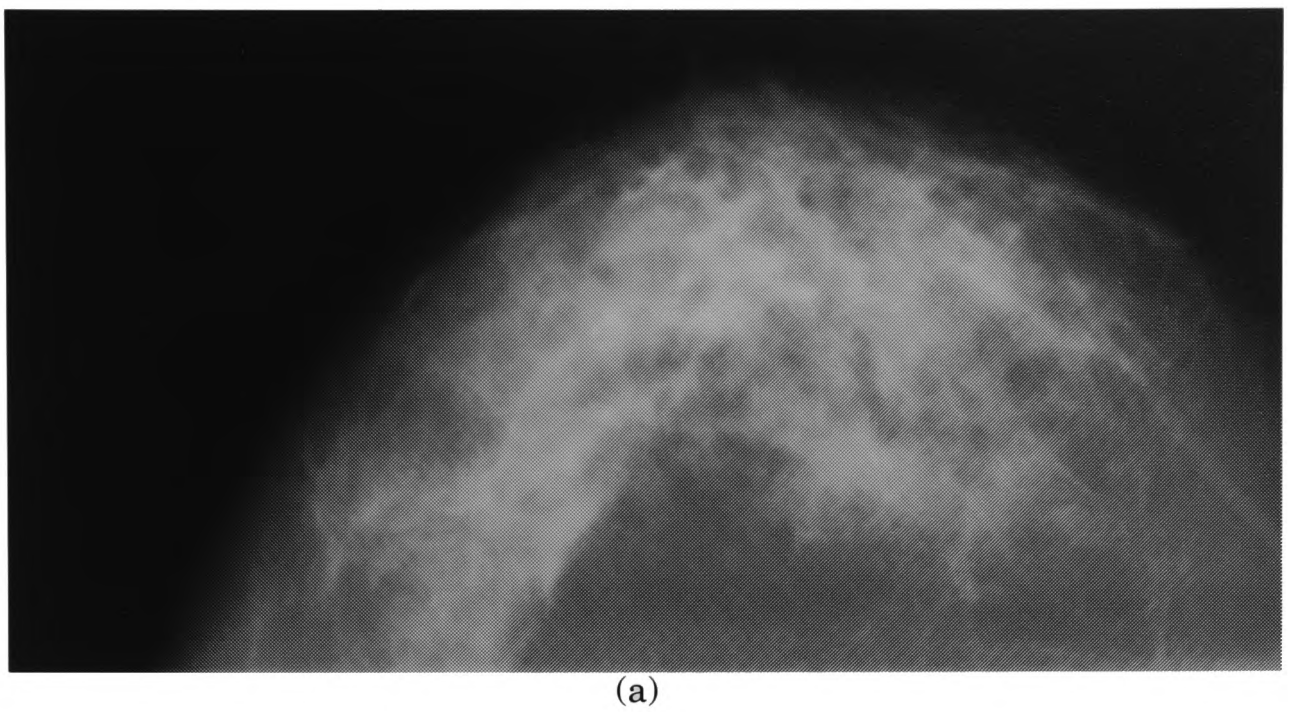
Fredfeldt et al. [72] enhance image contrast by locally estimating the standard deviation of the grey-level intensity for a pixel's neighbourhood and adjusting the pixel's intensity "so that this measure of variability reaches a given level".

Hale et al. [82] use Adobe Photoshop®to visually enhance the appearance of mammogram images for diagnosis. They support these and other enhancement techniques since "locally adaptive image filters can differentially affect only certain components of a lesion such as sharpness of border details". Clearly care must be exercised at this juncture since these are the very components of an image that are primary indicators of cancer, and thus their alteration by ad hoc procedures must be carefully evaluated, even in a prompting only situation. Furthermore, Hale et al. assert "computers can be programmed to recognise patterns of microcalcifications and render an opinion regarding histology". An opinion! This statement highlights the dangers of inappropriate computer based analysis in the hands of those who fail to appreciate the capabilities and limitations of computerised systems. As expressed earlier, unfounded faith in automated systems is dangerous and is to be avoided.

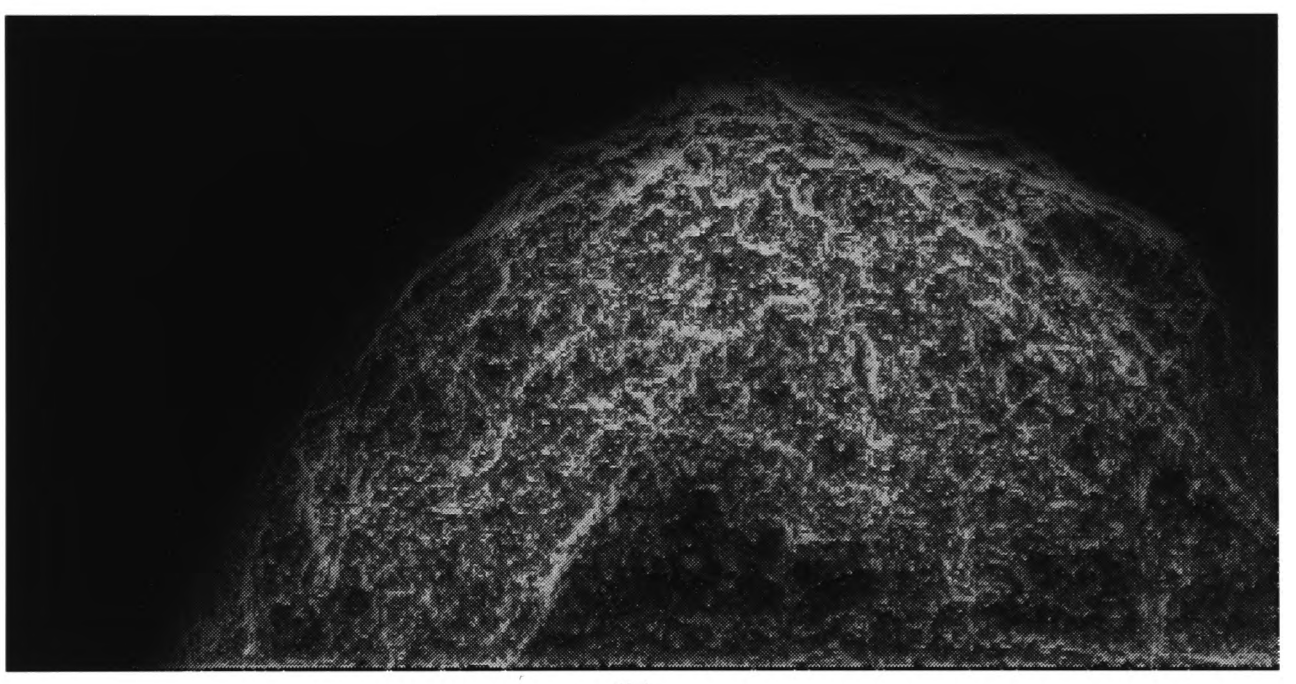
Tahoces et al. [209] describe an automatic spatial filtering technique. They smooth the original image twice, take a linear combination of the three images (original and two smoothed) and then perform a non-linear contrast stretch. The authors propose that this enhancement technique might be suitable for enhancing edge detail and general image contrast to aid the search for microcalcifications and assessment of general parenchymal processes.

Gordon and Rangayyan [76] propose a contrast enhancement method based upon mapping the image intensities to a contrast range between zero and one, taking the square root (preserving the range of contrast data) and then mapping back to image intensities. Dhawan and colleagues [51, 52, 53] also propose this square-root mapping, and additionally a number of piecewise mapping functions 'tuned' to enhance specific mammographic signs such as calcifications. With the exception of the square-root operator (which might be justified by considering that the luminance of screen-display hardware is proportional to the square of the image grey-level—but not the contrast—although the authors do not consider this), these operators appear to be simple heuristics for which there is no quantitative justification. The difficulty with this approach is that the enhancement is offered for automated analysis, not merely prompting.

To illustrate the effects of this class of contrast enhancement operators, figure 3.1 shows the results of applying the algorithms of Dhawan et al., and Gordon and Rangayyan. Note



(a)



(b)

(c)

Figure 3.1: Effects of contrast enhancement according to Dhawan et al. [51, 52, 53]: (a) original image, (b) image enhanced by mapping of Dhawan et al. optimised for microcalcification detection, (c) image enhanced by square-root mapping (proposed as a 'general' structure enhancement mapping) of Dhawan et al., and Gordon and Rangayyan [76].

the comparison between the relatively blurry and ill-defined edges of the central region of the image (figure 3.1(a)) and the 'general structure' enhanced version of figure 3.1(c). The effects of the enhancement, which to the eye give the appearance of 'focussing' a slightly 'defocussed' original, are difficult to quantify, particularly regarding the introduction of artifacts. It is clear that sharper edges have been introduced into the image and this may, for example, alter the mammographic signs of malignancy for masses (see sections 2.5 and 3.2). The difficulty of quantifying the effects of this algorithm is a direct result of its heuristic formulation. Dhawan et al. note that their method requires the careful balancing of thresholds as a higher value "causes saturation and over enhancement over many regions, ... while a lower threshold ... is more sensitive to small noise-like variations. It also causes some artifacts of highlighting spurious features (like dots) against the background." This is particularly obvious in the 'enhanced' image of figure 3.1(b). The assertion that this enhanced image is more suitable for detecting calcifications is, at best, questionable, and its inclusion in an automated analysis system would pose many new difficulties.

A great many of the enhancement techniques proposed in the literature are based on ad hoc procedures, and whilst this is not necessarily an undesirable feature, within the context of automated analysis (the context of the present work) this is undesirable. In conclusion, the use of general ad hoc enhancement routines based on non-existent or unfounded principles, as illustrated by the examples reviewed here, is to be discouraged within the context of automated analysis since there is no guarantee that the real data of the raw image will be logically represented in the enhanced image.

3.9 Other techniques

The techniques reviewed in the previous sections attempt to detect abnormalities by directly emulating one or more of the procedures undertaken by the clinician, or to overcome some factor that limits this approach.

Unlike this approach, Highnam et al. [97, 102] propose a technique that draws on additional information available in mammograms that is not utilised by the procedures employed by clinicians. They propose acquisition of mammograms of the same breast at different compressions. These differential-compression mammograms are then analysed for relative tissue deformation. In initial trials the technique has displayed the ability to give a "vivid impression of the movement and deformation of tissue". The authors expect the compressibility characteristics of different tissues (tumorous masses tend to retain their three-dimensional structure even under vigorous compression [6]) will be apparent, allowing differentiation of problematic cases.

3.10 Summary

There are a number of observations and conclusions that can be drawn from this review of the previous work in automated analysis of mammogram images. Although some ad hoc

procedures may have a place in prompting, such procedures with little or no justifiable links with the image data are to be strongly discouraged in automated analysis/interpretation as they have the ability to alter the mammographic signs of malignancy without that fact being necessarily understood by a clinician using the final results.

Those authors that have attempted to assess the robustness and stability of their algorithms have compared the algorithm results with the opinion of a radiologist. This process has however generally been undermined by the lack of images/cases to which the algorithm has been applied. Most algorithms have been explored over a handful of cases, a few over 50–100 only. Given that mammograms are acquired at the rate of approximately three million a year in the UK alone, that the incidence of abnormal cases is so low, and that the mammographic signs of malignancy are so varied and vague it is imperative that algorithms are assessed over as large a sample space as possible to try and achieve some statistical significance in the results. The fact that algorithms have not been examined over a large sample space is really a reflection of the enormous computing resources (and time resources with today's technology) required to capture, store and process mammogram images. With time these limitations have, and will continue to, diminish and a more realistic appraisal can be made.

Whilst on this point of statistical significance, the recent introduction of neural classifiers into this field in order to solve many of the classification problems associated with calcification, mass and parenchymal interpretation must be considered carefully. Given that the characteristics of these features are so varied and in reality, poorly specified (that is we cannot say 'this is definitely cancer, whilst this is definitely not'), any attempt to capture the statistics of these features in a classifier must expose the training phase of the classifier to an enormous variety of cases, preferably a large quantity (read much larger than one) of each case represented in the population space. Without exception the applications to date have tried to train networks on a handful of cases. This is definitely not representative of the variety of cases that the classifier will be called upon to analyse, and accordingly this effectively negates the entire process. A serious rethink of these procedures is necessary before any real advances can be made with their use. This point will be taken up again and considered in much greater detail in section 6.4.

4

Finding the curvi-linear structures (CLS)

4.1 Introduction

4.1.1 Anatomical attributes of the CLS features

There are many mammographic signs that typically attract the attention of the diagnostic radiologist during the search for breast abnormality and disease. Some of the main mammographic signs of breast abnormality are calcifications (as primary indicators of DCIS) and densities (indicators of masses, both circumscribed and spiculated). Other indicators such as bilateral asymmetry can present in a variety of forms such as asymmetrical densities and uneven parenchyma, but almost always these cases can also be classed as being spatially diverse, that is, of low spatial-frequency.

In the case of calcifications, their shape and arrangement are important in assessing the degree of malignancy. Additionally it is useful to be able to distinguish both the inter-ductal and intra-ductal calcification arrangement, and to differentiate arterial and ductal calcifications [168, 174]. Consequently it would be very useful to have a description of the locations of the ducts and blood vessels in an image, allowing for a comparative analysis of calcifications present in an image.

When assessing a detected mammographic density, a radiologist routinely checks the border surrounding the density for spiculations or radial fibrous tissue anchored to the density, since such spiculations are strong indicators of malignancy. Similarly the radial spiculations surrounding a radiolucent centre are possibly indicative of disease as they are the only signs of radial scar. In fact the detection of radial scar is a very difficult problem due to the subtle appearance of these features in an image. Clearly in these situations, a feature-space describ-

ing the spiculations would be useful for determining the degree of spiculation associated with a mammographic density identified by some other means, or as the primary analysis domain for identification of instances of radial scar.

Mammographic densities (spiculated or circumscribed) appear as spatially low frequency patches of higher film intensity. The search for these features is often complicated by the higher-frequency textural variations in image intensity due to the fibrous tissue, milk ducts and blood vessels which occur with much smaller spatial extent, but with high incidence (hence high-frequency). As noted in section 3.2, these limitations have consequently precipitated the introduction of additional measures and procedures during the formulation of many density detectors reported in the literature. Note that simple low-pass filtering of an image is an unacceptable method for removing these high-frequency features since they are not the only high-frequency components of an image. The border (edge) of a mammographic density is another region of high-frequency intensity variations, and consequently the indiscriminate filtering of these techniques would also remove (or blur) the information defining the density border. As previously mentioned, these attributes of a density are a key indicator of malignancy. By contrast, if a description of the location of the irrelevant (in a diagnostic sense) high-frequency features were available, it may be possible to compensate for their existence in a systematic fashion consistent with the goals of any image analysis or interpretation that may subsequently be attempted. This scenario is in fact the case, and is demonstrated in chapters 5 and 6, where a method for detecting mammographic densities is developed that relies crucially on identification and suppression of the textures associated with the irrelevant high-frequency features.

In summary there are applications that would benefit from a high-level description of the milk-ducts, blood vessels, mass spiculations and other higher-frequency fibrous tissues present in an image, allowing many subsequent tasks and further analysis to be undertaken. It is to the task of finding such a description that we now turn our attention.

4.1.2 Mammographic attributes of the CLS features

In order to identify the vessel/fibrous features in an image, it is necessary to determine the mammographic signs of these features, that is, the *footprint* left by these features in a mammogram, so that these attributes may become the subject of our search.

When imaged mammographically, the extra density of breast tissue that an X-ray must traverse when passing through the vessels/fibrous-tissues causes the beam to be attenuated slightly more than a beam that does not pass through such features. Consequently the film is less exposed in these regions and therefore has lower film density (appears brighter). By assuming that the cross-sectional profile of the vessels/fibrous-tissues in the compressed breast are elliptical, section 4.4.1 develops a model of the cross-sectional profile of the mammographic film density, and ultimately the image intensity for these features. From these considerations it is shown that such features increase the image intensity by a single grey-level for approximately

each $180\mu m$ of feature thickness. Due to the compression of the breast it is reasonable to assume that these features will be wider than they are thick (that is, the major axis of the ellipse is orthogonal to the incident X-ray beam rather than parallel to it). Therefore if we adopt the conservative view of considering the features to be of roughly equal extent in width and thickness, then it is expected that only features in excess of $180\mu m$ in width will be found in mammogram images. In fact, due to image noise (see section 4.1.3) resulting from both the imaging and digitisation processes, the expected minimum width of these features in an image should be substantially greater than $180\mu m$. Considering that blood capillaries are $7–10\mu m$ in diameter, this explains why the majority of the blood vessel network is not imaged by the mammography process, leaving only the larger vessels to be resolved in the film and subsequently, the image. An analogous argument can be developed for the milk ducts and fibrous tissues. Consequently there is no reason to search for features narrower than this limit as an image carries no such information.

Inspection of a range of mammograms (such as figure 2.7) reveals that the high-frequency features are locally (near) linear and nominally 1-dimensional. This is hardly a surprising result, since the vessels that can be imaged are the larger vessels in the respective arterial/venous/ductal networks and it is reasonable to assume that these will be locally linear (cf. the distribution of a typical tree root system where the largest roots are the straightest for the longest distances). Equivalently, the fibrous tissues, as functional components of the breast stroma, will carry their tensile loads over relatively straight paths (at least locally), and thus they will appear mammographically as linear structures.

Consequently the mammographic appearance of the high-frequency vessels/fibrous-tissues possess the attributes of local linearity, 1-dimensionality and slightly higher image intensity than the surrounding image regions. This description might lead one to the conclusion that it would be impossible to distinguish between the types of features—blood vessels, milk ducts and fibrous tissue—since they exhibit similar mammographic features. Although they are indeed similar, there are some distinctions between the respective features that would aid this task were such a segmentation desired. Recall from section 2.1 that the milk-duct network converges at the nipple. A common radiological tool for determining between ductal and arterial calcifications for instance, is to determine the orientation of the vessel they may apparently lie in. A vessel directed towards the nipple is likely to indicate a milk-duct, whilst a vessel in an orthogonal direction is likely to be a blood-vessel. Such orientation information is useful for determining between vessel types. Determining between vessels and fibrous tissues is generally much easier as the vessels form part of a connected network, whilst the fibrous tissues need not. Thus the vessel networks are expected to give rise to longer features than the fibrous one. Furthermore, since the vessels distribute fluid (blood/milk) they must necessarily display a far greater stability in feature width than the fibrous tissues (cf. the gradual reduction in diameter of air-conditioning conduits as they deliver fresh air to sites further away from the plant).

Thus the attributes describing the milk-duct, blood-vessel and fibrous tissue features are very similar on a local scale, and subtly different on a more global scale. Fortunately many of the applications which would benefit from a description of these features do not require to be able to distinguish between these types, but rather, simply being able to identify those areas of an image that are the milk-ducts, blood-vessels and fibrous tissues collectively will suffice. Considered in this fashion, the computer vision task to find the regions comprising these features can be defined as finding the regions with slightly increased intensity that are nominally 1-dimensional, and locally linear—but may diverge from strictly linear over larger scales, that is they may exist up to some maximum curvature in a continuous sense. For this reason, these features are commonly referred to by the collective term *curvi-linear structures* (CLS), and it is a description of these CLS features that we desire. To further appreciate this definition and task specification, the reader may wish to skip ahead at this point to inspect the CLS features identified in a number of images by the algorithm developed in the following sections. Figures 4.29 through 4.33 on pages 107 through 111 illustrate a number of such examples.

4.1.3 Image Processing Constraints

Any attempt to extract information describing the CLS features (or for that matter, *any* features) is bound by the limits of the information available in the image. A factor degrading the useful information content of an image is noise. Karssemeijer [112] states that the film noise associated with mammography can be high, and in combination with the digitisation process, can pose significant difficulties for image analysis. Although eliminating some of the film noise may one day be possible by using direct digital imaging (DDI), this technique is still a long long way from becoming a routine screening instrument. Additionally there are substantial inventories of records on film that may one day be digitised, and possibly require analysis. Therefore, the current practice of digitising film to acquire images of mammograms will find regular application for many years to come, and thus we must address (at least be *aware* of if we cannot *address*) the issue of digitisation noise. To try and get a rough estimate of the noise in an image, consider the facet-model introduced by Graham [77] and developed further by Haralick and colleagues [86, 87, 230].

The principle of the model assumes that an image can be thought of as an underlying continuum or piecewise continuous grey level intensity surface. The observed digital image is however, a noisy, discretised sampling of a distorted version of this surface. By fitting piecewise continuous surfaces (common forms include constant, and bivariate linear, quadratic or cubic) to neighbourhoods of each pixel it is possible to estimate both the parameters of the underlying surface for a given neighbourhood and the variance of the noise causing the observed deviation from the fitted surface (assumed to be representative of the 'real' surface).

Following Haralick, fit a sloped facet (bivariate linear) approximation to the image surface,

and determine the squared differences between the fitted surface and the image as:.

$$\epsilon^2_{(i,j)} = \sum_{(r,c)\in N} \left[\hat{\alpha}r + \hat{\beta}c + \hat{\gamma} - I(r,c)\right]^2 \tag{4.1}$$

where N is the 8-connected neighbourhood region over (r,c) surrounding the central pixel (i,j) to which the facet-model is fitted to I, the raw image data. A least-squares minimisation of ϵ^2 gives the model parameters at each (i,j) as

$$\begin{aligned}
\hat{\alpha} &= \frac{\sum_{(r,c)\in N} r\, I(r,c)}{\sum_{(r,c)\in N} r^2} \\
\hat{\beta} &= \frac{\sum_{(r,c)\in N} c\, I(r,c)}{\sum_{(r,c)\in N} c^2} \\
\hat{\gamma} &= \frac{\sum_{(r,c)\in N} I(r,c)}{\sum_{(r,c)\in N} 1}
\end{aligned}$$

Each neighbourhood's normalised squared residual error $\epsilon^2/(\sum_r \sum_c 1 - 3)$ can constitute an unbiased estimator for the variance σ^2 of the noise, as modelled by the facet-model. When averaged over all the pixels in an image ($i \in K$) this forms a stable estimator of σ^2, as:

$$\widehat{\sigma^2} = \frac{\frac{1}{K}\sum_{i=1}^{K} \epsilon_i^2}{\sum_r \sum_c 1 - 3} \tag{4.2}$$

Whilst this method may appear sound, there is the small matter of the assumptions upon which it is based, and in particular the assumption that the underlying (supposed 'true') surface *can* be approximated by a piecewise continuous (in this formulation, bivariate linear) surface. Clearly the very CLS features that are the subject of this chapter produce discontinuities in the image surface at their boundaries, as do mammographic densities etc. Across such image regions, an attempt to fit a continuous surface would, in the formulation of the facet-model, incorrectly class the discontinuity as the result of a noise process, and attempt to quantify the noise variance as some function of the contrast across the feature boundary. However, rather than evaluate ϵ^2 (equation (4.1)) at *each* pixel of the image I (and therefore include those pixels whose neighbourhoods include CLS boundaries) when calculating the noise variance estimate of equation (4.2), if ϵ^2 could be calculated at only those pixels that do not include a CLS pixel in their neighbourhoods, that is calculate it for 'background' pixels that do not include a CLS border in their neighbourhoods, then the assumption of a locally bilinear surface can be sustained. Since the noise is assumed to be white Gaussian (non-systematic) it should be equally distributed over the extents of the image. Thus measuring ϵ^2 only at known background pixels should not bias the calculation of $\widehat{\sigma^2}$ provided sufficiently large numbers of background pixels are used, but rather it should vastly improve the accuracy of the noise estimate $\widehat{\sigma^2}$ since the model actually fits the data to which it is being applied.

The remaining difficulty of how to determine which pixels have CLS pixels in their neighbourhoods is easily overcome by knowledge of which pixels in an image are CLS pixels, and then searching in the neighbourhood surrounding the pixel under investigation. Determination of the CLS pixels is, of course, the subject of this entire chapter, and accordingly, the

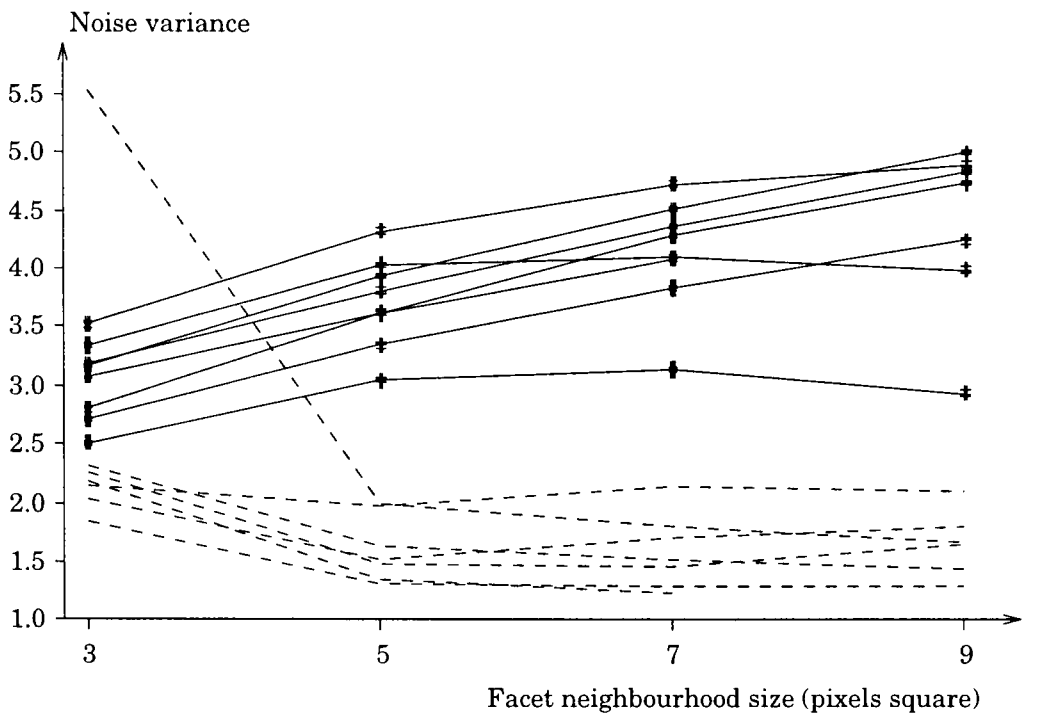


Figure 4.1: Facet model image noise estimates for variety of 'fatty' breast images calculated from averages over pixels without CLS pixels in their neighbourhoods. The solid curves show the estimate of the variance of the image noise, whilst the dashed curves show the standard deviation of the estimates. The standard deviations should really be plotted as error-bars, however since they would all overlap they have been plotted as curves for clarity.

details of finding the CLS pixels are deferred to the sections which follow. By using the methods developed in the following sections for finding the CLS pixels, the calculation of $\widehat{\sigma^2}$ from equation (4.2) proceeds directly.

There remains the difficulty posed by mammographic densities, and in particular including estimates of $\widehat{\sigma^2}$ calculated for pixels whose neighbourhood broaches the border of a mammographic density. As for the CLS features discussed above, this biases the estimation of the noise variance by fitting the noise to the contrast of the density border. Thus to further increase the accuracy of this model it is proposed to evaluate the model for images of 'fatty' breasts only. This assumes that the noise process varies little from image to image, which is not unreasonable.

Figure 4.1 shows the results of applying this noise estimation technique to a range of mammogram images that fit the assumptions of the noise model as outlined above. The figure gives both the noise estimate generated by equation (4.2) and the standard deviation of that estimate. Note that the noise estimate increases with facet neighbourhood size, since as the domain of the bilinear surface is extended there is less likelihood that a given patch of real image surface will fit the bilinear model over the extents of the domain. Thus the error terms of equation (4.1) will be large, increasing the estimate of σ^2. The estimate of the noise is clearly dependent upon the facet neighbourhood size and this size is selected by noting that the standard deviation of the $\widehat{\sigma^2}$ (as averaged over the many suitable pixels of an image) initially reduces at a facet size of 5×5 pixels and levels out, or maybe even increases again. Clearly the balance between too large a facet size (at which the surface variability of the model cannot capture the data sufficiently well producing unnecessarily large error estimates) and too small

a facet size seems to be at a facet size of 5 pixels. Observing the noise variance estimates of figure 4.1 at a facet size of 5 pixels, the general noise in a mammogram image is ≈3–4 intensity levels as modelled by the facet-model.

As mentioned previously, we develop a model for the expected CLS footprint in an image (see section 4.4.1) that allows calculation of the image intensity change caused by a CLS attenuating the beam relative to its immediate neighbourhood. Effectively a mammogram cannot resolve CLS features with X-ray path lengths (the distance travelled by an X-ray beam when traversing a CLS feature) below a few hundred microns. Anatomically the CLS features do exist in the breast at dimensions well below this minimum imageable size, and therefore an image will contain CLS features with differential intensities right to the limit of the quantisation. Our rough estimate of the image noise put its variance at about 3–4 intensity levels, and thus it is clear that for many CLS features the signal may be swamped by the noise in the image, giving very low signal-to-noise ratios. The ultimate consequence of these low signal-to-noise ratios is that an image *cannot* be smoothed prior to analysis. This pre-empts conventional noise suppression techniques which often employ a smoothing step under the assumption that the noise is white-Gaussian.

Another difficulty to consider when searching for the CLS features is revealed by simple inspection of a mammogram film. Often there are a number of CLS features that are roughly parallel and close by one another. Topologically, the adjacent features are quite distinct and it is necessary to reflect this in the extracted CLS feature description. Yet again this means that an image should not be smoothed prior to analysis since this would merge adjacent CLS features into a larger CLS feature spanning the original features.

4.1.4 Summary

Thus the image processing task at hand is to identify a description of the CLS features in an image, given that they appear as nominally 1-dimensional, locally linear regions with generally higher image intensity than their surrounding neighbourhood. The task is however complicated by very poor signal-to-noise ratios, possibly as low as $-6\,\mathrm{dB}$[1]. In light of this, it is inappropriate to smooth the images and thus alternative methods of suppressing noise must be considered.

4.2 CLS Detection — Related Work

In the terminology of the computer vision community, the CLS features can be thought of as *weak ridges* ('weak' due to the low SNR, and 'ridges' due to their local 1-dimensionality). There are a number of techniques available for identifying the ridges in images and it would therefore be logical to evaluate the suitability of these techniques for the present application of finding the CLS features in an image.

[1] Assuming the worst case of the discussion of the previous sections, we have a signal strength of 1 intensity level and a noise variance of 4, giving $\mathrm{SNR} = 10\log(\frac{1}{4}) = -6\,\mathrm{dB}$

The formulation of a ridge detector naturally proceeds by developing a model of the ridge and then implementing a strategy to identify instances of the model in an image, which are then interpreted to be the ridges of the image.

An early computer vision implementation of this algorithm design philosophy modelled a ridge as a pair of back-to-back edges. Clearly this approach needs to be able to identify the edges in an image, and then to be able to determine which edge pairs enclose a portion of a ridge. The difficulty here is that neither of these problems is trivial, nor can they in any sense be considered to have been 'solved' by the computer vision community. Considering the first task of identifying the edges in an image, one of the leading edge detector algorithms of recent times is that of Canny [28]. Figure 4.2 shows the edges identified in the images of figure 2.7. One of the difficulties of using the Canny edge detector in any application is determining an appropriate value of the smoothing variance for the first stage of that algorithm, and this is no exception for the present application. Figure 4.2 displays the edges found for smoothing variances of 1, 2 and 3. Clearly there are dramatic differences between the responses of this operator, and without a systematic method for selection of an appropriate value for the variance, there is no way to be sure of what the output from any analysis would be. This would make any attempt at subsequent processing a very difficult task. Note that as the smoothing variance is increased, the number of edges identified in the images decreases dramatically. This is a reflection of the introductory comments above warning against the introduction of a smoothing step into the processing. Since the CLS features have a weak signal strength, and are generally quite thin, the smoothing has both completely obliterated some CLS features, and merged other adjacent CLS features into single larger structures, and thus altered the topology of the image features. At even the lowest smoothing variance some of the smaller CLS features are simply too narrow for the Canny operator to identify two back-to-back edges, and consequently any subsequent analysis using these results would be confounded.

These difficulties are not unique to the Canny detector, nor are they unique to this application. In fact, the identification of ridges is a generic problem, with many diverse applications, for instance the inspection of printed circuit boards during manufacture, fingerprint identification, etc. A consequence of these limitations for finding ridges as back-to-back edges has precipitated the development of dedicated ridge-detectors designed to accommodate some of the shortcomings. The difference between the edge and ridge based approaches is that the ridge-detector develops a model of a ridge as a single entity (such as a top-hat in cross-section), rather than as a collection of smaller entities (such as back-to-back edges).

A leading exponent of such techniques is the ridge-detector of Haralick [85, 86] based on the facet-model introduced in section 4.1.3 above. The model of a ridge is developed as a continuous surface fitted to a (say, bicubic) facet arrangement that reflects the expected shape of the ridges. The operator is then configured to respond at regions in an image that fit the facet model of the ridge, thus identifying the local ridge-lines in an image. The difficulty with this approach is not so much modelling an expected shape of the ridges, but rather establishing a stable

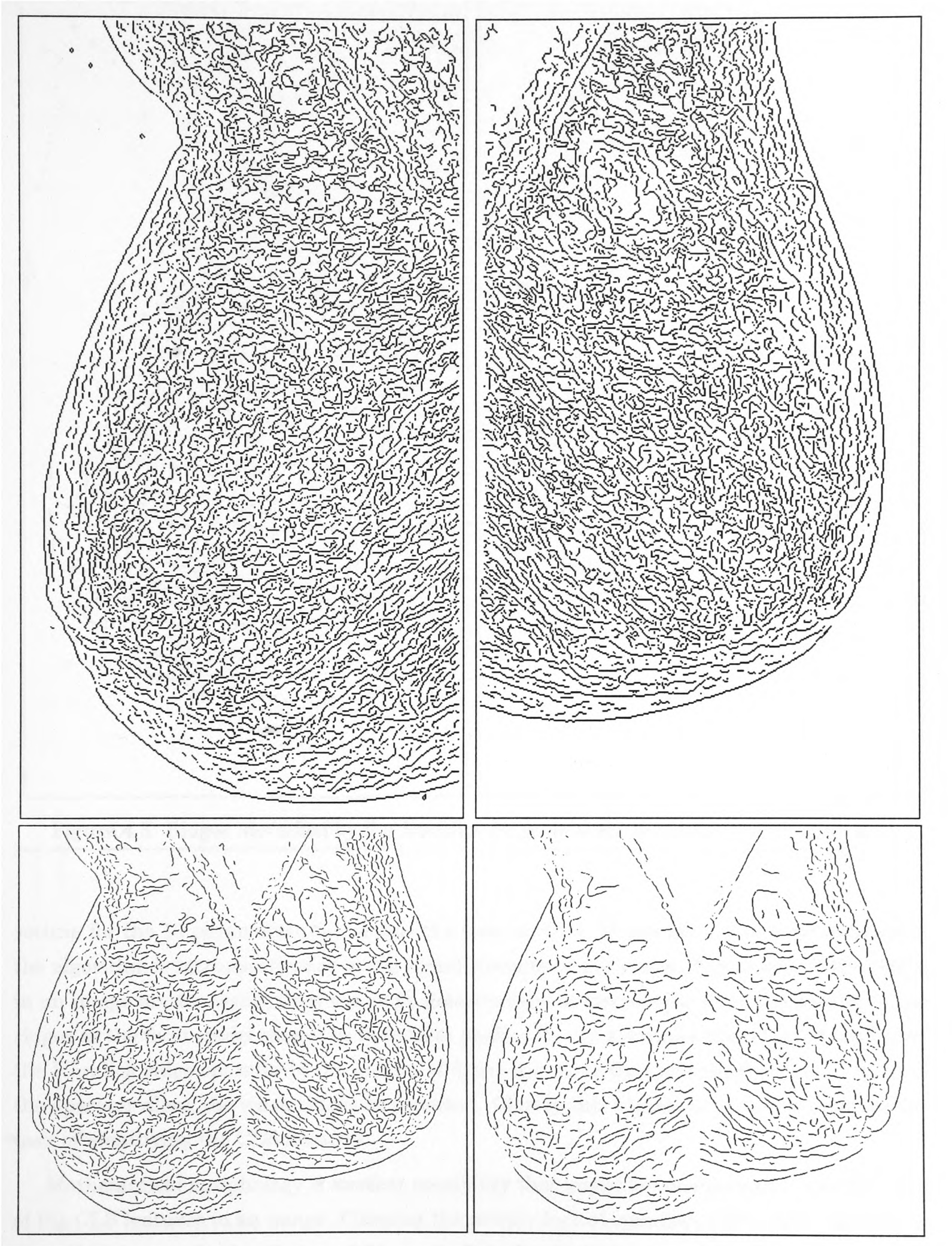


Figure 4.2: Edges identified by the Canny edge detector in the images of figure 2.7 for smoothing variances of 1 (top), 2 (bottom-left) and 3 (bottom-right).

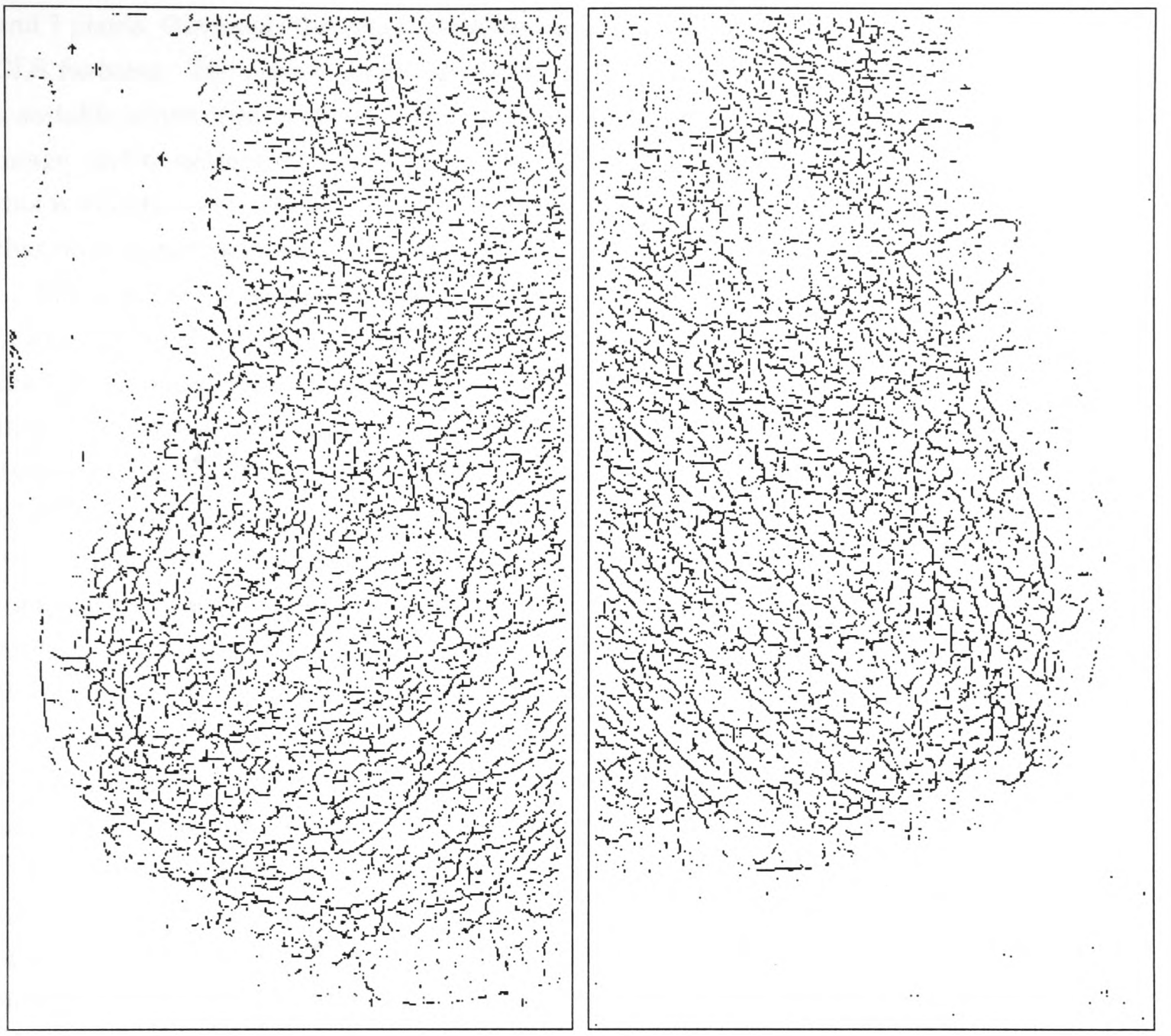


Figure 4.3: Ridges identified by the Haralick ridge detector in the images of figure 2.7.

setting for the 10 parameters that define the facet surface. Consequently the performance of the algorithm is intrinsically tied to the signal strength in the image. Essentially, if the ridges in an image are too weak, then the ridge detector fails to respond (or responds poorly) to the ridges as the facet model cannot be fit to their profile. Figure 4.3 shows the ridges identified by the Haralick ridge operator in the images of figure 2.7. Note the particularly poor response of the operator along the weak ridges as expected. Clearly this technique is not suitable for the task of identifying the CLS features.

Mathematical morphology is another possibility that might be considered for identification of the CLS features in an image. Consider the morphological operator of Wu [229] designed to identify the network of roads in aerial images of rural and urban scenes. The formulation of this operator is simply based upon identifying ridges from the closed residue of an image surface, a procedure that could equally apply to CLS detection. Figure 4.4 displays the ridges identified in the images of figure 2.7 for spherical morphological structuring elements of diameter 3, 5

and 7 pixels. Obviously the ridges identified by this procedure are unsuitable for detecting the CLS features. The reason these results are so poor is that it is very difficult to both match a suitable structuring element to the variations of CLS size and shape that are present in an image, and to achieve even a moderate signal in the closed residue of the image. Once again, this is simply an expression of the weak image forces defining the CLS features, a limitation that no structuring element can overcome.

Fleck [65, 66] describes a model-based edge detector (entitled *Spectre*) that builds a model of an edge as a region of high magnitude second derivatives of the image surface. Unlike the zero-crossing class of edge-detectors (which are based upon the zero-crossings between these large magnitude, opposite sign second derivatives) *Spectre* seeks to identify the second derivative peaks in the image, and then deduce the boundary locations from a consideration of the topology of the difference map. The directional second-derivatives of the image surface are estimated from evaluation of the second-difference images in six directions and then the information is assembled to robustly determine the local image curvature at each point from all the evidence available from the different directions. A further analysis of the first and third differences allows those responses due to noise to be segmented from those due to real image features. Note that this is achieved by assuming that the features under analysis by this technique impose a degree of continuity or smoothness in the results due to their extended size and generally locally continuous properties[2]. Figure 4.5 displays the second-difference map for the images of figure 2.7. Note that the regions of the images identified with a significant negative surface curvature (the black pixels) appear to capture a description of the CLS features of the image. To facilitate this observation/analysis, these regions of the images have been extracted and appear in figure 4.6. The ability of *Spectre* to seemingly identify many of the CLS pixels is due to a fortuitous set of circumstances, that is, fortunately, not difficult to reconstruct. The spatial extent of the second-difference filter kernels used in the *Spectre* detector is 5 pixels. Recall that the CLS features have higher image intensity than the surrounding regions due to the greater absorption of the incident X-ray photons by the CLS material. Consequently, when considering a pixel that is part of a CLS feature, yet is close to the edge of that feature (within a distance of half the kernel size), a filter kernel oriented in a direction normal to the direction of the boundary (or the local direction of the CLS feature) will straddle that boundary, and measure a negative second-difference at that location. Application of the algorithm over the entire image therefore results in identification of a thin strip of pixels adjacent to the boundary of, yet within, the CLS features as possessing a negative surface curvature. In a stroke of luck, the filter spatial extent of 5 pixels is large enough so that for most CLS features of the image, the identified pixels with negative surface curvature from one edge overlap with those of the opposite edge, effectively giving a solution to the 'back-to-back' edge problem mention in relation to the Canny edge detector above. Clearly there is some information here encoding a description of the CLS pixels, and although this is

[2]The algorithm has been formulated as a general image processing edge detector and has therefore earned its stripes in the analysis of (near) rigid bodies captured by say a CCD imaging camera (still or video).

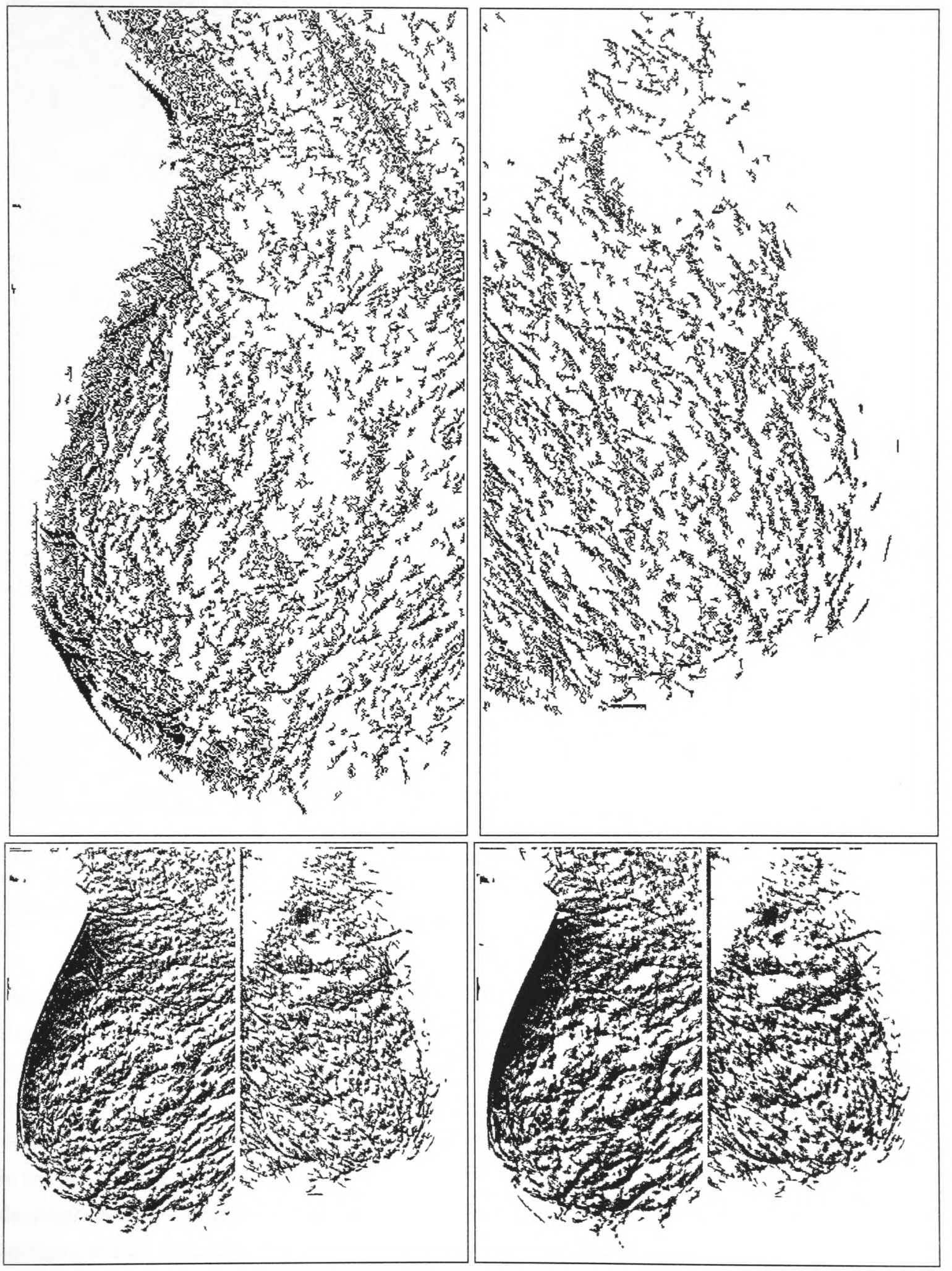


Figure 4.4: Responses to the morphological ridge finder of Wu for spherical structuring elements of diameter 3 pixels (top), 5 pixels (bottom-left) and 7 pixels (bottom-right).

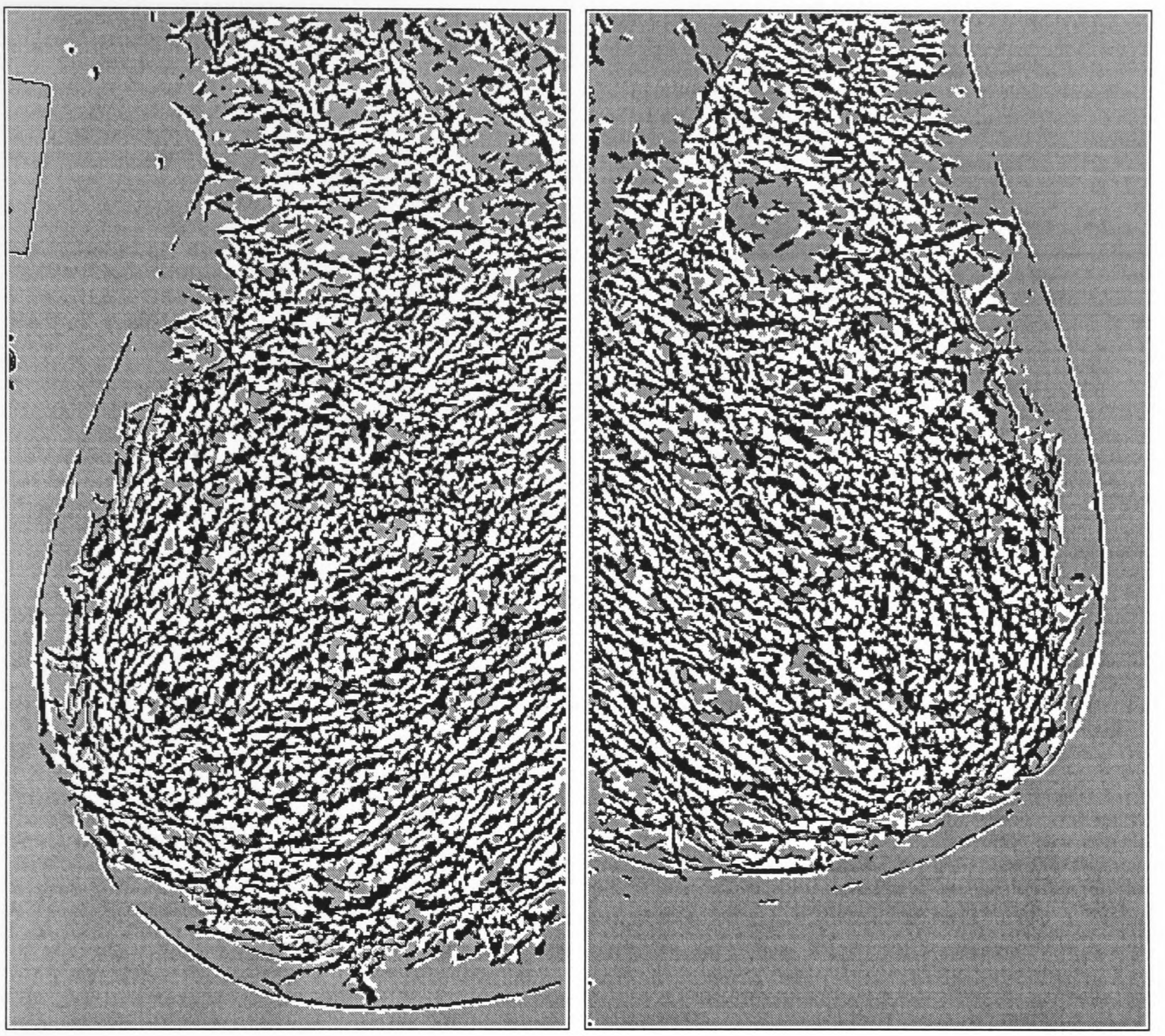


Figure 4.5: Results of the *Spectre* image surface curvature assessment for the images of figure 2.7. White coloured pixels indicate regions with positive surface curvature (ravines), grey indicates (near) zero curvature, and black negative curvature (ridges).

encouraging, there are a number of questions yet to be considered. Namely, although it appears that the second-difference of the image carries information identifying the CLS pixels, is this observation justified, that is, does the existence of a CLS feature robustly give rise to situations that can be identified by second-difference filters of the form used in *Spectre*? Furthermore, a description of simply the pixels that comprise the CLS features, although better than nothing, is hardly a description of the CLS features. How can such a list of pixels be developed into a description of the CLS features? It is these questions that we will address in the remaining sections of this chapter.

Before proceeding any further with this analysis, let us consider the implications of the last question just posed. Recall that we ultimately desire a high-level description of the CLS *features*, not just the pixels that comprise them. The possible techniques described above are all low-level, pixel-based methods that seek to identify the pixels that comprise the CLS features.

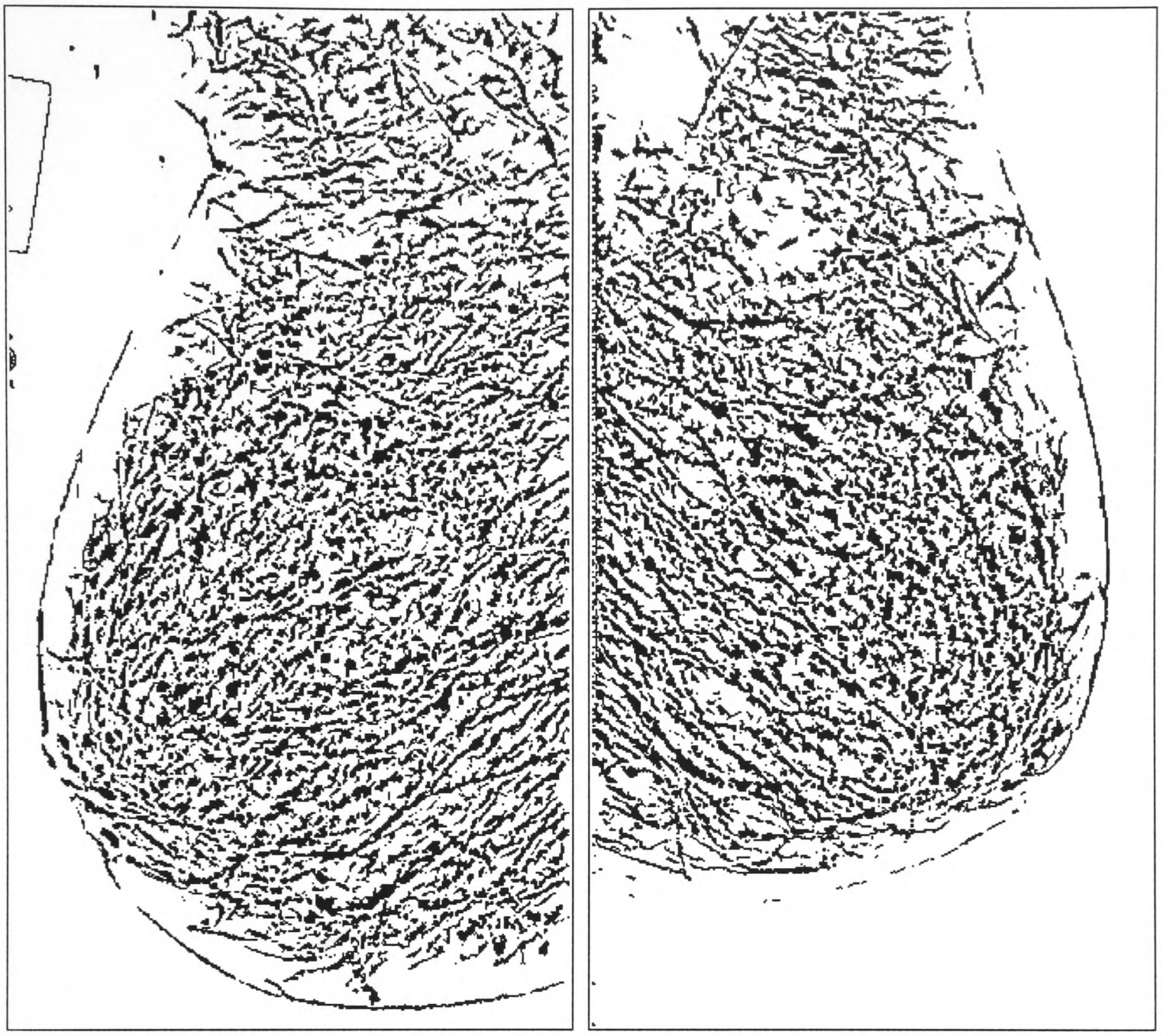


Figure 4.6: *Spectre*'s regions of negative surface curvature extracted from the images of figure 2.7.

In any such method a subsequent stage of processing would be required to determine a higher-level description. By contrast to this approach, can the high-level description be arrived at directly?

Hayton [89] has recently adapted the 'structural saliency' techniques of Sha'ashua and Ullman [186] in order to identify the salient structures (read CLS features) in mammogram images. The technique tries to 'fill-in' the missing information between components of features by searching for neighbouring features in the regions surrounding each feature. When two given features in close proximity reinforce the evidence available for each other's existence, they increment a measure of the global saliency at that location. Hayton used the results of the Haralick ridge finder (see figure 4.3) as the input tokens for this analysis, and thus the technique essentially aims to correct for the poor performance of that operator. Even if a more accurate representation of the CLS features were known, the algorithm still has many shortcomings within the current context. Suppose for a moment that a description of the CLS

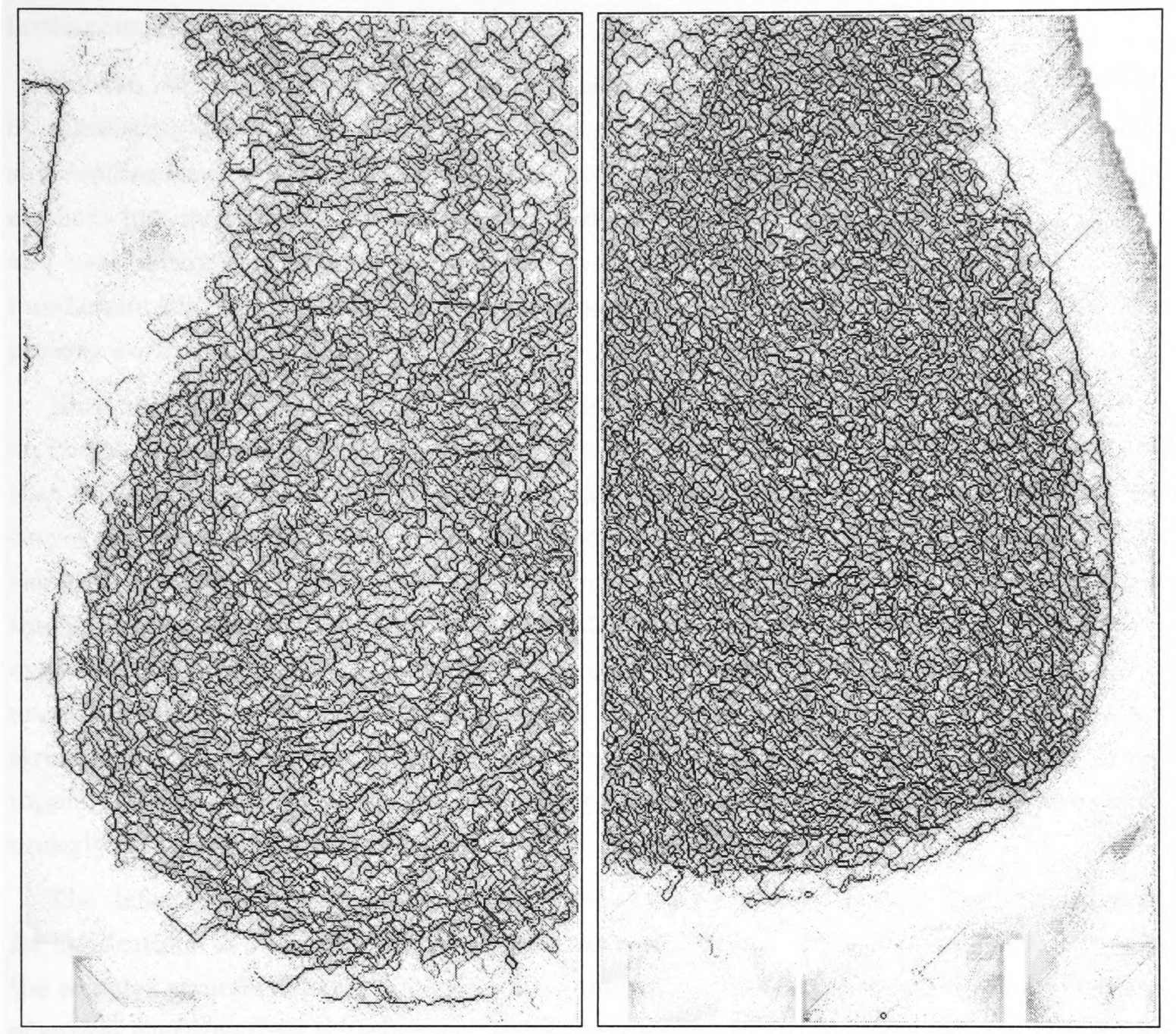


Figure 4.7: The relative saliency of each pixel for the images of figure 2.7 as identified by Hayton's [89] adaptation of Sha'ashua and Ullman's algorithm [186]. Higher saliency is indicated by darker pixels. Note the spreading of the support regions as the algorithm iterates.

pixels was known. By initiating Hayton's method at these pixels, one might endeavour to identify a high-level description of the relatively salient CLS features of an image by 'filling-in' the gaps between groups of these pixels. Despite the philosophical concerns that might be expressed over this approach (which will be considered shortly), using the results of the present work to identify the CLS pixels (see figure 4.29 on page 107[3]), figure 4.7 illustrates the relative saliency of each pixel calculated for the images of figure 2.7. Observe that the results from this method are saliency measures at each pixel, and that the determination of a higher level representation would still require significant interpretation, a non-trivial task considering the image complexity. Despite these very real difficulties, it is, ultimately, the graver questions surrounding the implications of 'filling-in' the gaps (see later) that reject this method from any

[3]In fact, the process is initialised on the skeletal pixels of each 8-connected region in the CLS pixel space. See figure 4.26 for details of acquiring the skeletal pixels from the CLS pixels of figure 4.29.

further consideration.

Duncan [56] describes a technique for extracting the salient linear structure from an image by reinforcing the evidence for such features at each location from the evidence within the surrounding neighbourhood. Clearly this is a technique with parallels to the structural saliency methods just mentioned. The method is initialised with the edges from the Canny operator and uses relaxation labelling as the framework within which the evidence is assimilated. Fundamentally, this procedure is an edge detector and consequently does not apply to the present work.

Kass and Witkin [113, 114] describe a technique for identifying the 'oriented patterns' in an image. Consider the patterns of the fingerprints on your own fingers and thumbs. Observe that there is a locally dominant pattern of parallel lines, deformed from linearity into the curves and spirals of each fingerprint. This technique attempts to extract a description of the salient features in images by decoupling the local patterns and global deformations, of which the fingerprint scenario is an example. An implementation of this algorithm revealed that, as expected, not only are the image forces (CLS signal strength) simply too low, but the implicit necessity that there be an underlying regular pattern is simply not supported by the CLS structures of the breast. Although the milk ducts for instance, do in general converge at the nipple, the variability of these networks, and the CLS features in general means that no such underlying regular pattern exists, and thus the algorithm necessarily fails.

The deformable contour, or 'snake' introduced by Kass et al. [115, 116] has been adapted for applications of biomedical image segmentation [8, 29, 135], since it can readily deform to the complex structures often present in these images. A difficulty common to all applications of snakes concerns their initialisation, and within the present application, not only is there no trivial solution, there is not even an option to pursue. Aside from initialisation, there are still further difficulties associated with these methods. An intrinsic assumption of these energy minimisation methods is that the update of the snake at each iteration is driven by the location of the *greatest* image force within the orthogonal search space at each point (or control point) of the snake. If image gradients are used as the measure of the image forces, then the update of the snake will drive it towards the strongest edge within the search space. Consequently to avoid oscillations and instabilities it is necessary to ensure that the size of the update search region is smaller than the distance between adjacent edges. Whilst this approach is acceptable within the scope of many typical computer vision applications, where the width of the region of the edge during the transition from the feature to the background is generally small compared to the distance *between* adjacent features, and thus the search space of the snake can be set to an intermediate value as a compromise between speed and stability, it is simply not an appropriate model of the CLS features in mammogram images. As has been noted previously, it is quite common for some CLS features to be spatially close and parallel, with separations of only a few pixels. If a snake were able to settle on at a given location, there is no guarantee that the boundaries so identified would be from the *same* CLS feature, as two adjacent features might

each be at the *locally* dominant edge gradient at different regions of the image. As previously noted, the image forces in the present application (edge strength at CLS boundary) are small, and they render this approach completely ineffective. An implementation of this approach has confirmed these sentiments, revealing the inability to stabilise on a given CLS boundary, even if initiated moderately close to it (compared to surrounding structures). Deformable contours are an inappropriate technique for identifying a high-level CLS feature description.

Malladi et al. [141] describe a contour based segmentation procedure that overcomes the difficulties that active contours (snakes) have in describing complex structures due to the penalties imposed by high plane curvatures of the feature boundaries (such as at the intersections of CLS features in the CLS networks). The method initiates an inflating/propagating contour/wave-front that is driven by the image gradient field, which is assumed maximal at the feature boundaries. The authors include an example of identifying blood vessels in an angiogram image within which the shortcomings of the technique for the present application are well exposed. Essentially, whilst a large portion of the vessels were successfully identified, there were many that were not. Considering the high contrast of these images compared to mammography, it is clear that even the more salient CLS features would not be identified, a reflection of the weak image forces in mammograms, and the failure of this technique to respond to them.

David and Zucker [47] describe a high-level ridge detection technique that identifies the global curves of an image. The technique builds a probability density function (pdf) over the surface of an image which describes the likelihood of a given location belonging to a ridge. The high-level image curve description is extracted from the pdf by using dynamic curve segments based on snakes to identify the 'ridges' in the pdf. This is a less convoluted approach than it might initially appear since the pdf is constructed as a sum of Gaussian distributions and, to machine precision, is a continuous function with sharper peaks at the ridge-line than the original image surface. The difficulty of the approach is that the input tokens from which the Gaussian summation is driven must not only identify the pixels on or near the ridge (CLS) features, but also the local tangential direction and the plane curvature. Note that this information is required *prior* to the development of the pdf. Obtaining the tangential direction may be possible with the aid of a standard edge detector, however the difficulty of identifying those edges, that are part of the features has already been mentioned. In addition, the estimation of the curvature is even more difficult, and thus within the current application this technique cannot be initiated, and must therefore be rejected.

Another limitation that generally applies to techniques attempting to directly determine a high-level description of the CLS features is that ultimately they have an insufficiently specified low-level description. Knowing that a CLS feature exists at a given location in an image, yet not being able to identify exactly which pixels are part of the feature and which are not, does not allow the subsequent inclusion of this feature description into further analysis. This has then become the inverse of the problem as formulated for the pixel-based methods above, however it

is more problematic in the sense that if the high-level description does not accurately reflect the structures of the image, then certainly the pixels subsequently identified in a search directed by the locations of the high-level structures will also be inaccurate.

As a general comment on these high-level methods, they are to be treated with caution, since our underlying assumption, which is also a component of the philosophy behind model-based analysis, is that the input data, i.e., the original image, is sacred, and that only information that is actually present in the image can be extracted. It is unreasonable to interpolate the existence of information from the perceived incomplete data of the image, since it is impossible to know if the perception of saliency is an accurate reflection of the physical structures in a breast, degraded and complicated by occlusions and saturation in the the two-dimensional projected image of the three-dimensional breast. For example, deformable contours and structural saliency techniques, etc., aim to identify features by 'filling-in' the gaps in the structures of the image, yet it is often impossible to determine from the image alone if a given fill-in is accurate, or which of a variety of connections between terminating features should be made. Essentially, how can one say that CLS features X and Y, which "look like they are the same feature except for that small gap", are actually the same feature vignetted or occluded by some obstruction? The 'gaps' which these algorithms seek to bridge are (since an implementation will naturally reflect the introspection of its creator) merely 'apparent' gaps which are a function of the perception of the human visual/reasoning system and in no way necessarily relate to the realities of the original image data. It is the philosophy of the present work that such questions cannot be answered, and to implement an algorithm that introduces such techniques is not only inappropriate, but considering the subject matter, quite unethical, such are the implications of inadvertently introducing artifacts that have unknown effects on any subsequently attempted image analysis/interpretation.

Since it is not clear how a high-level CLS feature description can be identified directly without a prior description of the pixels involved, and any subsequent application of the CLS feature description would involve a knowledge of the actual pixels involved, it would seem the pixel-based philosophy of initially determining the CLS pixels and then extracting a high-level description of the CLS features was a superior approach. Returning to this perspective then, the task of identifying the CLS features logically centres around accurately identifying the CLS pixels in the first instance. Clearly the results of Fleck's *Spectre* detector indicate that the CLS features appear to respond to measures of the second-difference of the image surface. It is to this point that we shall return during the development of the CLS pixel detector given in section 4.4. Supposing for a moment that a low-level description of the CLS pixels were available, it is then our task to extract a description of the high-level CLS features, a task considered in the next section.

4.3 The CLS detector algorithm

The algorithm developed in the present work to extract the CLS features is presented in this section in its most generic form, stripped of all the implementation details which have been deferred to following sections.

Consider a binary image segmented into *foreground* and *background* regions. Without loss of generality, in order to preserve the topology of a feature in the discrete digital representation of the image, we choose the common system [65, 130] of 8-connectivity for the foreground and 4-connectivity for the background. In the description of the algorithm which follows it is clear which pixels should be considered as foreground at any given time—they are members of the sets **N, S, C, B** and **CLS** (see below) whilst the particular set is under consideration (the background is implicitly defined as the conjugate of the respective set under consideration)—and accordingly these details are often omitted to avoid unnecessary clutter.

Consider an original grey-level mammogram image I. The algorithm for extracting the curvi-linear structures is:

1. Find the set **N** of all pixels $\mathbf{n}_i$ in the image I that are components of the CLS features.

2. Fill small bounded holes in set **N** by including the pixels of the holes in set **N**.

3. Thin the regions of set **N**, retaining in a new set **S** the pixels that comprise the simply-connected *skeletons* of **N**'s regions.

4. Break the skeleton into its segments by scanning the set **S** collecting into *cantons* $\mathbf{c}_i$ those 8-connected pixels that collectively share a common global heading/direction. At each branch (or junction) in the skeleton initiate another canton. Form the set **C** of all cantons $\mathbf{c}_i$ that result from classifying all $\mathbf{s}_i \in \mathbf{S}$.

5. Scan the set **C**, merging into *bones* $\mathbf{b}_i$ those cantons with both a (nearly) common endpoint and a sufficiently similar heading. Scan the bones $\mathbf{b}_i$ searching for merges involving a short intermediate canton joining two dissimilar cantons. Exclude the least convincing of the two merges, returning a junction at this location. At the completion of this phase, those cantons that are not merged with others to form bones are transcribed directly to the set **B** creating a new bone for each such canton.

6. Search the set **B** for the known irrelevant branches of the skeleton and reject them[4].

7. Search the set **N** in the region surrounding each bone $\mathbf{b}_i \in \mathbf{B}$, collecting the 8-connected neighbourhood pixels in the new set $\mathbf{cls}_i$ associated with $\mathbf{b}_i$.

8. Declare the set **CLS** of all $\mathbf{cls}_i$ (with corresponding bones $\mathbf{b}_i$) to be the curvi-linear structures of the original image I.

[4] In practice, the details of this step are highly dependent upon the output of the skeletonisation algorithm (step 3 of the algorithm), and in some instances this step may not be necessary.

Clearly, this simple description suppresses details that are crucial to the success of the algorithm (e.g., the precise definition of **N**). The algorithm design has evolved through a mix of analysis, introspection and experimentation. It is justified not on the grounds that it may be a plausible model of how humans extract such features manually, but on the grounds that it succeeds at the task it was formulated to deal with since the resulting CLS features:

- accord well with what radiologists perceive, and
- support subsequent processing, as described in the chapters which follow.

In the following sections, we analyse the less obvious steps of the algorithm presented here, developing theoretical models where necessary and providing sufficient information to allow an implementation of the algorithm to proceed directly. The following sections mimic the sequential order of the CLS detector algorithm presented here, and whilst each section is self-contained, they often draw on information developed in the preceding sections and the reader is encouraged to work through this material sequentially.

4.4 Finding the CLS pixels — Low-level description

4.4.1 Modelling the expected CLS profile

Before we can build a detector capable of finding the CLS pixels in an image, we need to develop a model of the expected CLS footprint so that we can implement a suitable strategy to locate them. Consider the anatomy of the CLS features. The milk-ducts and blood vessels are tubes, of generally (nearly) elliptical cross-sections, particularly in the compressed breast as presented at mammography. The fibrous tissues are not of any specific cross-sectional shape, although it would not be unreasonable to also classify them as generally near-elliptical (indeed, there are 2-dimensional 'sheets' of fibrous tissue—e.g., the Cooper's Ligaments—which do not fit this description in a 3-dimensional Euclidean sense, however in the sense of the X-ray attenuation cross-section of the CLS, this approximation is not unreasonable).

Consider the model of a section of homogeneous breast tissue of thickness H with an X-ray attenuation coefficient μ_1, as shown in figure 4.8(a). The incident X-ray beam in a modern mammographic imaging device is poly-energetic of the form shown in figure 4.9. To accurately analyse the attenuation of this beam as it traverses our model tissue sample, it is necessary to integrate the following analysis over all beam energies, which is analytically an intractable problem given the non-linearities of the beam-energy profile. Thus within the scope of the present analysis we assume a mono-energetic incident beam[5] of energy E_o, such that

$$E_o = \phi X_c \mathcal{E}_p$$

where $\mathcal{E}_p$ is the energy per X-ray photon (the beam 'power' if you will, (keV)), X_c describes the X-ray tube current and exposure time (in milliampere-seconds (mAs)) which effectively is

[5]The numerical simulation which follows this section does integrate over the beam energies of a typical poly-energetic beam, and so avoids this assumption.

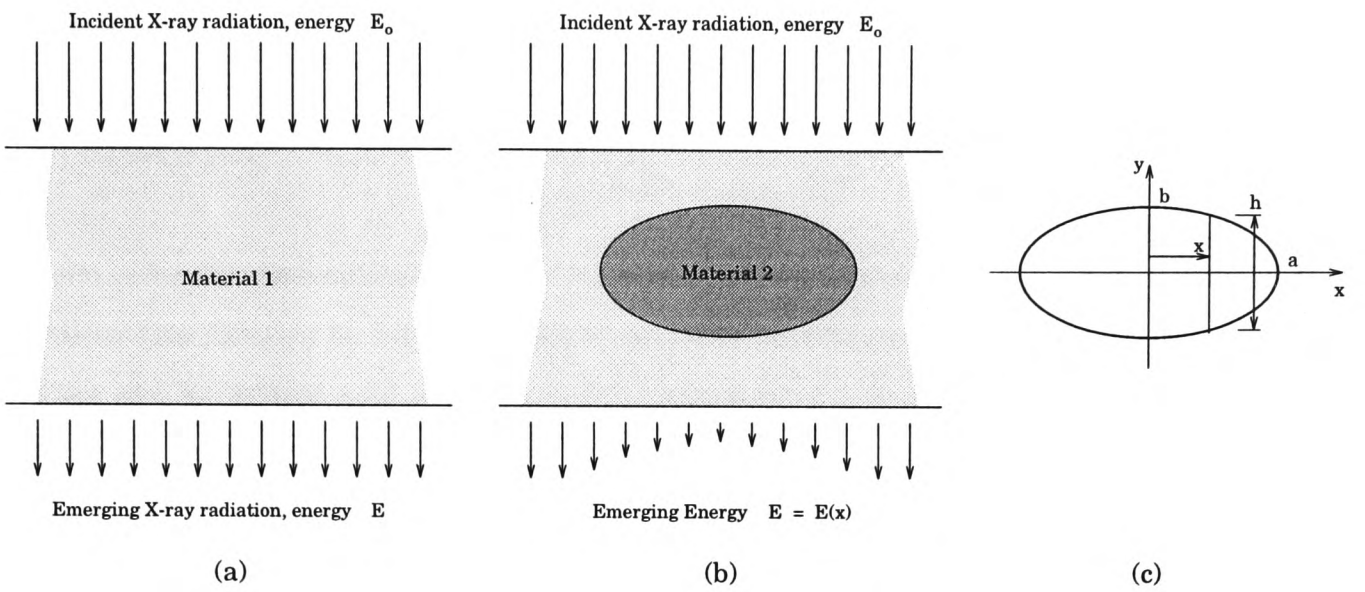


Figure 4.8: Modelling the X-ray attenuation of the CLS features. (a) The 'background' tissue, (b) The CLS feature modelled by an ellipse in the compressed breast, (c) an ellipse with major and minor radii a and b respectively. The model assumes that the CLS feature is 'long' in the direction orthogonal to the page surface relative to a and b, allowing the response of the cross-section to accurately represent any point along the CLS.

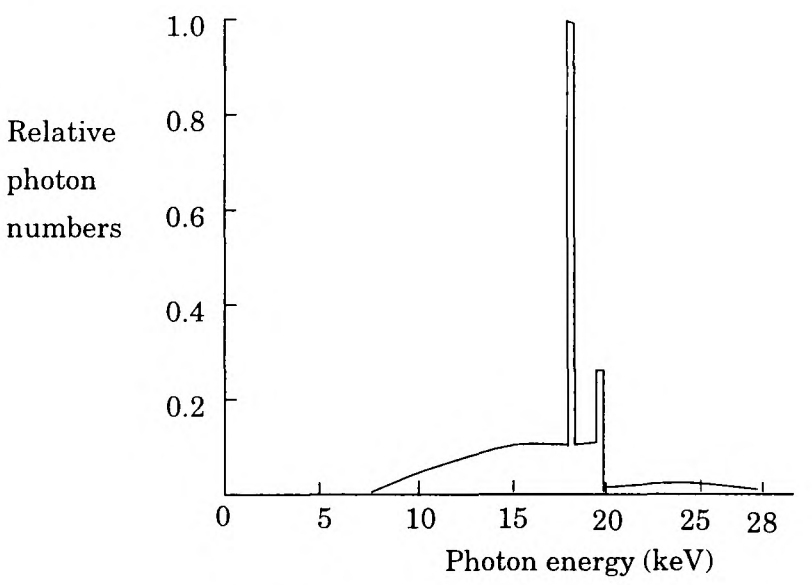


Figure 4.9: An X-ray spectrum typical of the beam produced by mammographic X-ray tubes (with tube voltage 28kVp, Mb anode, 0.8mm Be and 0.03mm Mb filters).

a measure of the *number* of photons in an exposure, and ϕ is a constant of proportionality. The assumption of a mono-energetic beam equates to setting $\mathcal{E}_p$ to a constant value. Given the relative occurrence of photon energies in a typical incident beam (see figure 4.9) we choose $\mathcal{E}_p = 17.4$keV during the following analysis.

The energy of the X-ray beam emerging from the sample has two components, the *primary* energy E_p that has traversed a straight path through the material from the X-ray gun to the imaging device (and been absorbed along the way), and the *scattered* energy E_s which has been scattered within the sample. The absorption of the primary beam can be described by $E_p = E_o \exp(-\mu_1 H)$. Analytically describing the scattered component of the beam energy arriving at a given point in the imaging plane (the plane at which the film is exposed) is very

difficult. Thus we assume in the present analysis that there is no scatter component in the beam energy exposing the film[6]. Thus we can write:

$$E = E_p + E_s = E_o e^{-\mu_1 H}.$$

Now consider the introduction of a CLS feature into the cross-sectional model of the breast tissue. Allow the feature to have an elliptical cross-section, since that is the model of the CLS features we are adopting, and allow it to be composed of a material with an X-ray attenuation coefficient of μ_2, as shown in figure 4.8(b).

From the general equation of an ellipse with major and minor radii a and b respectively (see figure 4.8(c)), and assuming a collimated incident X-ray beam we can derive the path length the beam traverses through the feature as:

$$h(x) = \frac{2b}{a}\sqrt{a^2 - x^2}, \quad -a \leq x \leq a, \tag{4.3}$$

and the net beam attenuation for the combined sample of figure 4.8(b) as

$$E(x) = \begin{cases} E_o e^{-\mu_1(H-h(x))} e^{-\mu_2 h(x)} & -a \leq x \leq a, \\ E_o e^{-\mu_1 H} & \text{otherwise.} \end{cases} \tag{4.4}$$

Once the beam has been attenuated during its travel through the breast (or in this case, the cross-sectional sample) there are a number of stages before the digitised image is available. Each stage must be considered so that the absolute pixel intensities of the image, and thus the image surface, can be estimated.

The film-screen apparatus can be modelled by considering a typical film-screen characteristic curve as shown in figure 4.10. By utilising the linear portion of the curve, we can model this as:

$$D = \gamma \log_{10}(\beta E) \tag{4.5}$$

where D is the film density, γ the film gradient, β is related to both the film speed and (within the mono-energetic assumption) the anti-scatter grid and intensifying screen response, and finally E is the X-ray energy imparted to the film-screen system.

To estimate the image intensities resulting from the digitisation of the film, we now consider the digitisation process itself. There are a number of different ways to digitise a film, however they generally can be classed as those that quantise the film density D, and those that quantise the light transmitted through the film T_l for a given back illumination I_l. These representations can be related by the following equation:

$$D = \log_{10}\left(\frac{I_l}{T_l}\right). \tag{4.6}$$

The direct digitisation of the film density (for example by a scanning microdensiometer) returns image pixel values $P_d(x, y)$ in an approximately linear relationship with D as:

$$P_d(x, y) = mD(x, y) + q \tag{4.7}$$

[6]Once again, the numerical simulation to follow does model the scattered components of the beam energy.

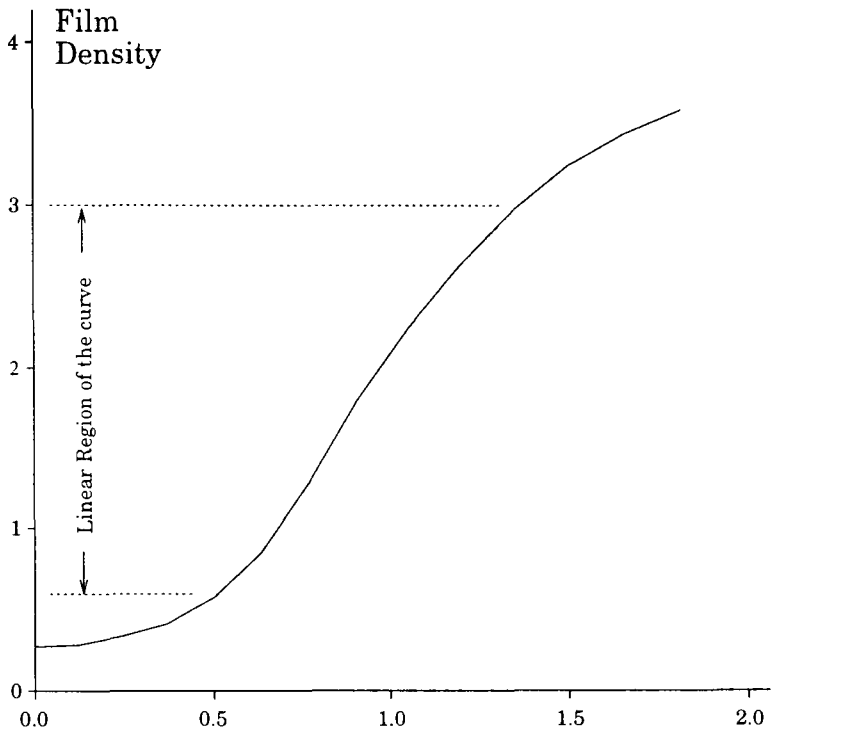


Figure 4.10: Typical film-screen characteristic curve.

whilst the digitisation of the transmitted light (for example using a CCD camera and a film back lit on a light-box) relates the image intensity $P_l(x, y)$ with T_l as:

$$P_l(x, y) = \alpha T_l(x, y) + \lambda$$

where m, q, α and λ are digitisation calibration constants. Substituting equation (4.6) yields:

$$P_l(x, y) = \alpha I_l 10^{-D(x,y)} + \lambda. \tag{4.8}$$

At this point the analysis of the overall system diverges down the respective paths headed by the choice of image representation. Many people advocate the use of the film density as the domain of choice for analysing mammogram images since it can be linearly equated to the raw data of the film. Although the clinical practice is to view a mammogram mounted on a light-box, and thus observe the transmitted light T_l, it is believed the human visual system responds logarithmically to light intensities falling on the retina [79, p85]. In effect the human visual system implements the logarithmic transform of equation (4.6) and thus responds linearly to the film density D. Since the clinical domain of analysis for mammogram images has to date been the human visual interpretation of light-box mounted films, it is naturally within the experience of this domain that the interpretative radiological techniques and rules for reading films have evolved, and it is for this reason that this domain is often advocated as the domain-of-choice for automated analysis systems. However, since it is possible to form digital images with intensities representing either film-density or transmitted-light (and to convert between them with relative ease by the application of equation (4.6)), we proceed by analysing the consequence of both digitisation methods within the context of the present task of finding the CLS features.

Tissue type	$\mu(\text{cm}^{-1})$	Beam energy (keV)		
		17.4	20	25
Fat	Minimum	0.538	0.441	0.314
	Mean	0.558	0.456	0.322
	Maximum	0.585	0.476	0.333
Fibrous	Minimum	1.014	0.791	0.499
	Mean	1.028	0.802	0.506
	Maximum	1.045	0.816	0.516

Figure 4.11: X-ray linear attenuation coefficients for 'fat' and 'fibrous' breast tissues, from Johns and Yaffe [106].

Consider initially the direct digitisation of the film-density D given by equation (4.7). Substitution of equations (4.3), (4.4) and (4.5), and restriction of the domain to the region of interest, i.e., $-a \leq x \leq a$, yields:

$$P_d(x) = \frac{m\gamma}{\ln 10}\left(\ln(\beta\phi X_c \mathcal{E}_p) - \mu_1 H - (\mu_2 - \mu_1)\frac{2b}{a}\sqrt{a^2 - x^2}\right) + q. \tag{4.9}$$

From Fleck's *Spectre* detector (see section 4.2) it is clear that there is useful information in the second-difference of an image for finding features such as the CLS features. In the continuous sense this equates to the second derivative, or curvature of the image surface. Differentiating equation (4.9) gives:

$$\frac{d^2 P_d}{dx^2} = \frac{2m\gamma}{\ln 10}(\mu_2 - \mu_1)ab(a^2 - x^2)^{-3/2}. \tag{4.10}$$

Within the assumptions of the model developed in this section (mono-energetic beam, no scatter, no noise), equation (4.9) returns the *absolute* pixel intensities in the digital image across the section for a feature with a cross-section as shown in figure 4.8(b) (subject to the intensity scaling due to the imaging parameters β, ϕ, X_c, $\mathcal{E}_p$ which are functions of H and μ_1), while equation (4.10) returns the line-curvature of the section[7].

In order to analyse the expected response from these equations, we insert typical values for the constants as given in figure 4.12. To obtain the best results possible we use the linear attenuation coefficients for various breast tissues reported by Johns and Yaffe [106] and tabulated in figure 4.11. Setting the substrate of the model to an attenuation coefficient of 'fat' and the CLS feature to that of 'fibrous' tissues is the most obvious approach since this reflects the actual composition of the CLS features and the 'background'. The digitisation constants m and q can be measured by solving the set of two simultaneous equations resulting from substitution of known (film-density, digitised-value) data pairs obtained from calibration data. Typically digitisation at 8-bit resolution utilising the linear region of the film-screen characteristic curve (figure 4.10), gives two data points at $(0.6, 255)$ and $(3.0, 0)$, yielding $m =$

[7] Whilst it is the surface-curvature in the two-dimensions of the image that is considered in the real case, we assume that the surface variance orthogonal to the cross-sectional plane is small compared to the variation within the plane, and thus the image surface curvature will deviate only slightly (if at all) from the line-curvature calculated here.

Constants	*Value*
Film density range	$0.6 \leq D \leq 3.0$
Image intensity range (8-bit)	$0 \leq P \leq 255$
Film gradient	$\gamma = 3$
Photon energy (keV)	$\mathcal{E}_p = 17.4$
Attenuation coefficients (cm^{-1})	$\mu_1 = 0.558$, $\mu_2 = 1.028$
Breast thickness (cm)	3.5
Size of feature (cm)	$a = 0.10$, $b = 0.02$

Figure 4.12: Parameters for calculation of the CLS model surface equations.

-106.25 and $q = 191.25$. Obtaining a value of ϕ (actually $\beta\phi X_c\mathcal{E}_p$) is harder, since it is simply a scaling factor that accommodates the different imaging parameters. This scaling factor is necessary as it allows similar images to be recorded (i.e., remain within the non-saturated regions of the film-screen recording system) for breasts of quite different compressed thickness and density/attenuation (i.e., a thicker breast needs a longer exposure to give a similarly satisfactory image). Without loss of generality, $\beta\phi X_c\mathcal{E}_p$ is set by looking ahead to the results of the numerical simulation and assigning it a value that gives equal background pixel intensities for the analytical and simulated results (effectively assign it the same value that was used in the simulation, which was taken from a particular known set of imaging data gleaned from calibration of the experimental setup). This then allows the analytic model to be verified against the simulation by directly comparing the differential pixel intensities due to the CLS's attenuation of the beam. Finally, evaluation of equations (4.9) and (4.10) for P_d and P_d'' yield the curves of figure 4.13. A discussion of these curves is deferred until the equivalent curves for the transmitted-light image representation have been developed.

We now return to the alternative approach of analysing the expected surface curvature using image intensities based upon the light transmitted image representation. Proceeding as above, substitute equations (4.3), (4.4) and (4.5) into equation (4.8). Again, restrict the domain to the region of interest, i.e., $-a \leq x \leq a$, to give:

$$P_l(x) = \frac{\alpha I_l}{(\beta\phi X_c\mathcal{E}_p)^\gamma} \exp\left[\gamma\mu_1 H + \gamma(\mu_2 - \mu_1)\frac{2b}{a}\sqrt{a^2 - x^2}\right] + \lambda \tag{4.11}$$

Differentiating twice with respect to x yields:

$$\begin{aligned}\frac{d^2P_l}{dx^2} &= \frac{\alpha I_l\gamma(\mu_2 - \mu_1)}{(\beta\phi X_c\mathcal{E}_p)^\gamma}\left(\gamma(\mu_2 - \mu_1)2bx^2\sqrt{a^2 - x^2} - a^3\right) \times \\ &\quad \frac{2b}{a^2(a^2 - x^2)^{3/2}} \exp\left[\gamma\mu_1 H + \gamma(\mu_2 - \mu_1)\frac{2b}{a}\sqrt{a^2 - x^2}\right]\end{aligned} \tag{4.12}$$

As before, substitute the typical values of figure 4.12 for the constants of the equations to yield numerical examples. Once again the calibrated digitisation constants can be estimated (using equation (4.8)) to give $\alpha I_l = 1019.2$ and $\lambda = -1.0192$.

Figure 4.14 shows the results of the numerical evaluation of equations (4.11) and (4.12). From figures 4.13 and 4.14 it can be seen that the image surface corresponding to a CLS

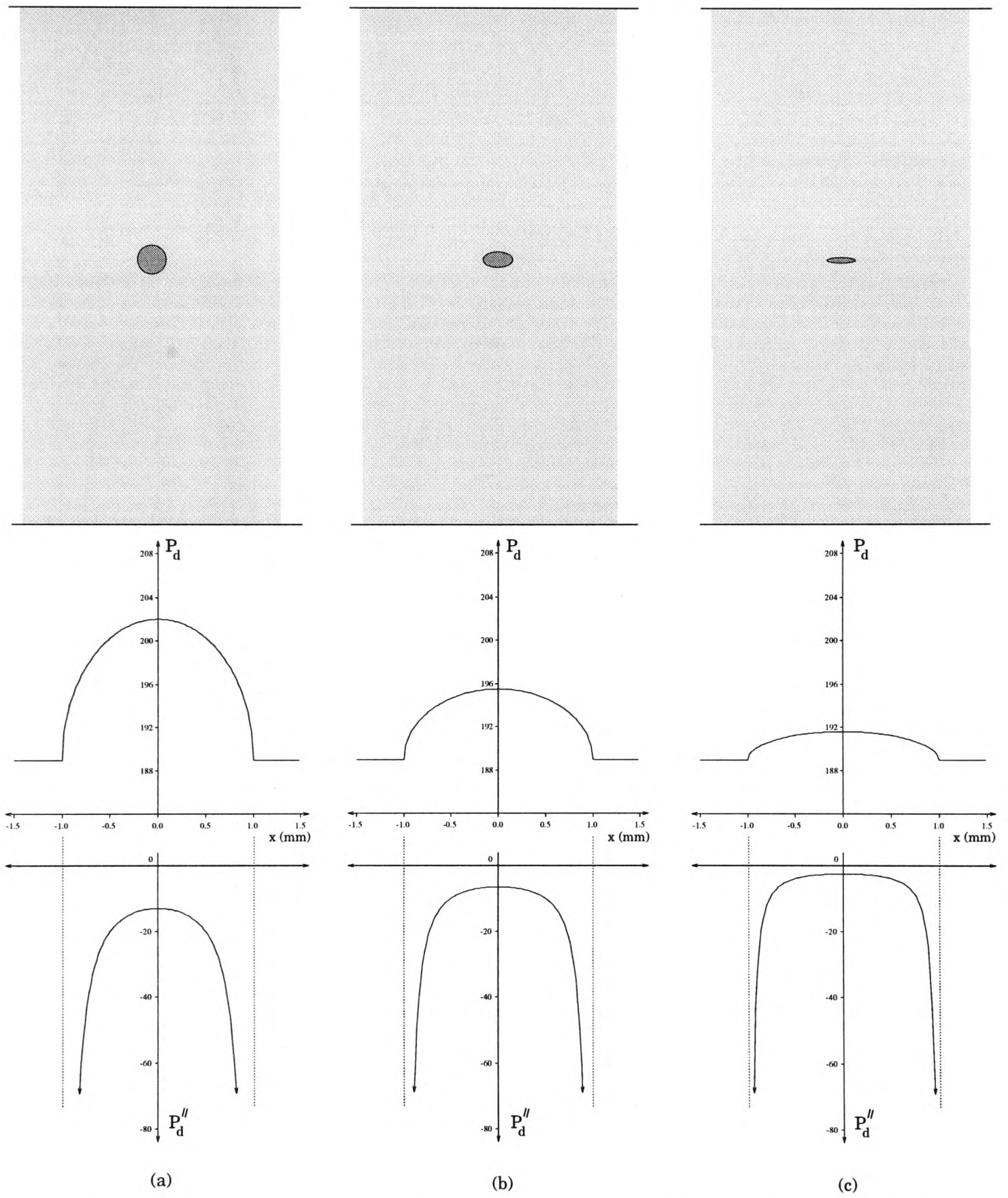


Figure 4.13: The image and second-derivative curves for idealised elliptical CLS features where intensity represents film-density. In all cases the major radius of the ellipse is $a = 1mm$, however the minor radius is (a) $b = 1mm$, (b) $b = 0.5mm$ and (c) $b = 0.2mm$. The top row of each column shows the cross-sectional proportion of feature to breast-tissue, the middle row shows the image pixel intensity values expected across the feature, and the bottom row shows the second-derivative of the intensity curve above it.

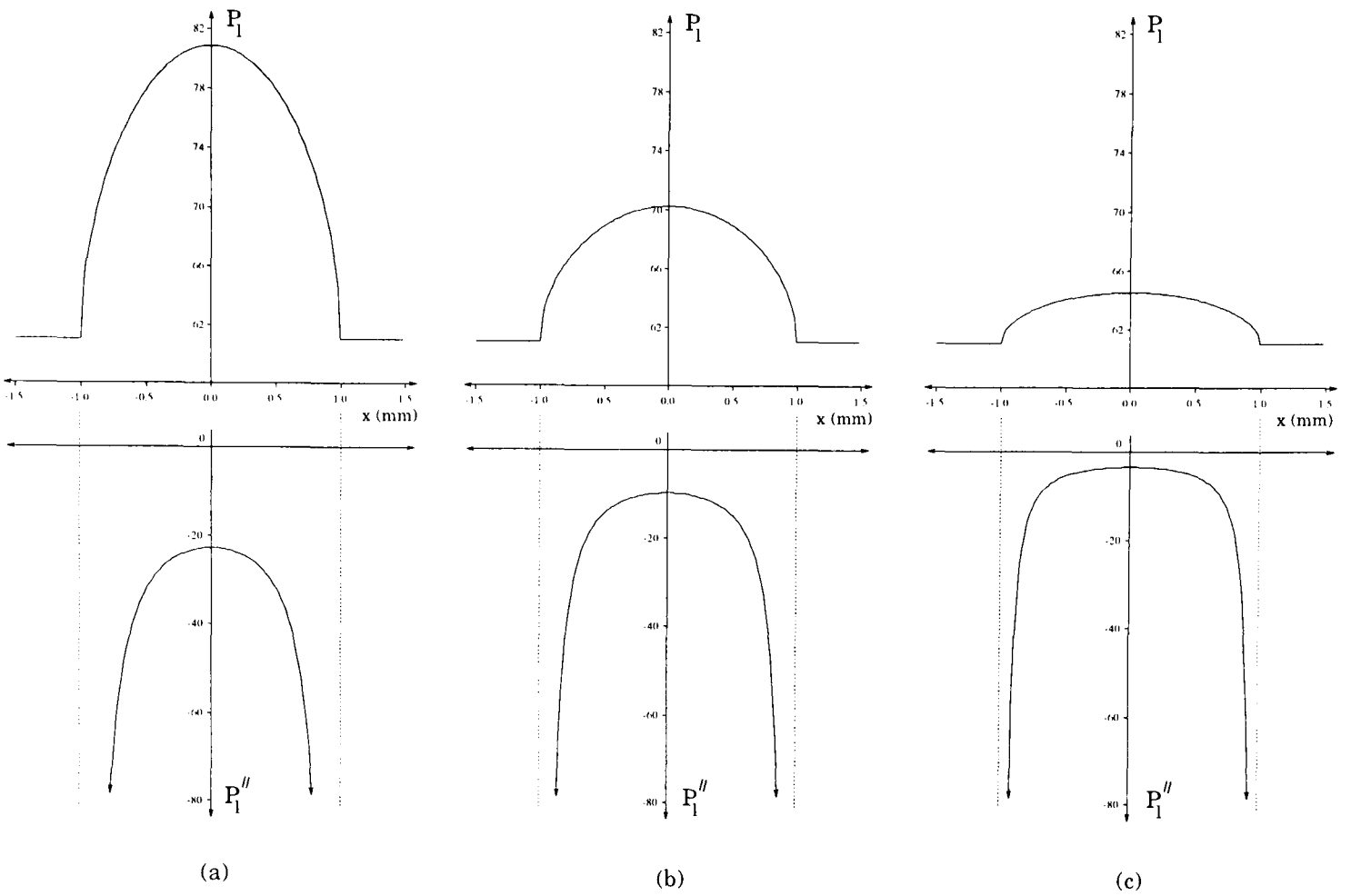


Figure 4.14: The image and second-derivative curves for idealised elliptical CLS features where intensity represents transmitted-light. In all cases the major radius of the ellipse is $a = 1mm$, however the minor radius is (a) $b = 1mm$, (b) $b = 0.5mm$ and (c) $b = 0.2mm$. The top row of each column shows the image pixel intensity values expected across the feature, and the bottom row shows the second-derivative of the intensity curve above it. Refer to the top row of figure 4.13 for a view of the shape and relative proportion of the feature in each case.

feature has a strong second derivative in the continuous domain and this may be useful in detection of the CLS features. Note that the magnitude of the transmitted-light representation is larger that of the film-density (that is $|P_l''| > |P_d''|$) due to the more 'pointed' peak of the P_l curve compared to P_d (compare P_l in figure 4.14 with P_d in figure 4.13). Since it is the magnitude of the second difference that would be useful for finding the CLS features, this initial assessment indicates that the transmitted-light representation would be a superior domain in this particular case.

Before endeavouring to find the CLS features in an image based on this observation, there remains a number of issues to address. These include: determining the effects of moving the analysis from the current continuous domain into the discrete digital domain; whether the assumptions of the current model—in particular the assumptions of an elliptical CLS cross-section, mono-energetic beam and non-scattered radiation—affect the conclusions in practice; and finally in which of the two commonly available image domains—film-density or transmitted-light—should the search for CLS features be conducted?

4.4.2 From continuous to discrete image surfaces

To determine the effect of moving to the discrete digital domain of analysis, consider the models developed by Highnam [97]. Highnam builds a suite of models which analyse the various components and effects during the mammogram imaging and digitisation processes, including the X-ray tube/beam characteristics, the anode-heel effect, beam-hardening, scattering, and then finally the digitisation. Highnam [101] has developed a numerical implementation of these models which allows any situation (combination of the above listed effects and breast geometry and composition) to be simulated, returning the *digital* image that would have been created in reality. The digital image can be furnished in either the film-density or transmitted-light domains.

Initially this method is applied with the conditions of a mono-energetic beam, non-scattered radiation and zero noise as specified during the development of the CLS modelling procedures of the previous sections. Thus the simulation returns the digital image surface that would be obtained if the sample of figure 4.8(b) were imaged and digitised, allowing a direct comparison between the discrete digital results obtained and the results of the continuous analysis of the previous section. Figure 4.15 shows the simulated intensity profile for a feature with major and minor radii of $1mm$ and $0.5mm$ respectively, imaging constants of figure 4.12, and at digitisation resolutions of $50\mu m$, $125\mu m$ and $250\mu m$ square per pixel. The specific digitisation resolutions shown were chosen since $50\mu m$ and $125\mu m$/pixel are commonly available with current digitisation techniques, and $250\mu m$/pixel is a size convenient for analysis of CLS features as mentioned in section 4.1.2.

This digital image surface can be directly compared to the continuous surface (figures 4.13 and 4.14) given by equations (4.9) and (4.11). To facilitate this comparison, figure 4.15 includes the respective continuous curve (the dotted curve) for each case.

Figure 4.15 shows the (simulated) digital image surface resulting from digitisation of both film-density and transmitted-light representations. Note that although there is quite a difference between the continuous image surfaces (dotted lines) of the two representations, the digital surfaces do not reflect the difference in this instance, that is, the digital curves of the film-density representation are *exactly* the same as those of the transmitted-light representation. This inability to resolve the difference between the two digital representations is due mostly to the relatively large intensity quantum. Although the assumptions of the present model (in particular that of no noise) would accommodate a refinement of the intensity quantum (i.e., digitise the image to greater depth, say 12-bits) to more faithfully represent the respective image domains in the digital case, in practice this is not possible. Recall from section 4.1.3 that a typical image digitised to 8-bit resolution from a film-screen mammogram film contains a total noise component of variance $\approx$3–4 intensity levels, mostly resulting from the film-screen imaging processes, rather than the digitisation. Although improvements in digitisation technology may have the capacity to more accurately digitise a signal at the expense of introducing relatively little noise, the digitised signal will always be limited by the noise contained in the

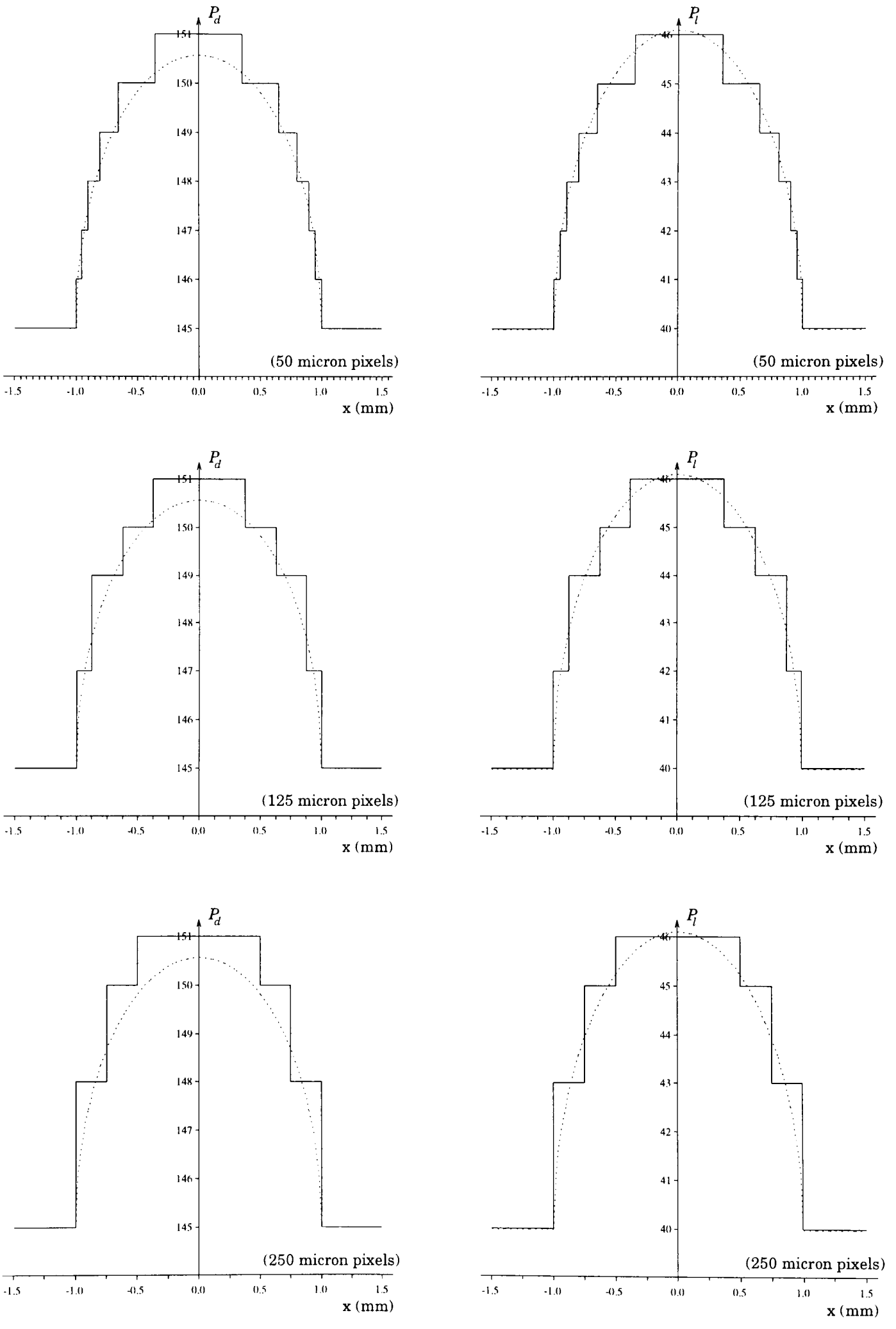


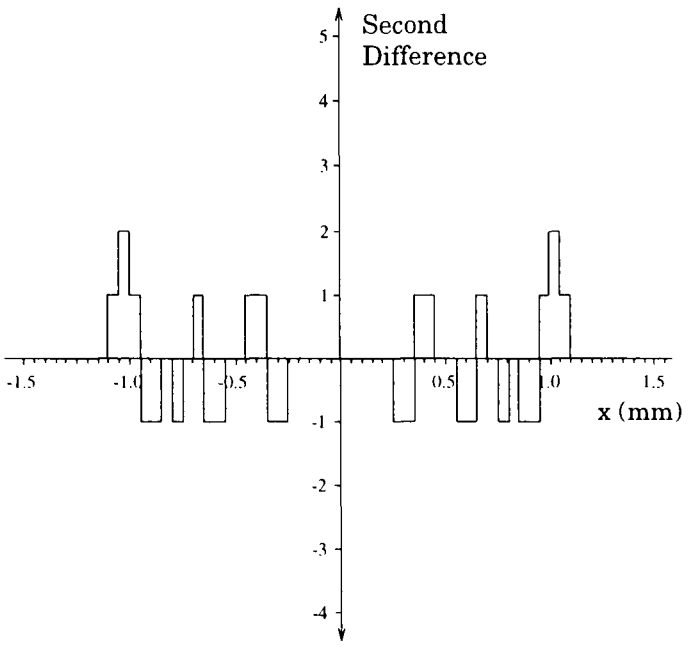
Figure 4.15: Comparison between the analytical (continuous) expected image intensities (dotted lines) verses the simulated discrete approximation (solid lines) for an elliptical CLS feature with major and minor radii $1mm$ and $0.5mm$ respectively. The top row shows the simulation at a resolution of 20 pixels/mm (50μm pixels), middle at 8 pixels/mm (125μm pixels) and bottom at 4 pixels/mm (250μm pixels). Note that the x-axis ticks indicate the pixel size.

original signal. Karssemeijer and van Erning [112] have shown that CCD sensor noise (and digitisation noise in general) is small compared to the noise present in the original analogue signal of the mammogram film. Thus an attempt to digitise a film to greater depth may simply measure the film-noise with greater accuracy, and therefore may not capture the image surface characteristics as desired. Consequently, we have chosen to confine the analysis of the present work to images digitised at the routinely available digitisation depth of 8-bit. Due to the lack of suitable hardware, it was not, within the scope of the present work, possible to digitise a mammogram film at both 12-bit and 8-bit depths to assess the difference that this factor might have.

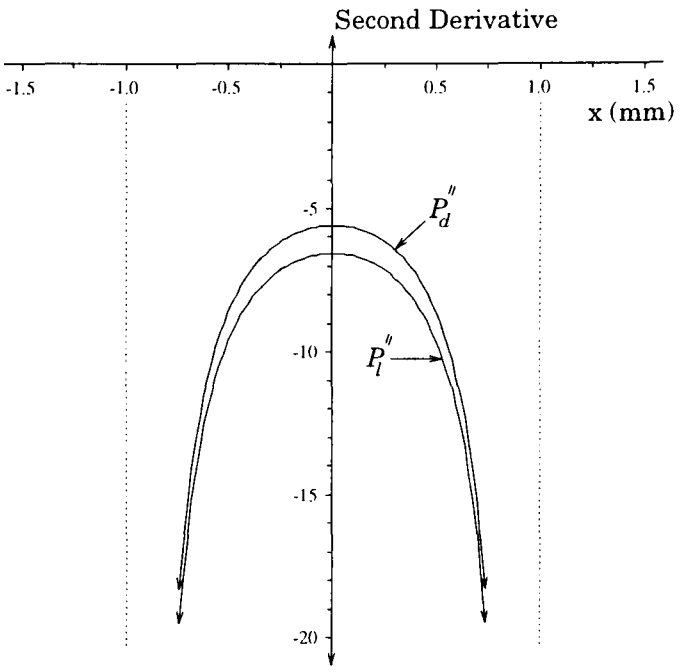
Now that a discrete approximation to the expected image surface profile has been obtained, it is a simple matter to calculate the surface curvature for the discrete case, by taking second differences (see section 4.4.8 following) and compare it with the continuous second derivative. Since there is no difference between the image surface of the film-density and transmitted-light representations, the discrete image surface curvature for each case will also be equal. Figure 4.16 shows the second difference responses obtained from application of the Δ''_N kernel of figure 4.23 to the image surface profiles of figure 4.15.

Note that the discrete second difference curves of figures 4.16(a), (b) and (c) are quite different to those of the continuous case of figure 4.16(d). Observe however, that the second difference responses more closely resemble the continuous case (excepting the positive second difference response in the tails of the discrete case due to the spatial extent of the Δ''_N kernel) as the pixel size is increased from $50\mu m$ through $125\mu m$ and finally to $250\mu m$/pixel. There are a number of forces at work here—namely the digitisation resolution, intensity quanta size, the spatial extent of the feature and the spatial extent of the filter kernel—and they each contribute to the observed responses just described. At the current digitisation quanta size (8-bit), feature size ($2mm$ wide and $1mm$ thick) and kernel extent (5 pixels) an image resolution of $250\mu m$/pixel gives the best results of the resolutions shown. If further image digitisation depth were available, and in the absence of noise, one would expect the smaller image resolutions to give progressively better results. Similarly the smaller image resolutions would progressively improve if the feature size was reduced in width, or the feature thickness *increased* (giving greater beam attenuation and therefore better image contrast, and ultimately producing much the same results as for increased digitisation depth!). Alternatively the second difference kernel size itself could be adjusted to give either better performance for variations in image resolution given a constant feature size under investigation, or more importantly, given a fixed image resolution, adjusting the filter size will give responses matched to specific feature sizes. This important concept of kernel scale-space and its effect on the kernel's response to a range of feature sizes is noted only in passing at this stage. A fuller discussion of this concept is deferred to section 4.4.8.

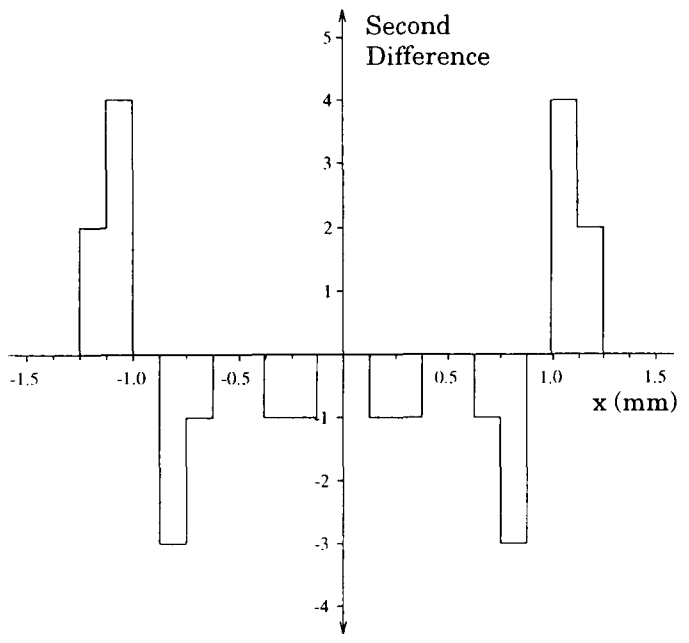
In conclusion it is noted that for a given digitisation resolution and intensity quanta it is possible to tune the second difference kernel size to return second difference responses



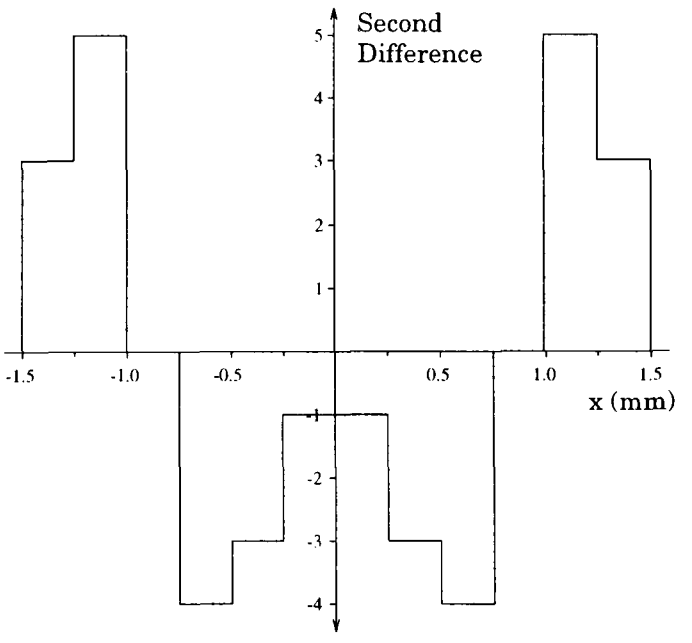
(a) 50 micron pixels



(d) continuous



(b) 125 micron pixels



(c) 250 micron pixels

Figure 4.16: Comparison between the calculated second differences of the discrete case and the second derivative of the continuous model.

consistent with the continuous model developed in the previous sections. Therefore if the real CLS features in an image can be adequately represented by the model so developed, then the second difference (or curvature) of the image surface can be used to locate the CLS features.

4.4.3 Is the elliptical CLS model adequate?

The calculated image surface profiles and image gradient profiles developed in the previous sections are derived from a model which assumes that the CLS features in a real breast are elliptical in cross-section. Naturally the validity of the results which have followed from adopting this model is dependent upon the ability of the elliptical model (figure 4.8) to faithfully represent the actual CLS cross-sections.

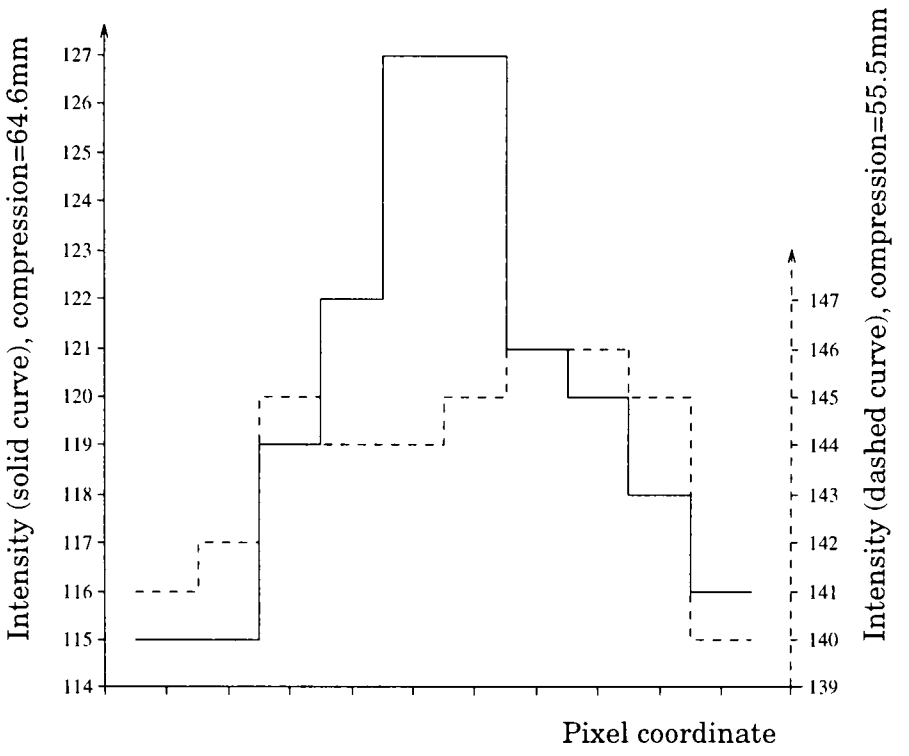


Figure 4.17: Comparison of image surface profiles for a given CLS feature identified in two mammograms of the same breast acquired at different compressions. The solid curve and solid left-hand y-axis indicate the surface profile for the CLS at light breast compression, whilst the dashed curve and dashed right-hand y-axis are for the breast at greater compression. Observe that the profiles have been shifted vertically to match their base-lines (note that this shift is accommodated directly within the *constants* of equation (4.9) and thus plays no part in the image curvature, and are therefore irrelevant within the context of this comparison).

Further to the initial arguments presented in support of this premise during the development of the elliptical model in section 4.4.1, consider the retrospective empirical evidence now available by comparison of the model with real CLS features extracted from an image. To facilitate this analysis, consider the 'differential-compression' techniques of Highnam et al. [102]. They suggest acquiring mammograms of the same breast at different compressions and then using the mechanical deformation of the tissue as a cue to aid the detection of local disease processes. Rather than adopting the procedure of Highnam et al. for detection of disease processes, extract from two mammograms of the same breast at different compressions the cross-sectional profile of a given CLS. For this particular situation we solve the correspondence problem by eye, returning the cross-sectional profiles for a CLS feature in each image deemed to be the same CLS feature by a radiologist.

As compression is increased, we would expect a given CLS feature to spread out and get thinner, that is, the major radius a of the elliptical cross-section would increase, whilst the minor radius b would decrease. Consequently we would expect the image surface profile orthogonal to the CLS direction would reflect these changes if indeed the real CLS features were deformable. Figure 4.17 shows the CLS surface profiles extracted from the images of a breast imaged at different compressions. The solid line is at low compression, whilst the dashed line is at higher compression. Thus we would expect the dashed line to have lower overall magnitude and have wider spatial extent than that of the solid line, and indeed this is the case.

Note that this only shows that the particular CLS chosen was a deformable body consistent

with the notion of a blood-vessel or milk-duct (in fact the particular CLS chosen was a milk-duct). For completeness it is noted that a similar analysis for the fibrous stromal tissues produces far less deformation of the surface profiles, again consistent with the expected stiffer nature of the fibrous tissues.

Still the question of the validity of the *elliptical* cross-section remains. The real question is: does it matter if the actual CLS is elliptical or not, provided it has some basic curvature? Clearly the answer to this question is that it does not matter. If the image surface has some negative curvature, then that is sufficient. That it comply with the curvature of an ellipse is unnecessary.

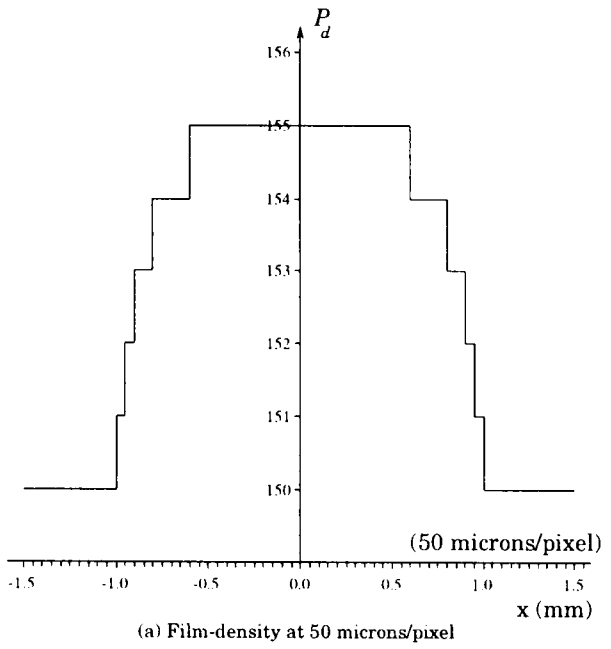
This last point is further aided by choosing to reduce the images to $250\mu m$ pixel resolutions since the majority of CLS features will then only be a few pixels in size. In this situation there are a large range of continuous image surface shapes (including the elliptical one of this model) that will quantise into an image surface with negative curvatures. Consequently the model is in fact more stringent than it needs to be for the purposes of detecting such features, and therefore any necessary relaxation of the modelling to more accurately describe the precise shape of the CLS image surface profiles can be accommodated. Obviously adopting this approach means that all regions of negative surface curvature in an image will be identified as CLS features. An important consequence of this is that features such as edges (for instance at the border of densities) will also be identified by this procedure since these regions also give rise to regions of negative image surface curvature. This point will be considered further in section 5.1.

4.4.4 Poly-energetic beams and scattering

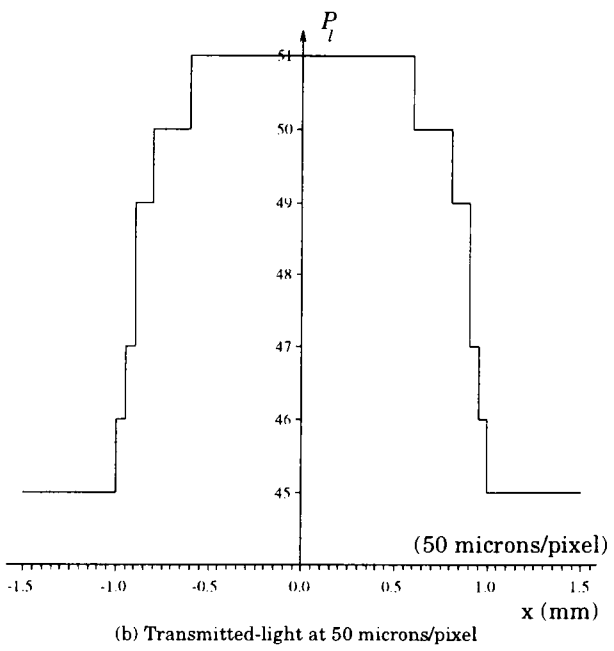
The previous sections have established the applicability of the model of the CLS cross-section and the use of the image surface curvature, as calculated by the second difference, for detection of the CLS pixels within the assumptions of a mono-energetic beam and the absence of any scattered radiation component in the energy imparted to the film during exposure. In reality however, the X-ray beam is a poly-energetic beam of the form shown in figure 4.9, and despite the use of anti-scatter grids, there is a scattered radiation component of the energy incident on the film. To analyse the impact these effects have on the analysis conducted so far, consider the numerical simulation methods of Highnam referred to in section 4.4.2.

In addition to the mono-energetic and scatter-less simulation of that section, Highnam's methods have the capacity to fully model and simulate the effects of poly-energetic beams and scattered radiation. Submitting our standard CLS model of that section to this analysis returns the results displayed in figures 4.18 and 4.19.

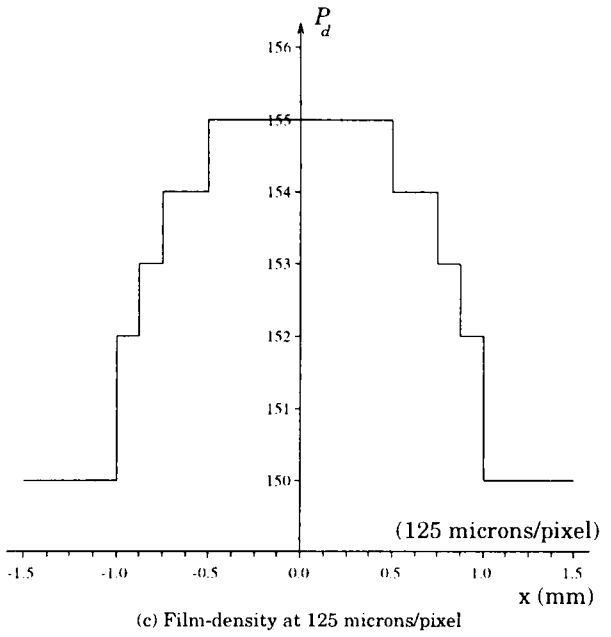
Comparison with figures 4.15 and 4.16 reveals that unlike the mono-energetic/non-scattered case, there is a difference between the film-density and transmitted-light representations, and in particular this is reflected in the second difference curves of figure 4.19. Thus following on from the similar situation identified in the analytic mono-energetic/non-scattered case of section 4.4.1, we must determine in which image representation the analysis should be conducted.



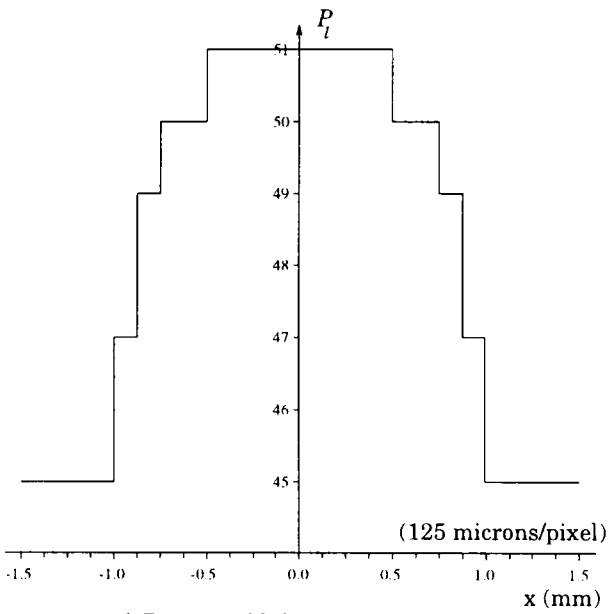
(a) Film-density at 50 microns/pixel



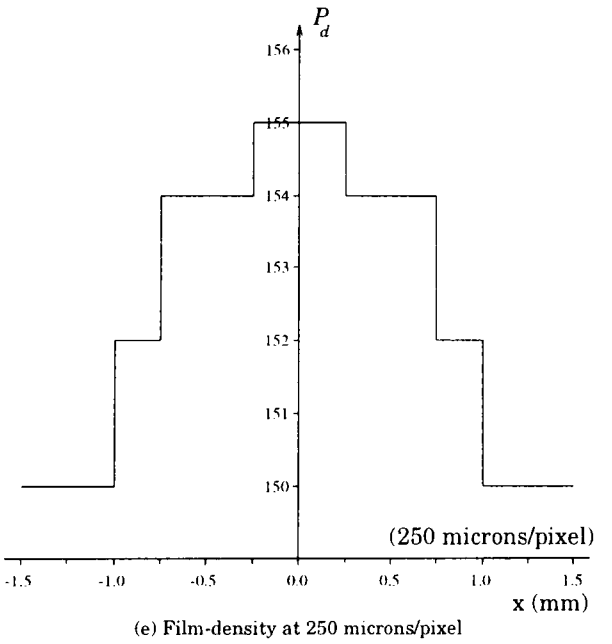
(b) Transmitted-light at 50 microns/pixel



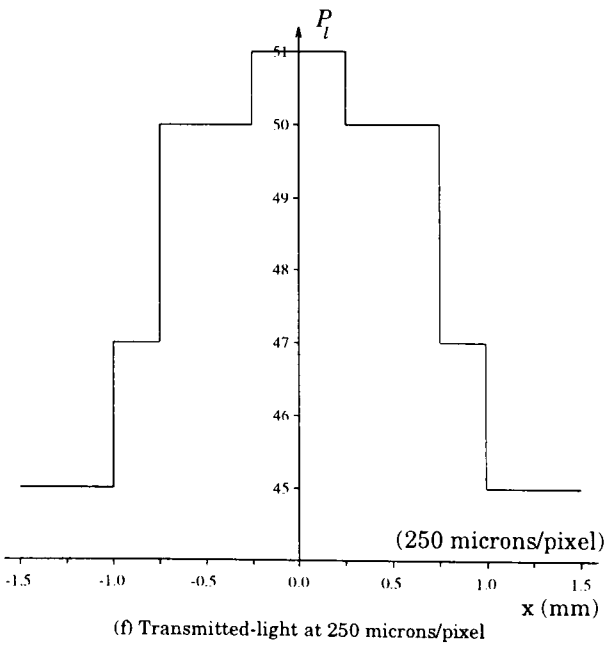
(c) Film-density at 125 microns/pixel



(d) Transmitted-light at 125 microns/pixel

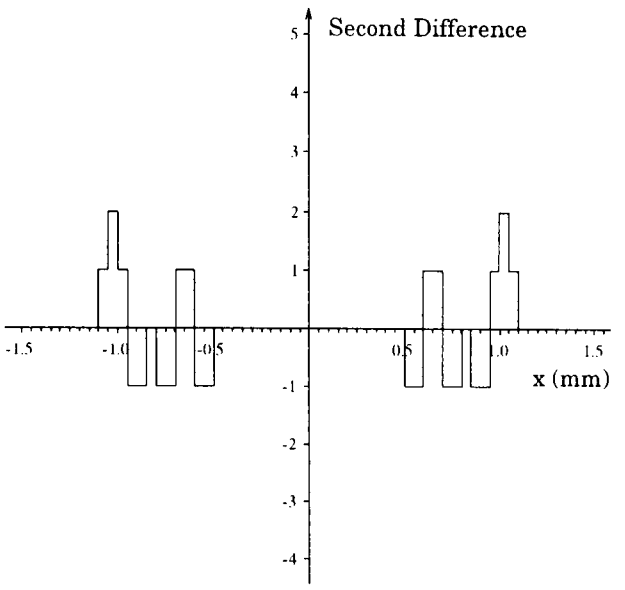


(e) Film-density at 250 microns/pixel

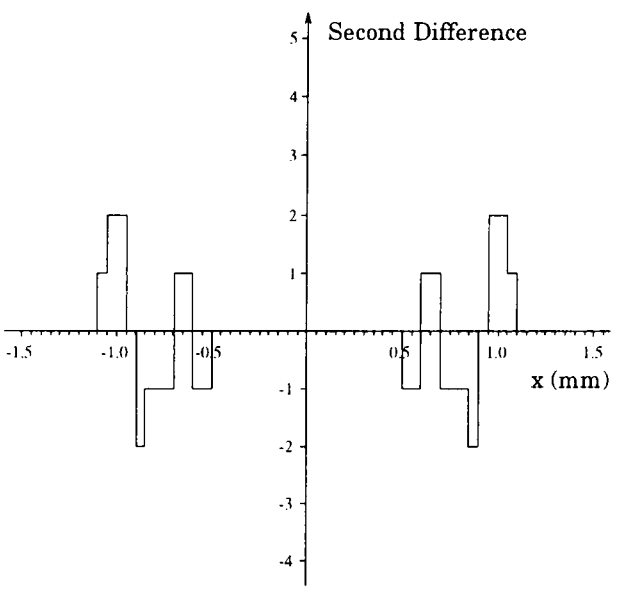


(f) Transmitted-light at 250 microns/pixel

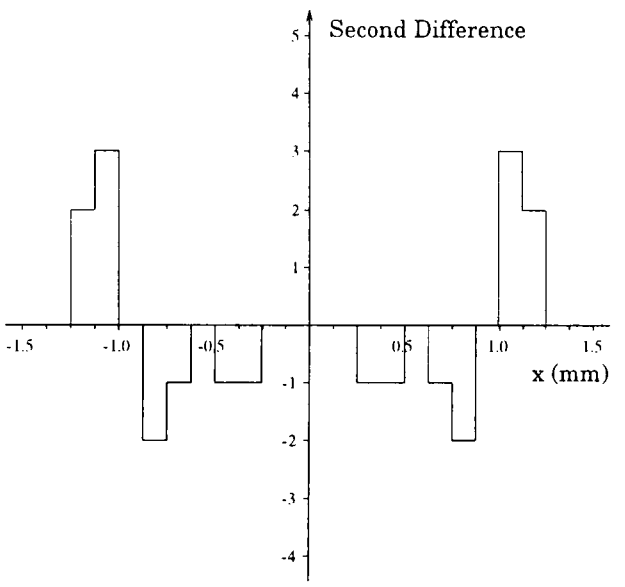
Figure 4.18: The simulated image surfaces for a poly-energetic beam and scattered radiation. Compare with the mono-energetic/scatterless simulations of figure 4.15. Note that the x-axis ticks indicate the pixel size used in the simulation.



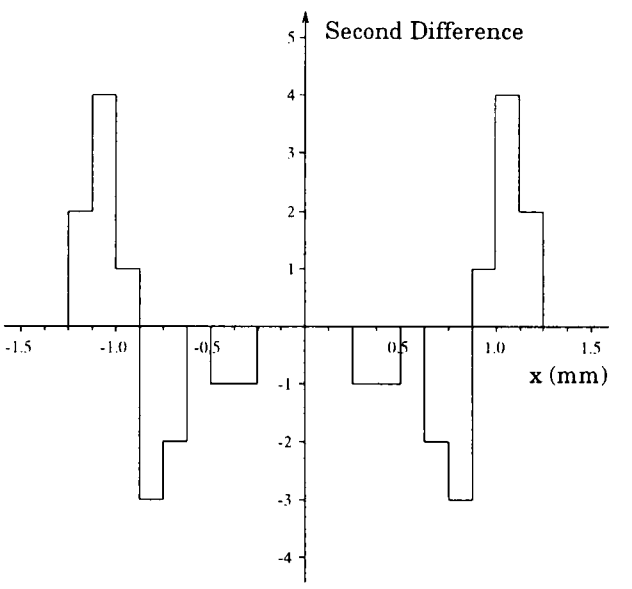
(a) Film-density at 50 microns/pixel



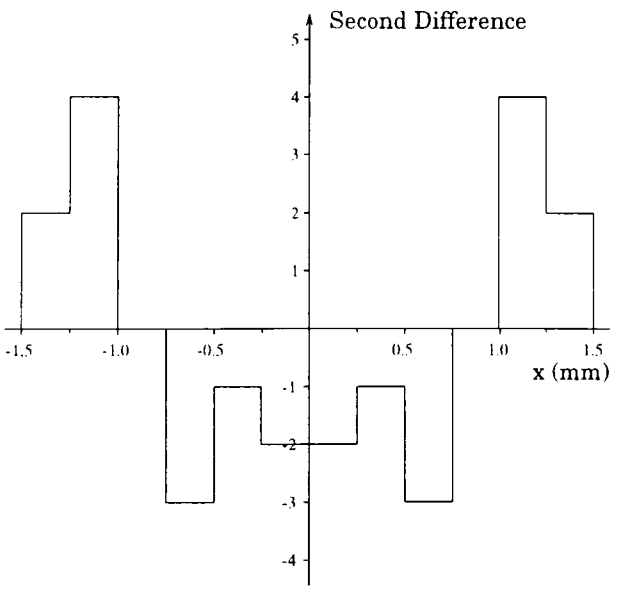
(b) Transmitted-light at 50 microns/pixel



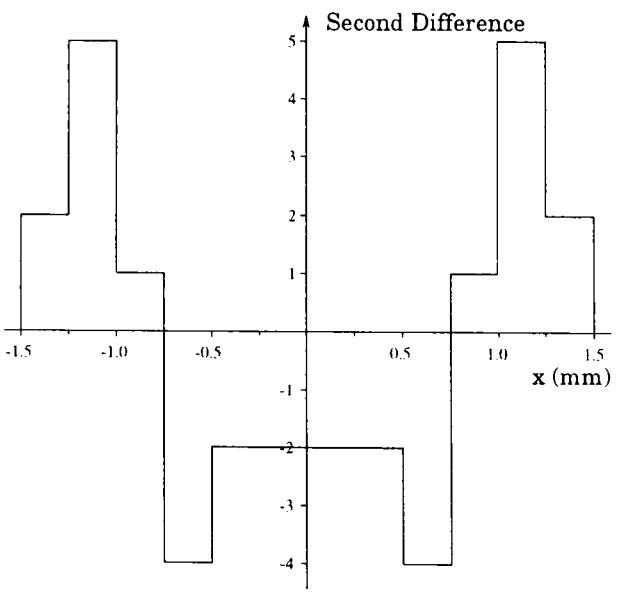
(c) Film-density at 125 microns/pixel



(d) Transmitted-light at 125 microns/pixel



(e) Film-density at 250 microns/pixel



(f) Transmitted-light at 250 microns/pixel

Figure 4.19: The second-difference responses to the poly-energetic/scattered-radiation image surfaces of figure 4.18. Note that the x-axis ticks indicate the pixel size used in the simulation.

4.4.5 Film-density or Transmitted-light Image Representations?

Although the appearance of equations (4.10) and (4.12) are quite different, in practice they produce very similar curves of the form shown in figures 4.13 and 4.14, always rising to a (negative) maximum at $x = 0$ (note the derivatives of P_d'' and P_l'', that is P_d''' and P_l'''). Described alternatively, one could say the respective curves achieve their *minimum* magnitude, or displacement from the x–axis, at $x = 0$. Since it is the magnitude of P_d'' and P_l'' that is important when searching for CLS features, it is instructive to consider the logarithm of the ratio of $P_d''(0)$ and $P_l''(0)$ as a measure of the relative performance index for the two analysis domains where a positive value of the index reveals a stronger signal in the film-density domain and hence that this domain should be preferred, whilst a negative value indicates that the transmitted-light domain should be preferred. From equations (4.12) and (4.10) we can write:

$$\frac{P_d''(0)}{P_l''(0)} = -m(\beta\phi X_c \mathcal{E}_p)^{\gamma}(\ln 10\alpha I_l)^{-1} \exp\left(-\gamma\mu_1 H - 2b\gamma(\mu_2 - \mu_1)\right). \tag{4.13}$$

As described in section 4.4.2 the combined imaging parameters $(\beta\phi X_c \mathcal{E}_p)$ are really a function of the breast substrate parameters H and μ_1. Setting these parameters as before and substituting into equation (4.13) reduces it to:

$$\frac{P_d''(0)}{P_l''(0)} = \kappa e^{-2b\gamma(\mu_2 - \mu_1)} \tag{4.14}$$

where $\kappa = 0.97886$ incorporates all the imaging variables. Figure 4.20 plots the characteristics of this equation, displaying the logarithm (to base 10) of $P_d''(0)/P_l''(0)$.

Within the limitations of the current modelling assumptions, inspection of equation (4.14) and figure 4.20 reveals that for all clinical imaging conditions and CLS parameters the transmitted-light domain outperforms the film-density, although the performance of either method is comparable.

Although developing an analytic model incorporating the effects of a poly-energetic beam is an intractable problem, it is possible to glean from equation (4.14) what effects a poly-energetic beam might have. Consider the nature of the poly-energetic beam typical of mammography shown in figure 4.9. Note that of the (poly-energetic) beam energies other that 17.4keV (the energy assumed in the mono-energetic analysis), the majority occur at energies *lower* than 17.4keV, and thus the net effect of using a poly-energetic beam will be dominated by the effect of the lower energy photons. Furthermore consider the variation of linear attenuation coefficients for fat and CLS-like tissues shown in figure 4.21. As the energy is reduced below 17.4keV, the difference between the substrate and CLS tissue attenuation coefficients $(\mu_2 - \mu_1)$ increases. Substitution into equation (4.14) shows that as $(\mu_2 - \mu_1)$ increases the value of $\log(P_d''(0)/P_l''(0))$ will decrease. Since all typical clinical operating conditions give $\log(P_d''(0)/P_l''(0)) < 0$, the effect of moving to the poly-energetic beam will swing the balance *further* in favour of the transmitted-light image representation, a result observed during the simulation of the previous section.

In regard to the selection of the appropriate image representation for detecting CLS features, observe that although the balance appears, on the basis of the evidence presented thus far, to

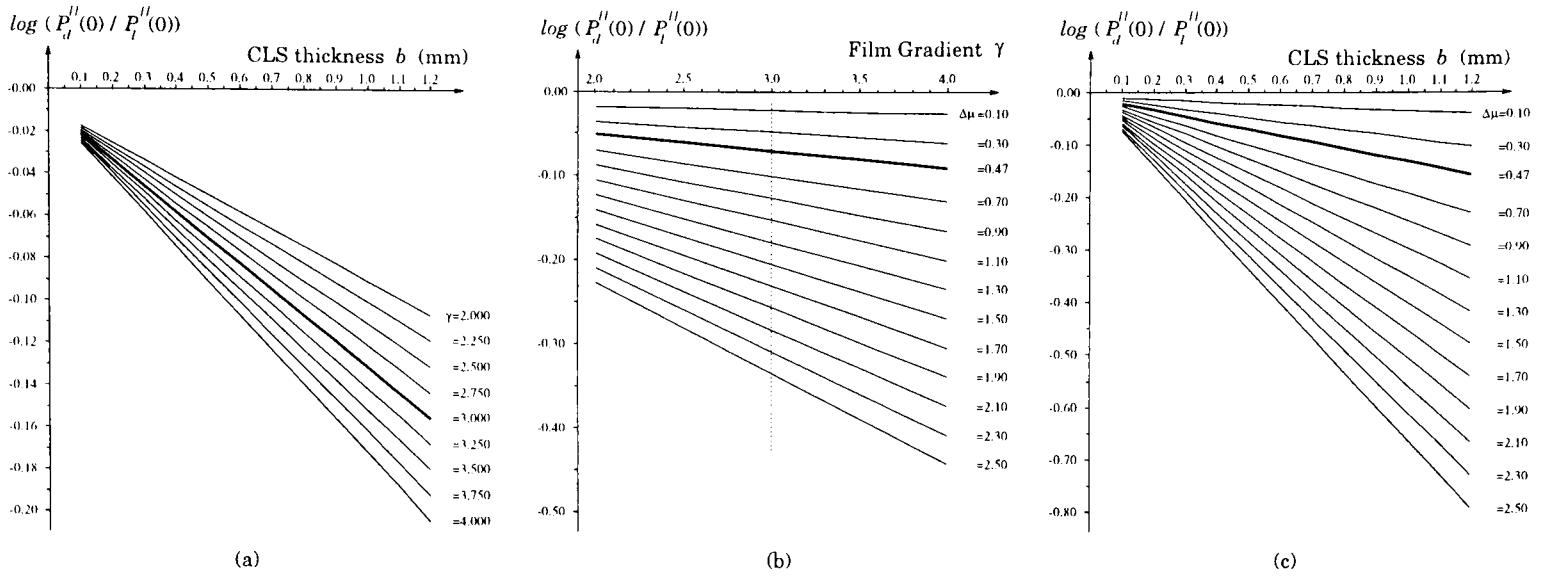


Figure 4.20: These curves display the characteristics of equation (4.14), which rates the relative performance of the film-density against the transmitted-light image representations for the task of extracting the second derivative of the image surface. Positive values of the range $\log(P_d''(0)/P_l''(0))$ indicate that the second derivative of the film-density image representation has larger magnitude than that of the transmitted-light representation, and therefore the film-density image representation is the preferred domain. Conversely negative values indicate that the transmitted-light representation is the preferred domain. The constants for each graph are (a) $\mu_2 - \mu_1 = 0.47/cm$, (i.e., assuming $\mu_2 = 1.028/cm$ and $\mu_1 = 0.558/cm$ (b) $b = 0.5mm$ and (c) $\gamma = 3.0$. Where the cross-reference is appropriate, these standard parameter values are identified in each graph by the **bold** typeface of the curve and the dashed line in (b). Note that the curves *always* return negative values of $\log(P_d''(0)/P_l''(0))$, and on this basis it would appear that the transmitted-light domain would be the domain of choice. Note also however, that the magnitude of the response is very small, that is, although the transmitted-light domain might be favoured, the difference between the two representations is very minor.

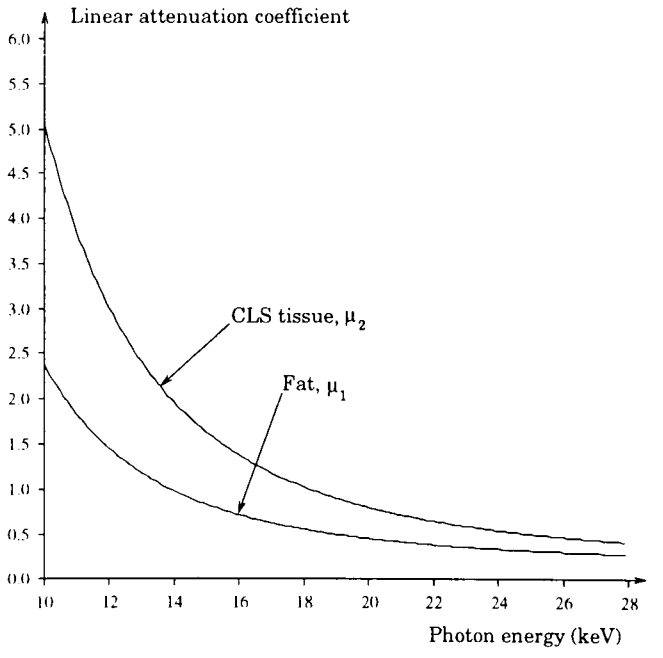


Figure 4.21: Values of the linear attenuation coefficients for fat and CLS-like tissues. From Highnam [97, p59].

lie in favour of the transmitted-light representation, it is only marginally so. Further to this, observe that P_d'' (equation (4.10)) is a function of the digitisation calibration (m), the film-gradient (γ), the material properties (μ_1, μ_2) and finally the CLS geometry (a, b). By

comparison, observe that P_l'' (equation (4.12)), in addition to the variables above, depends upon many additional terms, most notably the imaging constants (β, ϕ, X_c, $\mathcal{E}_p$). Thus the film-density representation has the advantage of being decoupled from many of the imaging conditions under which the image was formed, a luxury not enjoyed by the transmitted-light representation, as evidenced by the need to set these parameters (in combination) explicitly during the analysis presented previously. It is commonly known that the imaging parameters of β, ϕ, X_c and $\mathcal{E}_p$ can vary with time, and this instability can lead to non-linearities and inconsistencies in the evaluation of equation (4.12). Therefore, despite the relatively minor evidence available in support of the transmitted-light representation, there are grave concerns regarding its stability due to the coupling with the imaging parameters. Consequently it is recommended that the film-density image representation be used when analysing image second-differences, for example when searching for CLS pixels as described in these sections.

To conclude this discussion concerning the preferred image representations, consider a preview of the results available from the methods for finding the CLS pixels in digital images developed in the following sections. A small portion of a real mammogram image, furnished in both film-density and transmitted-light representations, is searched in each domain for the CLS pixels, and the results are displayed in figure 4.22. This figure also shows the difference image between the two 'CLS pixel' images. Firstly, note the similarity between the CLS pixels extracted from the two image representations. Also note that the difference between the two representations certainly does not contain any structural information and therefore there is little between the two representations.

4.4.6 Spatial extent of the CLS features

In order to determine the spatial extent of the CLS features in an image so that we might select the second difference kernel size accordingly, consider the noiseless, mono-energetic and scatter-less image intensity surface across a CLS feature described by equation (4.10) (we only consider the case of the film-density representation in light of the conclusions of the previous section).

From this equation we can write the image intensity representing the background fatty tissue as:

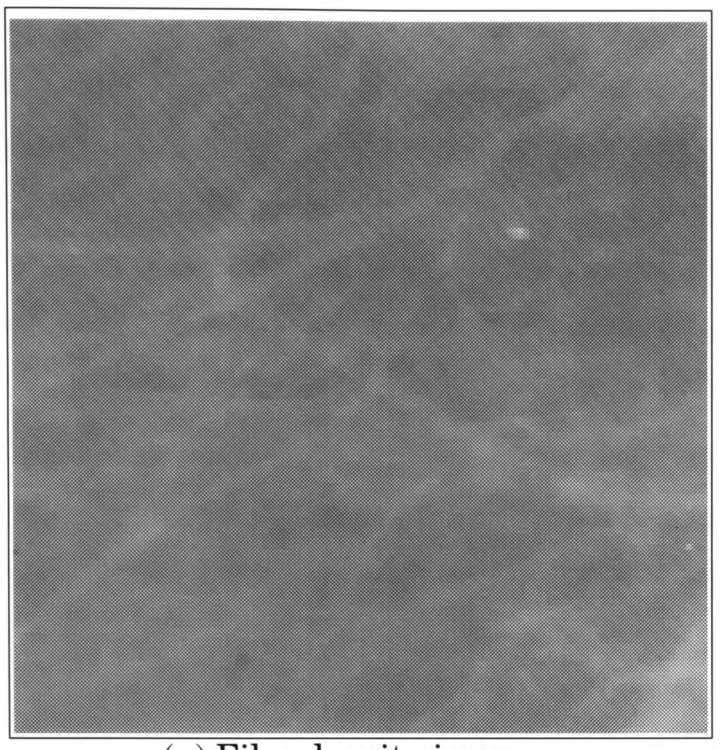
$$P_{\text{bg}} = \frac{m\gamma}{\ln 10}\left(\ln(\beta\phi X_c\mathcal{E}_p) - \mu_1 H\right) + q,$$

and the maximum image intensity of a CLS feature (i.e., at $x = 0$) as:

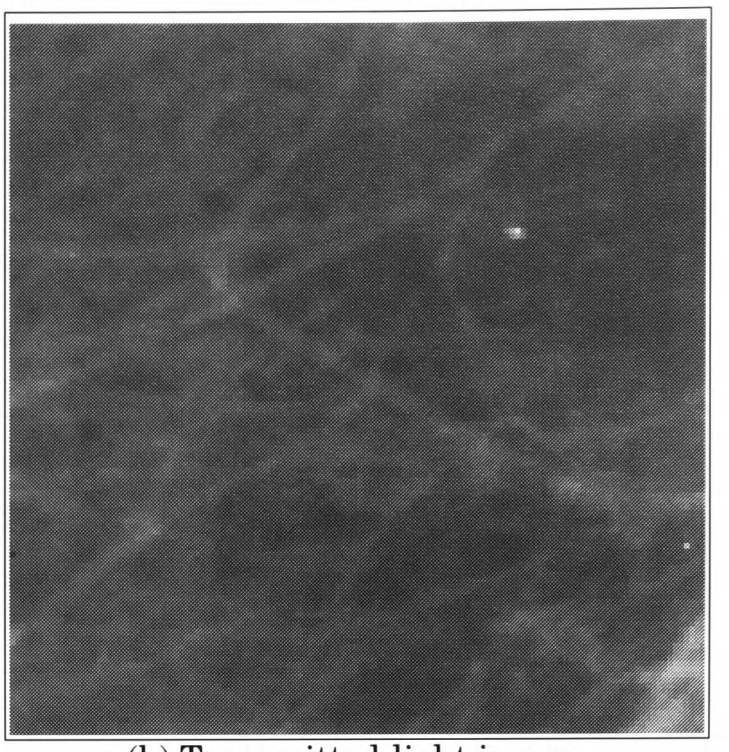
$$P_{\text{fg}} = \frac{m\gamma}{\ln 10}\left(\ln(\beta\phi X_c\mathcal{E}_p) - \mu_1 H - 2b(\mu_2 - \mu_1)\right) + q.$$

Let the image surface difference between the background and the CLS feature be $\Delta P = P_{\text{fg}} - P_{\text{bg}}$, then rearranging these equations gives:

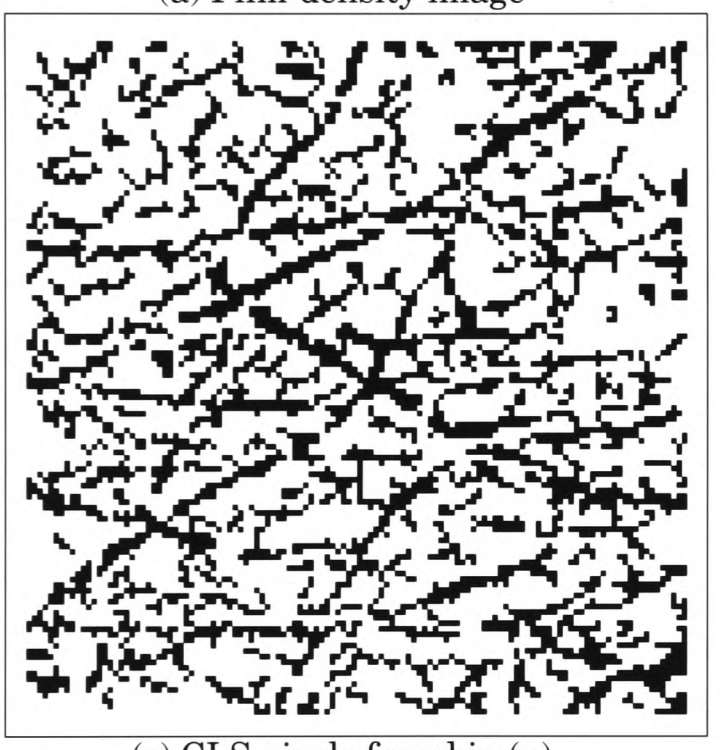
$$b = -\frac{\ln 10 \Delta P}{2m\gamma(\mu_2 - \mu_1)}. \tag{4.15}$$



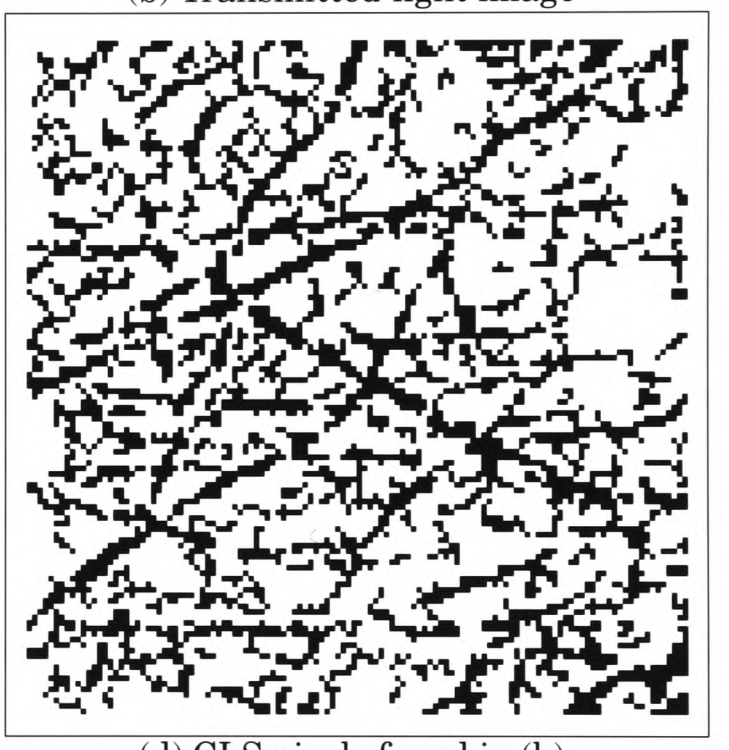
(a) Film-density image



(b) Transmitted-light image



(c) CLS pixels found in (a)



(d) CLS pixels found in (b)

(e) Difference between (c) and (d)

Figure 4.22: A comparative analysis of the pixels found in images with intensities representing film-density and transmitted-light.

From this equation it is a simple matter of allowing $\Delta P = 1$ and substituting the other constants from figures 4.12 and 4.11 to determine the minor radii (or half the thickness) of a CLS feature that will sufficiently attenuate the X-ray beam to produce a change in the image surface of a single intensity level. Execution of this procedure reveals that $b = 90\mu m$, or that the minimum possible CLS thickness capable of producing a detectable change in the image surface (assuming the absence of noise) is $180\mu m$.

Before proceeding any further we must again consider the limitations of the assumptions under which this figure of $180\mu m$ was calculated (mono-energetic beam and no scattered radiation). Observe that the poly-energetic and scattered radiation simulations of figure 4.18 have a lower ΔP magnitude than the mono-energetic and scatter-less simulations of figure 4.15. Thus for a given resultant ΔP in an image, a CLS feature needs to be thicker in the poly-energetic/scattered case compared to the mono-energetic/scatter-less case. Accordingly, this means that the value of $180\mu m$ can be considered as a lower limit for the thickness of CLS features which simply cannot be successfully imaged.

To develop a measure of the expected width of CLS features presenting in an image, consider the view introduced in section 4.1.2 that due to breast compression at imaging, it is unlikely for a CLS to be thicker than it is wide relative to the X-ray beam path. Adopting a conservative approach of allowing the CLS features to be of equal extent in either direction (i.e., circular cross-sections) surely accounts for nearly, if not all the possible CLS features. In this situation then, one would only expect CLS features of greater than $180\mu m$ width to be present in an image, and given the unlikely nature of the circular cross-section, they would be expected to be substantially wider than $180\mu m$.

Note that this analysis has assumed the absence of any degrading effects due to image noise, finding an absolute lower limit of the CLS width expected in an image as it gives rise to a single grey-level increase in image intensity. Recall from section 4.1.3 a rough estimate of the noise present in mammographic images was evaluated as $N(0, 3\text{–}4)$. On this basis then, and the reasoning given above in this section, it is reasonable to assume that in order to consistently detect the pixels of a CLS feature by this analysis, the features must somewhat larger than this lower limit. This is an important result, since it allows us to minimise the computational overheads of any analysis we undertake by tuning the resolution of the image to suit the expected size of the features. The majority of images digitised with currently available technology are done so at resolutions of approximately $50\text{–}125\mu m$ square/pixel, substantially smaller that the smallest feature we aim to detect. Thus we have the option of reducing the resolution of the images to say $250\text{–}300\mu m$ square/pixel. The actual resolution chosen depends on a number of factors, including the expected spatial extent of the features as mentioned above, the image noise, and the desire to sample the signal at a sufficiently high frequency to capture its details. Given the variation in signal amplitude (CLS image surface intensity) accompanying the various CLS features one must sample the signal more frequently than the Nyquist limit of Shannon's sampling theorem. For a binary signal of wavelength $360\mu m$ (approximately the minimum

detectable wavelength as it is twice the $180\mu m$ minimum CLS width expected), coupled with the influence of the estimated noise (variance $\approx$ 3–4 intensity levels), and the arguments of the previous paragraph equating intensity with spatial extent, this puts the Nyquist limit at $720\mu m$. Of course these comments apply only to a binary signal, something that the image intensity surface across CLS features can definitely not be described as. Consequently we must sample the image at higher frequencies than this limit, and we therefore choose to work with images at $300\mu m$/pixel. This figure also has the advantage of giving images that are of a realistic size when displayed on standard display hardware[8]. The real significance of reducing these images, is that with an appropriate selection of the filter chosen to reduce the images, we can not only acquire images of a suitable resolution for detecting the CLS feature, but we can also reduce the image noise present in the original. The median-filter has found wide application in situations such as these as it can reduce image resolutions and image noise whilst preserving image structure. Without giving a full description of this filter operator as such descriptions can be found elsewhere [86], for completeness it is noted that the reduction of image noise is simply due to the cancellation of the effects of the randomly (assumed Gaussian) distributed noise over the extent of the filter kernel, whilst the preservation of structure is due to the selection of the image intensity at each location most likely to represent the underlying data. Simpson and Boywer [195] have compared the median filter against ten other filters (both established and peripheral) for the task of accommodating noise in mammogram images. Tested over many types of artificially introduced noise, the median filter, although not the most successful in all cases, was found to perform remarkably well overall, even resulting in the authors concluding that it is "important to notice how well the plain median filter did". On this basis then, we choose to median reduce our images to a resolution of $300\mu m$ square/pixel and conduct a search for CLS features of one pixel or more in width.

4.4.7 CLS Modelling Conclusions

By assuming compressed CLS features have approximately elliptical cross-sections, the previous sections have developed a model of the imaged CLS features that reveals the following

- Under further assumptions of a mono-energetic incident beam and no scattered radiation, the second derivative of the (analogue) image surface is strongly negative at locations representing CLS features.
- By numerically simulating the effects of poly-energetic beams and scattered radiation, a correspondingly strong negative response is apparent in the second difference of the (digital) image surface at locations representing CLS features.
- The response of CLS features in an image surface to the second difference kernel is strongly dependent upon matching the image spatial resolution, image intensity resolu-

[8]Standard monitors available on common workstations today have screen pixel resolutions of $339\mu m$/pixel (75dpi), whilst those on personal computers are typically at $280\mu m$/pixel (90dpi). Thus if an image was reduced to a resolution of $339\mu m$/pixel and displayed on a 75dpi monitor, then the image and the film would be the same physical size.

tion, kernel size, and the range of CLS feature sizes.

- Image intensity resolution is limited by the image noise which is mostly due to the imaging process itself rather than the available A/D hardware. Subject to improvements in imaging and film-screen technologies, the standard 8-bit intensity resolution commonly available today provides a suitable framework for identifying the CLS features.

- Under the conservative assumption that one expects a CLS feature to be no thicker than it is wide whilst under compression (relative to the X-ray path), only CLS features wider than some 200–300μm can attenuate the incident X-ray beam sufficiently to produce any change in the image intensity surface for an 8-bit intensity quantisation.

- Reducing the resolution of the images to 300μm with the median-filter has the advantage of reducing both the image noise and the computational overheads of any analysis.

- Due to different digitisation techniques available it is possible to obtain images whose intensities represent either the film-density or the luminance of the light transmitted through the film. For a given CLS feature, and in a perfect world, the magnitude of the second difference of the transmitted-light representation is larger than that of the film-density, although the difference between the representations is quite small for normal clinical imaging conditions. Since the second difference magnitude is larger, it should be easier, and more reliable, to detect CLS features in transmitted-light images. However the transmitted-light image intensities are strongly coupled to the imaging parameters whilst those of the film-density intensities are not. Considering the magnitude disadvantage of the film-density representation is quite minor, the decoupling from the imaging constants swings the equation firmly in favour of the film-density image representation as the domain of choice for searching for CLS pixels.

Consequently it is proposed that the search for CLS pixels be undertaken by looking for pixels with negative image surface curvature. The analysis should be conducted in the film-density image representation, looking for features as small as a single pixel in width in images reduced to a resolution of 300μm square/pixel.

4.4.8 Finding the image surface curvature

As previewed during the surface curvature modelling of section 4.4.2, within the digital domain we implement the standard approximation to the continuous second-derivative given by the convolution of the image with the second-difference kernels of the form $\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$. This kernel can be developed directly from first-principles by considering the definition of the first derivative of a continuous function $y = f(x)$ as:

$$y' = \frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Constrained to the discrete domain of the image, $\lim_{h \to 0} h = 1$ pixel. Substitution gives $y' = f(x+1) - f(x)$, the common 'rise-over-run' formulation for the gradient of a line. Rewritten as $y' = -f(x) + f(x+1)$ it is clear that this difference approximation to the continuous first derivative can be represented by the convolution kernel $\begin{bmatrix} -1 & 1 \end{bmatrix}$. Iterative application of this procedure, that is take the derivative of the first derivative, leads directly to the second-difference kernel given above.

There are two issues which must be considered before implementing this second-difference measurement of the surface curvature, the kernel scale and angle. The image surface is a function of two variables $z = f(x, y)$, thus the directional derivatives of the surface z must be measured. This is particularly important since we certainly do not expect to find the CLS all aligned, with say the x-axis, but rather distributed at all angles of rotation. Consequently a rotational family of second-difference kernels based on $\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}$ are convolved with the original image to return the second difference response in a number of directions. Obviously the assembly of this decoupled information is the key to extracting a measure of the second-difference at any location, and this will be considered in the next section. In the first instance however, the question of the kernel scale warrants further attention.

Recall from section 4.4.6 that the task is to identify CLS features as small as a single pixel. It is desirable that the CLS detector should be able to locate all CLS features irrespective of their width. Inspection of many mammogram images reveals that although the vast majority of features are up to about 5 or 6 pixels in width (at $300\mu m$ resolution, i.e., 1.5–1.8mm), a few features, almost always large veins at the upper surface of the breast suffering a 'magnification' effect, are larger than this. An example of such a feature can be seen extending from the pectoral muscle region in the image of figure 4.30(b). In any complete analysis of mammogram images these less frequent cases would need to be accommodated. When selecting the scale, or spatial extent, of a filter kernel so that it responds to the features of interest, there is a balance that must be struck between the size of, and the spatial frequency between, those features. A large scale has the capacity to identify large objects, yet lacks the resolution to discriminate closely spaced features. Conversely a small scale can identify narrow, closely spaced features, yet shows no response to features broader than its scale. Fleck [67] develops the notion of filter-kernel scale-space within the context of determining reliable estimates of finite-differences. By measuring finite-differences at a number of different scales, and then combining the information available from each scale in a systematic fashion, the advantages of both large and small scales can be utilised. It is just such an approach that is suitable for accommodating the less frequent cases of wide CLS features in mammogram images. The penalty for adopting this approach is increased implementation complexity. In a minor limitation of the current implementation of the CLS detector, we choose, for the sake of simplicity, to determine the second-differences at a *single* kernel scale only. Given that this compromise between simplicity and completeness/robustness will exclude some CLS features from identification, it is desirable that, within the available options, we seek to exclude only those CLS features that

are less important, where 'importance' is defined in terms of the intended application of the CLS description. Were it necessary (for whatever reason) to identify the broad surface veins in an image for instance, we would therefore require a filter kernel of large spatial extent. Conversely the identification of narrow, perhaps closely spaced features would require a smaller filter scale. Recall that a major motivation for identifying the CLS features in the first place is to identify and possibly accommodate the high-frequency textural clutter associated with mammogram images. It is the smaller features that give rise to the high-frequency clutter and it is therefore imperative within this context that the smallest features are detected. We have shown in section 4.4.6 that these features are as narrow as a single pixel in width, and accordingly we require a filter kernel capable of responding to features at that size. As noted above, the vast majority of CLS features in our images appear at widths up to ≈5–6 pixels. Since it is inherently more convenient to implement filter kernels of odd dimension, we therefore choose the dimension of the filter kernels to be 5 pixels. Note that such a filter is capable of detecting CLS features of widths larger than this filter scale provided the image surface maintains a negative curvature over the full width of the feature. Figure 4.18(e) and figure 4.19(e) respectively illustrate the surface profile of a CLS feature of 8 pixels in width, and the response of the second difference filter kernel of dimension 5 pixels to that surface profile. Observe that the kernel has responded over the width of the feature, particularly in the centre. The lack of response at the extremities of the CLS feature will be considered further in section 5.1.

There are three possible rotational options that could be chosen for implementing second-difference filters at a scale of 5 pixels for rectangular image tesselations (that is, 2, 4 or 8 directions). Although arriving at the selection of 5 pixel filter scales for the *Phantom* and *Spectre* detectors[9] for an entirely different reason[10], Fleck was therefore forced to make a similar choice. An evaluation of these options and another one of 6 directions for a hybrid tesselation (which apparently harnesses the advantages of a hexagonal tesselation without going to the trouble of converting the standard rectangular one) revealed that very little real net difference existed between the options of 4, 6 and 8 directions, whilst they were all superior to that of 2. In experiments conducted within the present work, the options of 4 and 8 directions were considered, and similarly very little difference was observed. For simplicity of implementation then, we choose to set the number of directions to 4. Figure 4.23 shows the second-difference directional convolution kernels as implemented in the CLS detector.

We have established our intention to use second difference filter kernels at a single scale of five pixels and in four directions. The consequences of adopting this single-scale will be indicated at the occurrence of appropriate examples throughout the CLS detector results and its applications which follow.

[9] *Spectre* is an improved, though very similar, version of *Phantom*.

[10] Edge detectors seek to accurately localise an identified edge and therefore often choose filter scales as small as possible. Images analysed in these environments are often acquired from CCD imaged and digitised scenes, either sequences (video) or still pictures. To accommodate the interlacing that is often (but not always) associated with CCD or vidicon image capture, second difference filters of the type used here are employed.

$$\begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}^T \quad \begin{bmatrix} \sqrt{2}-1 & 0 & 0 & 0 & 0 \\ 0 & 2-\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 2-\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2}-1 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 & 0 & \sqrt{2}-1 \\ 0 & 0 & 0 & 2-\sqrt{2} & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 2-\sqrt{2} & 0 & 0 & 0 \\ \sqrt{2}-1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Δ''_{N} Δ''_{NE} Δ''_{E} Δ''_{SE}

Figure 4.23: The second difference directional convolution kernels for estimating the second derivative of the image surface.

4.4.9 Merging the directed-difference information

With the second difference responses available in four directions at each pixel (vertical, horizontal and the two diagonal directions between), our task is to assess each location for a significant negative second-difference response (negative in the sense of 'less-than-zero' rather than the logical null response), since it is this measure that allows the determination of the CLS features as described previously.

Consider a CLS feature lying in say, the vertical direction of this page. In accordance with the CLS model developed above, this feature would naturally exhibit maximal image surface curvature in the horizontal direction, and near zero image surface curvature in the vertical direction. Thus regions of this feature that are close to the boundary (relative to the kernel scale) would clearly display a large negative response to the Δ''_{N} kernel, and a near zero response to the Δ''_{E} kernel of figure 4.23, as we would expect from the simulations of section 4.4.2. The values of the Δ''_{NE} and Δ''_{SE} could be either large negative, intermediate negative, near zero or positive ,depending upon the details of the neighbouring region. In these cases, this information in addition to the Δ''_{N} and Δ''_{E} kernels previously mentioned, would indicate respectively, support for the current assessment of strong negative response, and similarly for intermediate values which by contrast do *not* reject the notion of a strong negative response, as is the case for near zero. Finally, a positive response would indicate the presence of a *saddle*. In all but the case of a saddle point, the identification and labelling of a significant negative response is relatively clear. In cases involving a saddle point in the image surface it is less clear what should be done. By observation, we have noted that the location of such saddle points in mammogram images is nearly always associated with overlapping CLS structures. In this situation the increased image brightness in the region of the overlap (due to increased X-ray beam attenuation resulting from the increased thickness of CLS material traversed by the beam) causes a *positive* surface curvature for those regions of the CLS features immediately *adjacent* to the overlapping region. Coupled to the negative curvature in the direction parallel to the feature overlapping the current feature, this is therefore identified as a saddle point. Observe that in these situations the magnitude of the positive difference will necessarily be smaller than that for the negative curvature (due to the location of the central pixel and the relative spatial weighting of the kernels), and this can be used to identify these instances from saddle points resulting from simply a noise process. With this formulation then, we can now specify the algorithm for merging the directed difference information.

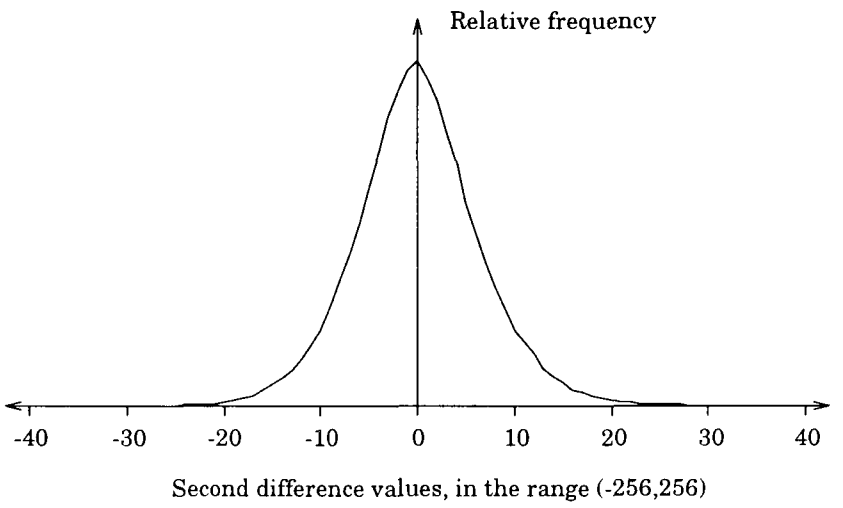


Figure 4.24: Histogram of the responses to the Δ''_{N} filter for the breast area pixels of the right image (left breast) of figure 2.7. Note that the response at zero has been suppressed due to its excessive magnitude, and only the central portion of the domain has been displayed.

Let Δ''_i be the n-second difference filters with orientations α_i, where $i = 1, \ldots, n$. Let the convolution of these filters with the image I be represented as $I''_i = \Delta''_i * I$. In a slight abuse of notation, allow I''_i to represent the second-difference value at pixel (x, y) in the i-direction, as appropriate. A pixel (x, y) therefore has a strong second-difference response if:

$$\min\left(I''_1,\, I''_2,\, \ldots,\, I''_n\right) < T \quad \text{and} \quad \max\left(I''_1,\, I''_2,\, \ldots,\, I''_n\right) \approx 0 \tag{4.16}$$

or:

$$\begin{aligned} \min\left(I''_1,\, I''_2,\, \ldots,\, I''_n\right) + \max\left(I''_1,\, I''_2,\, \ldots,\, I''_n\right) &\approx 0^- \quad \text{and} \\ \min\left(I''_1,\, I''_2,\, \ldots,\, I''_n\right) &\ll 0 \end{aligned} \tag{4.17}$$

where T is some threshold. Equation (4.17) identifies instances of the saddle points of interest, whilst equation (4.16) identifies the other simpler cases mentioned above.

The main task now is to set an appropriate value for the threshold T. Figure 4.24 shows an example of the typical histograms of the second-difference values returned by the kernels of figure 4.23. By far the vast majority of responses are at values close to zero reflecting the smoothness of the majority of the image, that is the 'background' regions surrounding the CLS features[11]. Naturally the responses at the CLS pixels lie in the tails of this distribution, and we therefore set the threshold T to be a function of the distribution width. By adopting this approach, we incorporate a self-tuning feature into the selection of CLS pixels, allowing for instance, a lower threshold in images of particularly fatty breasts showing little breast structure for instance. More importantly, if a new value of the threshold T is calculated at *each* of the kernel directions from the corresponding distribution, then the evaluation of equation (4.16) using the appropriate value of T will naturally give results more consistent

[11]The image from which this histogram has been calculated has been segmented into breast and non-breast tissues (by finding the breast border/edge) prior to calculating the histogram, which is measured over pixels that are part of the breast-tissues only. This avoids the biasing of the histogram towards the central zero response due to the 'smooth' image regions beyond breast edge.

with the data, than if T were assigned a global value. The obvious procedure here is to set T to some fraction of the standard deviation σ of the second-difference distribution. That is:

$$T = -t\sigma.$$

We have found through experimentation that a value of $t = 0.7$ gives a good compromise between sensitivity and specificity, and it is this value that we have used in the implementation developed as part of the present work. Section 4.9 reviews the effects of varying this threshold whilst analysing the overall stability of the algorithm.

With the procedure defined and the thresholds set, an image can now be assessed for the existence of pixels with significant second-difference responses, producing a bitmap segmentation of the image.

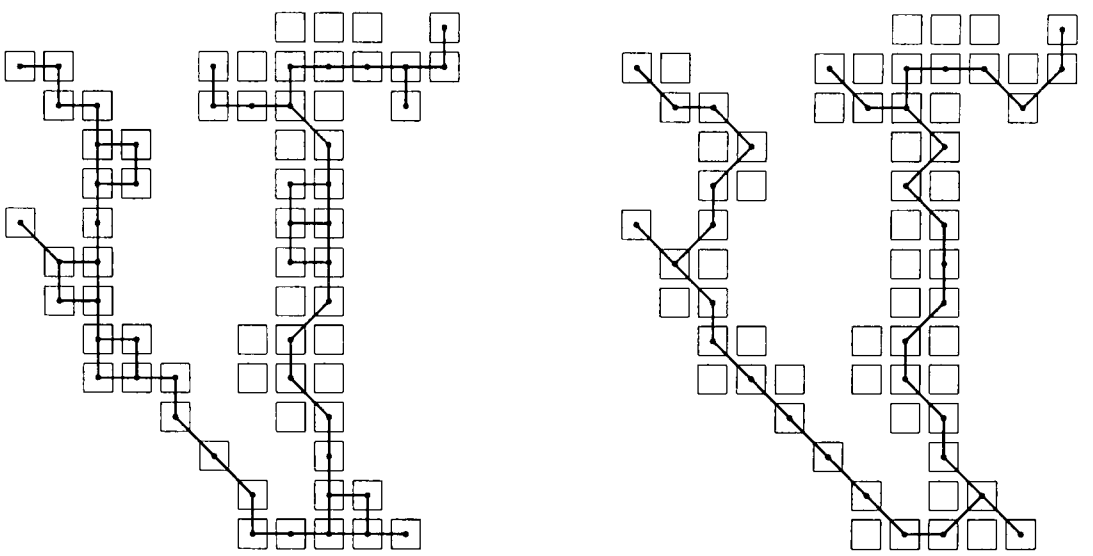
4.4.10 Forming the set N of all CLS pixels

To conclude this low-level analysis of an image in search of the CLS pixels, recall that step 1 of the CLS detector algorithm of section 4.3 sought to construct the set **N** of all pixels in an image constituting the CLS features.

Sections 4.4.1 through 4.4.7 developed a model of the CLS features which revealed that their locations in an image could be identified as those regions with negative image surface curvature. Sections 4.4.8 and 4.4.9 developed a technique for identifying the regions of negative image surface curvature, subject to the constraints imposed by the simplifications introduced to reduce the complexity of the implementation. The most notable of these limitations, due to the single kernel scale employed, is that not all CLS widths can be identified, although the vast majority can.

When forming the set **N** from the pixels identified during the last section it is possible to identify some of the pixels that were missed as a result of using the single finite kernel scale. It is this step that is introduced into the CLS algorithm as step 2 of section 4.3. Clearly the pixels that are not identified by the second-difference techniques of the previous sections will be those pixels towards the centres of the CLS features.

Let the regions of an image with significant second-difference responses as identified by the previous section be termed the *foreground*, whilst those neutral pixels not so identified be the *background*. Thus if a foreground region completely encloses a small background region it is possible that the pixels of that background region are actually components of real CLS features. Care must be exercised here, as clearly a number of CLS features may form a closed loop that completely encloses a real background region. Thus the cut-off threshold on the size of background regions eligible for reinstatement via this process must be very small. In forming the set **N** of all CLS pixels we implement the reinstatement of isolated background regions of areas up to 3 pixels only, adding them to the pixels identified during the last section. It is upon this set of pixels identifying the CLS features that the remainder of the CLS detector algorithm is based.



(a) (b)

Figure 4.25: Example results of the thinning algorithm (explained in detail in appendix A). The square boxes represent a single foreground pixel each whilst the surrounding plain regions are representative of the background. The solid black line indicates the 8-connected skeleton identified within each region. (a) Results of the Haralick and Shapiro thinning algorithm, displaying the complex connectivity that is unsuitable for the current application. (b) Results of the thinning algorithm developed in the present work, displaying the simply-connected nature of the skeleton that is necessary for the current application.

4.5 Region Thinning

The task of reducing the binary pixel map found by the methods discussed above, to a simply-connected skeleton (as required by the CLS detector algorithm—step 3 of section 4.3) is quite a complex problem. Rather than clutter the present description of the search for the CLS features with the details of the thinning algorithm developed in the present work to achieve this task, they have been deferred to appendix A. For the reader who wishes to skip the details of the appendix, it suffices to note that the application of the skeletonisation algorithm of that section returns a simply-connected topologically equivalent skeleton lying at most a single pixel from the medial-axis of the region under consideration.

It is important to achieve a *simply*-connected, topology-preserving skeleton since the subsequent steps of the algorithm for detecting the CLS features assume that each branch of the CLS network is represented by a *single* 8-connected span of pixels extending from one junction to the next. Figure 4.25 illustrates the difficulties of a general thinning algorithm, and the results of the algorithm developed in the present work to circumvent these difficulties.

As a more complete example, consider the analysis of a real region from a mammogram image. Figure 4.26 illustrates both the CLS pixels identified in an image (the grey squares), and the corresponding skeletons identified for each region (the thin black lines). Also illustrated are the identified *terminations* and *junctions* of the skeleton, marked as solid black blobs. Note the generally simply connected nature of the skeleton, except for two instances (can you spot them?[12])

[12]Centre the origin of a normally oriented Cartesian frame at the bottom-left corner of the frame surrounding the

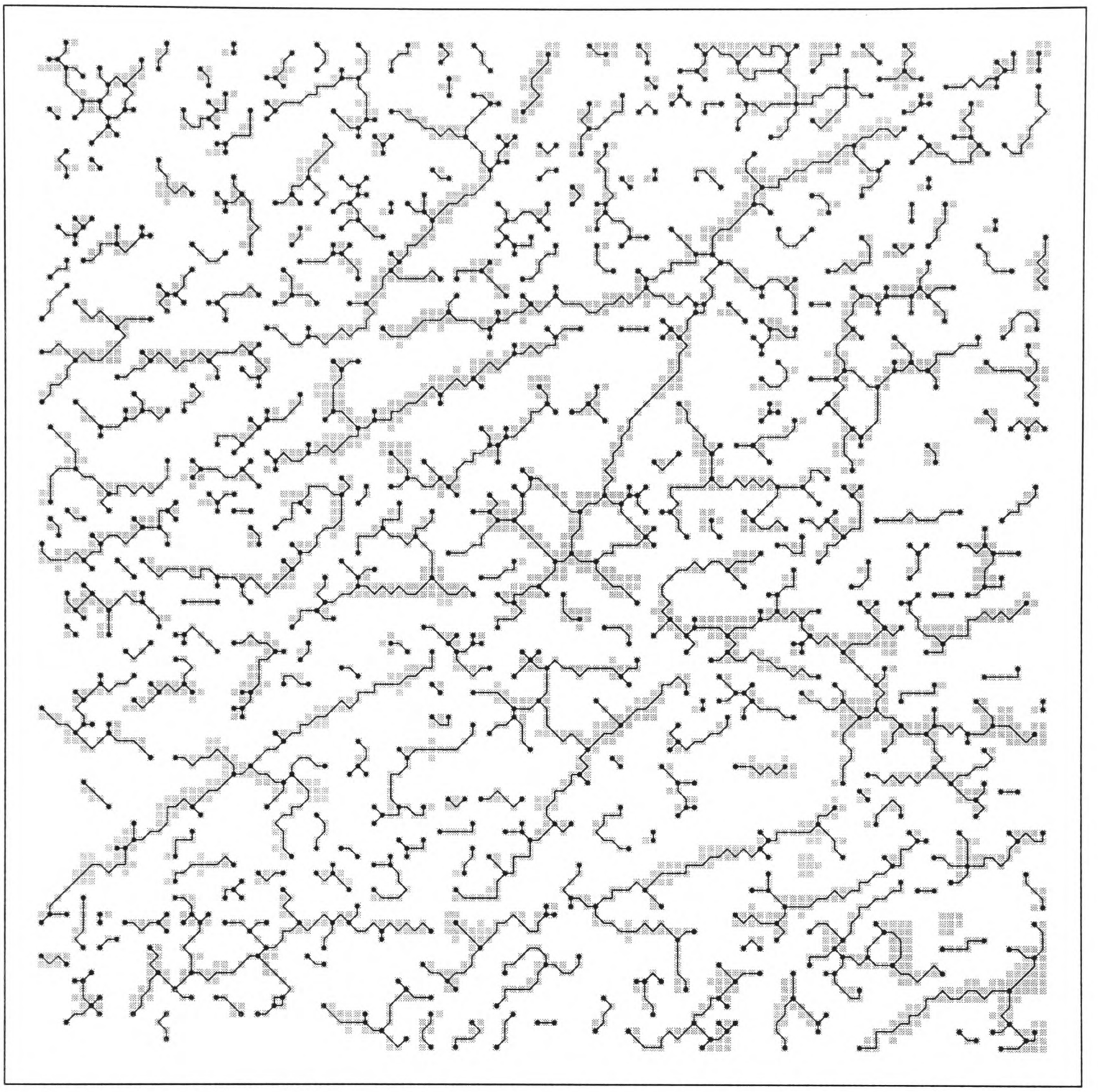


Figure 4.26: Results of the thinning algorithm developed in appendix A. The light grey squares represent the foreground pixels of the region for which the skeleton was sought, whilst the surrounding white areas should be interpreted as the background. The thin black line indicates the skeleton returned by the algorithm, and the solid dots indicate the *junctions* and *terminations* of the skeleton. Note the preservation of region topology, the simply connected nature of the skeleton, and that the location of the skeleton is at the medial axis of the regions where possible (subject to the constraint of simple connectivity).

Depending upon the intended application of this skeleton (and thus the intended application of the CLS structures so acquired) the very infrequent anomalies that do occur will either have no significant effect, or may need to be explicitly considered in the analysis. For the present work (see chapters 5 and 6) they do not influence the analysis and they are therefore simply

figure. The special cases of complexity in the skeleton are at co-ordinates $(38, 12)$ and $(97, 107)$ millimetres. These anomalies are two of the three 'mistakes' (the third is a reflection of the second about a line with gradient $= 1$ in the current frame) that the thinning operator can possibly make. See appendix A for a fuller discussion.

treated as another branch of the skeleton.

4.6 Linking Cantons into Bones

4.6.1 Introduction

Breaking the skeleton returned by the previous section into its smallest components, allows the formation of the skeletal *cantons* (see step 4 of section 4.3). This is achieved by initiating a new canton at each unclassified branch of each skeletal junction, and linking the connected pixels of that portion of the skeleton into a canton until either the branch terminates, or intersects another branch at another junction. This is best illustrated by considering the skeleton of figure 4.26. The pixels beneath each thin black line joining two solid dots are the components of a canton, including the two end-point pixels beneath the solid dots. Therefore, a junction pixel will be a member of all the cantons that terminate at that junction[13]. As a further example observe that this procedure would break the skeleton of figure 4.25(b) into seven cantons.

Step 5 of the CLS detection algorithm sifts through this list of cantons linking those that are part of the same CLS into the *bone* of that CLS. A new bone is created for each new CLS detected.

At this point in the analysis, it is worth noting that the required degree to which the cantons are successfully linked into bones (that is, have we linked *all* the cantons of a given CLS feature into a single bone, or do we have a number of bones describing smaller segments of the CLS?) is highly depended upon the intended application for the CLS description. Consequently the algorithmic sophistication required at this stage may vary considerably. A consequence of the intended application for the CLS structures in the present work (see chapters 5 and 6) is that we are allowed a large degree of freedom in the specification of this algorithm. Section 5.1 describes a technique for 'removing' the CLS features from an image. The only requirements of that section are a description of the CLS pixels (identified previously) and knowledge of the local normal direction of each CLS. Consequently, it is of no significance if our local measurement of the CLS normal direction is based on a description of the entire CLS, or merely segments of it. The only requirement is of course, that it is necessary that the segments accurately describe the local direction.

Inspection of the cantons of figure 4.26 shows that in many regions the local direction of a canton[14] does not fit the local direction of the CLS pixels, particularly for the shorter cantons (of 2 and 3 pixels only). Therefore we cannot simply declare each canton to be a CLS bone. Furthermore, many of these smaller cantons are simply irrelevant in the broader picture of the overall CLS pixels, yet they have been identified by the thinning algorithm as a consequence of preserving the topology of the CLS pixel regions. It is these cantons (later, bones) that step 6 of the CLS detector algorithm (section 4.3) seeks to eliminate. We will return to this point in section 4.6.3.

[13]Note that a junction is naturally defined as any pixel with three or four skeletal branches attached to it.
[14]It helps if you eyeball a line of best fit to the pixels of a canton.

From a purely practical consideration, if the intended application of the CLS feature description was simply those of the subsequent chapters, then a description that merely allows an accurate calculation of the local CLS normal direction would be sufficient. That the description accurately describe the CLS features per se would be an unnecessary redundancy. That, of course, would be a foolish option to pursue, since the non-portability of the feature description would be an undesirable feature.

There is a very limited quantity of material in the literature from which to draw a solution to our current problem. The difficulty with the interpretation of the data that we seek is that it is very application specific, constrained by the insistence that we do not stray beyond the boundaries of the information actually contained in the image. Clearly a relaxation of this constraint would allow the use of techniques such as the structural saliency, deformable contours, etc., mentioned in section 4.2.

To conclude these introductory comments, we aim to link as many cantons from a given CLS feature together, yet are not too concerned if incorrect or incomplete merges are made, provided the local normal direction of the underlying CLS features can be estimated with reasonable accuracy.

4.6.2 The algorithm

With regard to the task of linking cantons together, clearly it is the definition of the criteria that determine if two cantons are of the same CLS that is important, and it is at this stage of the entire CLS detector algorithm that the prior knowledge of the expected CLS shape is incorporated. By specifying a linearity (or plane-curvature, not to be confused with the surface-curvature of section 4.4.1 modelling the CLS cross-sectional profile) criterion, we can incorporate the knowledge that the CLS shape is nominally curvi-linear.

The method employed to determine if two connected cantons are components of the same CLS feature, and if they should therefore be linked together, is a simple rule-based procedure as follows:

- Fit a low order (typically third order) polynomial to the pixels of each canton. Care must be taken to ensure that the range of the fitted polynomial is single-valued so that it can adequately describe the canton. In practice this is achieved by fitting the pixels of the canton to two polynomials, one with the x-axis as the domain, the other with the y-axis. A χ^2 analysis of the fitting residual in each case indicates which domain should be chosen.

- At each pixel in an image that is a skeleton junction, determine which cantons *terminate* at that junction. (Termination is important in the iterative technique discussed below.)

- Determine for each canton so terminating at the current junction the local gradient at the junction from the corresponding fitted polynomial.

- Determine the included angle between each combination of canton pairs at the junction. (Necessarily between 0 and 180 degrees.)

- For those canton pairs with an included angle exceeding some threshold (typically 135 degrees) record this pair as a candidate pair for merging. If a particular canton is involved in two (three or more is impossible) such pairs then choose the one with the maximum included angle.
- Once *every* junction has been assessed in this fashion, search the list of recorded merges for any cantons of length less than some threshold (typically 4 pixels) that bridges two dissimilar cantons (assessed by their projected included angles as above). Retain on the list of candidate pairs for merging only the pair with the highest included angle, returning a junction at the site of the rejected pair.
- Merge the list of candidate pairs into bones. Where a string of cantons are identified for merging such that the head of one joins the tail of another, and so on, merge all the cantons into a single bone.
- Any remaining cantons that have not been merged with another to form a bone are directly transformed into bones.

To facilitate an explanation of some of these steps consider the CLS pixels and set of cantons illustrated in figure 4.26. Figures 4.27 and 4.28 illustrate respectively the low order polynomials fitted to the cantons (only those cantons involved in a junction have been fitted with a polynomial as they are the only ones to be assessed during the algorithm), and the bones subsequently identified by this procedure.

The reason a polynomial is fitted to the canton, rather than just estimating the local heading of the canton at each junction from the closest pixels, it that the thinning algorithm of the previous section (which provides the raw material for this operation) can return a skeleton that is *locally* quite jagged, yet is *globally* relatively accurate (to within a pixel of the medial-axis). This approach allows the measurement of the canton's direction at a junction as a trade-off between the global direction of the canton, and the local influence of the immediately adjacent pixels. We have found that using quadratic polynomials (order 3) provides a good balance between ensuring the end points of the polynomial reflect the global heading of the canton, whilst allowing enough flexibility to accommodate the local deviation from this global direction that may occur at the canton end.

As this procedure returns a feature description (the bones) of the same form as the input (the cantons), that is, both data structures are simply lists of connected pixels, it is technically possible to apply this procedure iteratively. In fact, there is a trade-off within this algorithm that can actually use this iterative capability to good effect. If the included angle threshold is initially set quite high (say at 170 degrees—enforcing a near-*linear* criterion), and then with each successive application of the algorithm the threshold is reduced (allowing for successively greater curvature of the CLS features), then the sequence of merges identified will progress from those with good agreement, and therefore highest probability of belonging to the same CLS feature, through to those with much poorer agreement, and hence with lower certainty

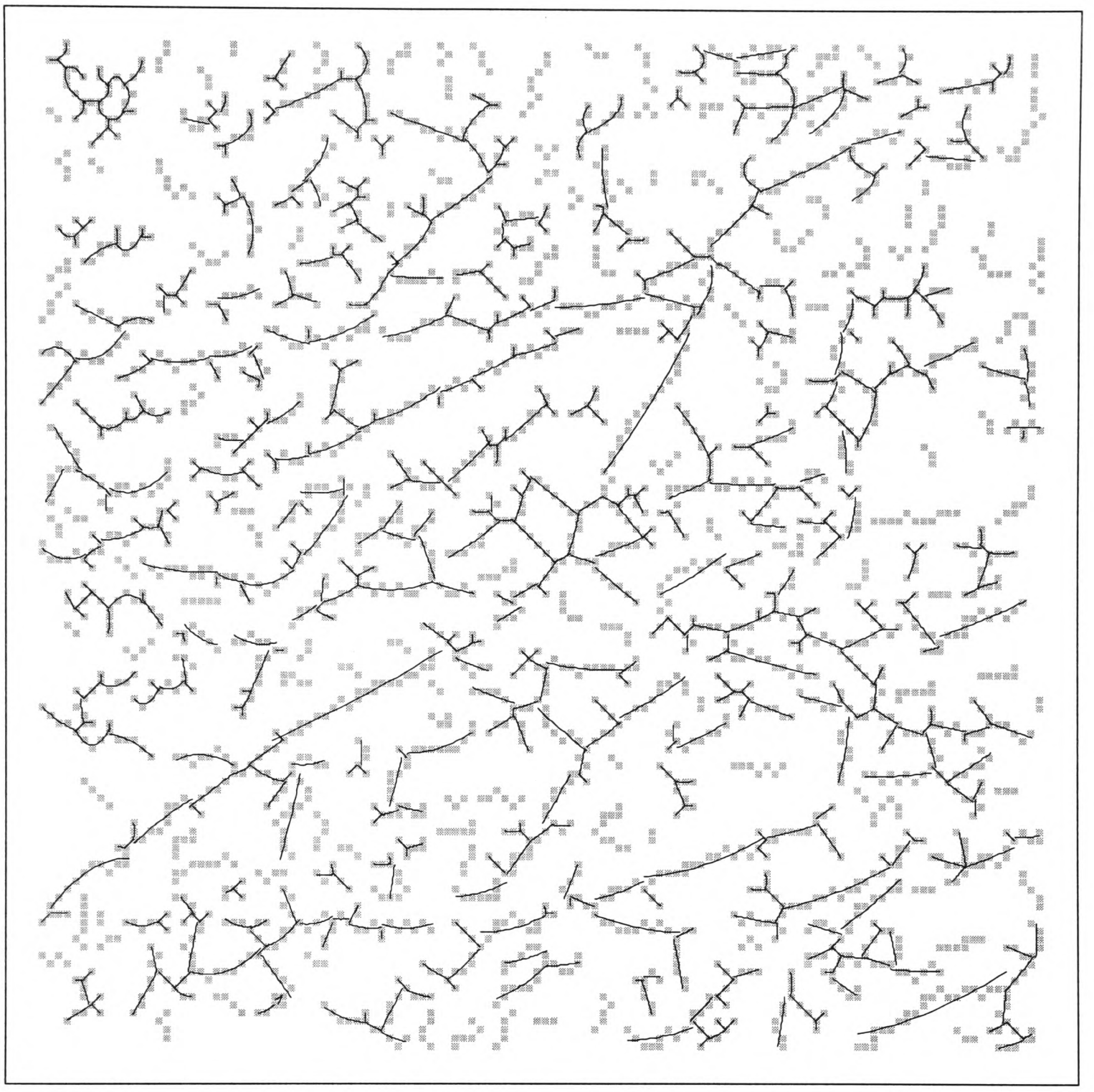


Figure 4.27: Quadratic polynomials fitted to the pixels of the cantons identified in figure 4.26. The background grey regions indicate the pixels of the skeleton that has been broken at each junction to form the cantons (compare with the solid line of figure 4.26). Observe that only those cantons involved in a junction, and are therefore eligible for consideration for merging with another canton, have been fitted with a polynomial. At canton junctions (see solid dots of figure 4.26) the included angle between all canton pairs at the junction is estimated from the local tangent to the polynomial at that point.

of belonging to the same CLS feature. Although an iterative implementation and its consequences/applications has not been explored, we have in practice found that a single pass of the algorithm at an included-angle threshold of 135 degrees gives an acceptable curvi-linear response in the resultant merged CLS features, whilst retaining algorithmic simplicity.

As mentioned above, we desire that the description of the bones be suitable for extracting an estimate of the local normal direction of any CLS feature. For completeness, if the reader

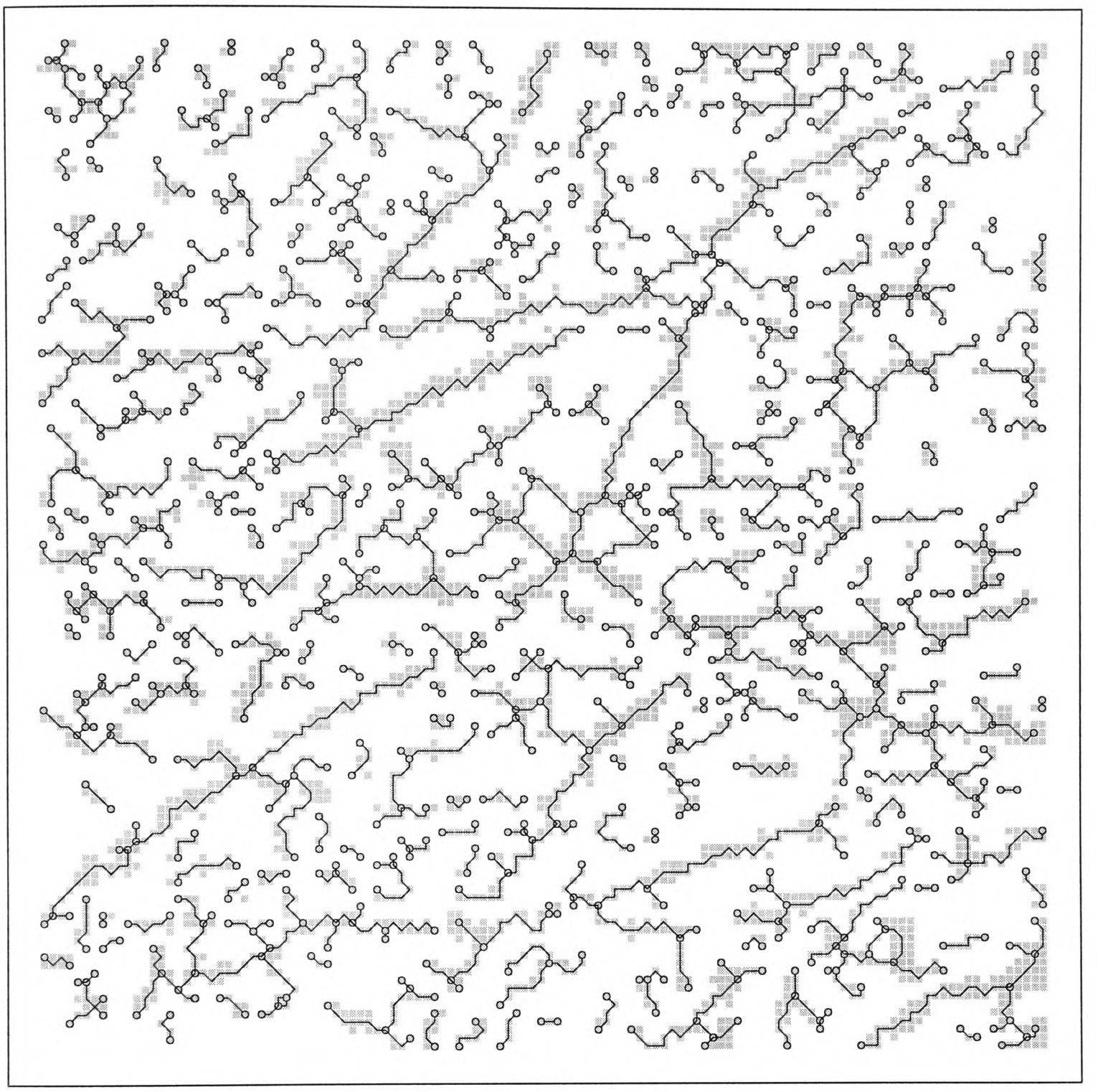


Figure 4.28: The bones identified from the cantons of figure 4.26. The grey background pixels display all the CLS pixels for which we desire a set of bones describing them. An open-circle indicates a termination at that location. A bone that passes right through a circle does not terminate at that location, but continues on until it stops at the boundary of a circle.

wishes to skip ahead to figure 5.2, an illustration of the polynomials fitted to the bones for estimating the local normal direction will be found. This figure also gives a clearer 'quick' indication of the bones displayed in figure 4.28 (although this figure is more accurate in the sense that it shows the proper connectivity of the bones, preserving the original topology of the CLS pixels).

4.6.3 Deleting known insignificant cantons/bones

Observe that a number of the smallest cantons of figure 4.26 have been completely deleted by the time the bone description of figure 4.28 has been acquired (for instance the canton at co-ordinate $(62mm, 61mm)$ in the frame of that figure introduced previously). This is a reflection of the procedure introduced as step 6 of the CLS detector algorithm of section 4.3.

It is known from the formulation of the thinning algorithm that the skeleton it returns, which gives rise to the cantons, is topologically equivalent to the original figure. For the benefit of those readers that did not skip ahead to appendix A to delve into the details of the thinning algorithm, the skeleton is acquired by successive erosion (not in the strict mathematical morphology sense, but rather in its spirit) of the original figure until no further pixels can be removed without breaking the topology of the figure. A consequence of this procedure is that if a small appendage protrudes from an otherwise nominally one-dimensional feature then this may initiate a branch of the skeleton that must remain connected to the skeleton of the locally dominant feature. In particular, if the length of the protrusion is a single pixel (such as the canton identified above) then although this is completely irrelevant as a bone in comparison to the larger bone to which it is attached, the above procedure will certainly identify it as a bone. Therefore we wish to remove these bones from the current list of bones. Consequently all bones of length 2 pixels (two adjacent pixels) that terminate in free space at one end and terminate at a junction with another bone at the other end are removed, provided that the other bone does *not* terminate at the junction. This last condition is necessary to ensure that the final little bones at the ends of some global CLS features are not deleted just because they were unable to be merged with their adjacent cantons (an example of such a situation can be seen in figures 4.26 or 4.28 at co-ordinates $(37mm, 87mm)$, where the co-ordinate system has been previously described).

4.7 Finding the CLS 'flesh'

In order to find the 'flesh' of the CLS features now that the bones describing each CLS have been found, we initiate an inflation process at each pixel of each bone, and classify each CLS pixel (detected at step 1 of section 4.3) as it is swept over by the inflating boundary defining the CLS features.

For those readers who jumped ahead at section 4.5 to consider the region thinning algorithm of appendix A, this inflation process can be considered as the inverse of the grass-fire (sugar-cane fire!) algorithm. By initiating a fire at each point along the bone, and recording the area that each bone's fire burns, one recovers the local area (the CLS) associated with each bone. There are a few points to note regarding this analogy; flame fronts simultaneously arriving at a common point share that point in each area's (CLS) description, flame fronts extinguish each other upon meeting, and flame fronts cannot 'jump' the roads between the cane-fields, that is, the pixels describing a CLS feature must be within an 8-connected region of the original set of

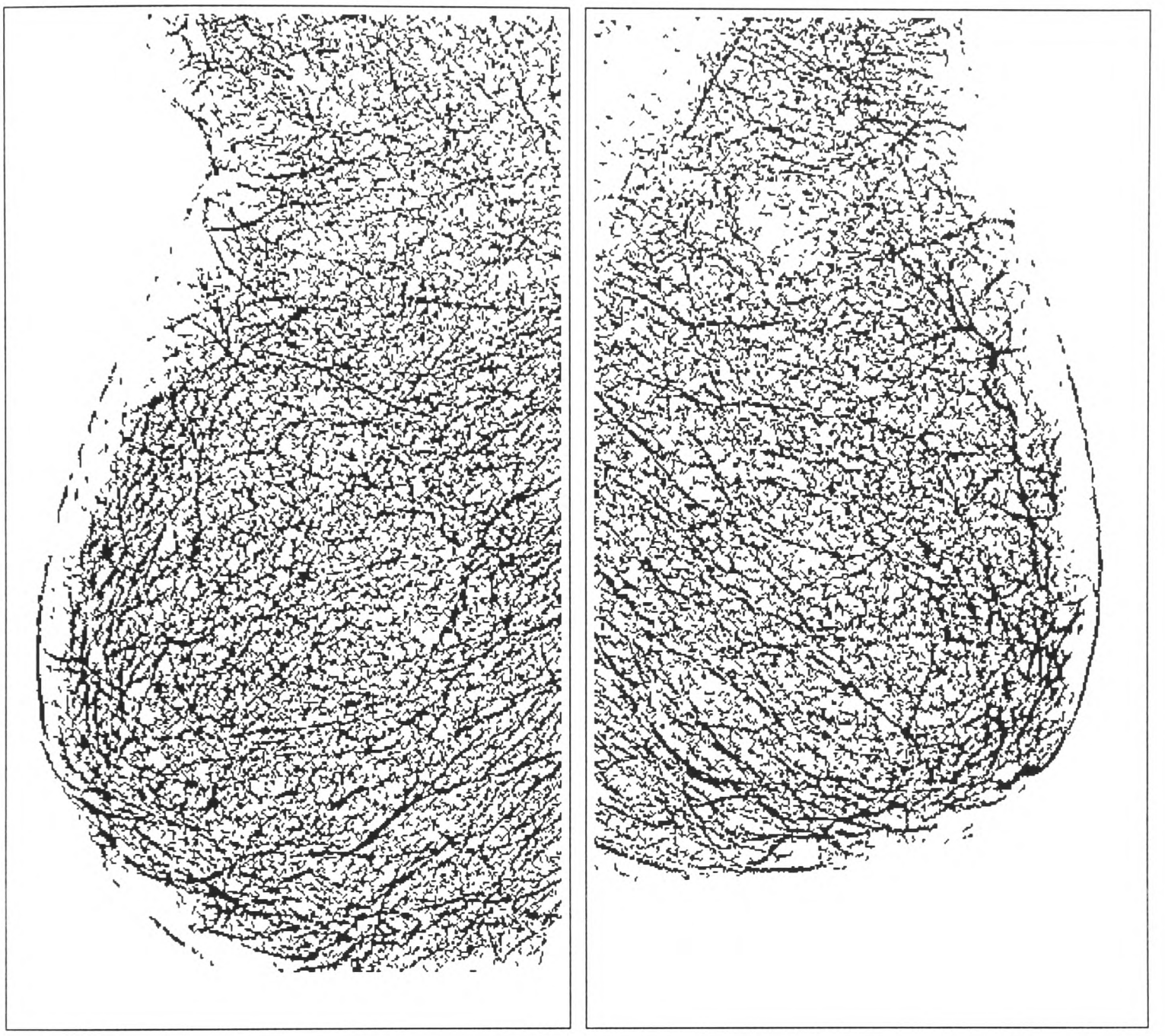


Figure 4.29: CLS pixels identified in the images of figure 2.7.

CLS pixels identified in section 4.4.10.

Following execution of this process the CLS pixels associated with each bone form the components of the CLS feature for that region of the original mammogram image. Note that this list of pixels identifies the entire area of the CLS feature, whilst the list of pixels of the corresponding bone identifies the local skeleton of the CLS.

The entire collection of such CLS features from an image, describes the CLS features of that image, and this is the result we have sought.

4.8 Results

An implementation of the CLS detector algorithm as described in this chapter has been developed, and applied to some 350 mammographic images for evaluation (322 of the images comprise the *MIAS Mammographic Database* [204]). Without exception, the algorithm appears to have identified within these images the CLS features as described by the model developed in this

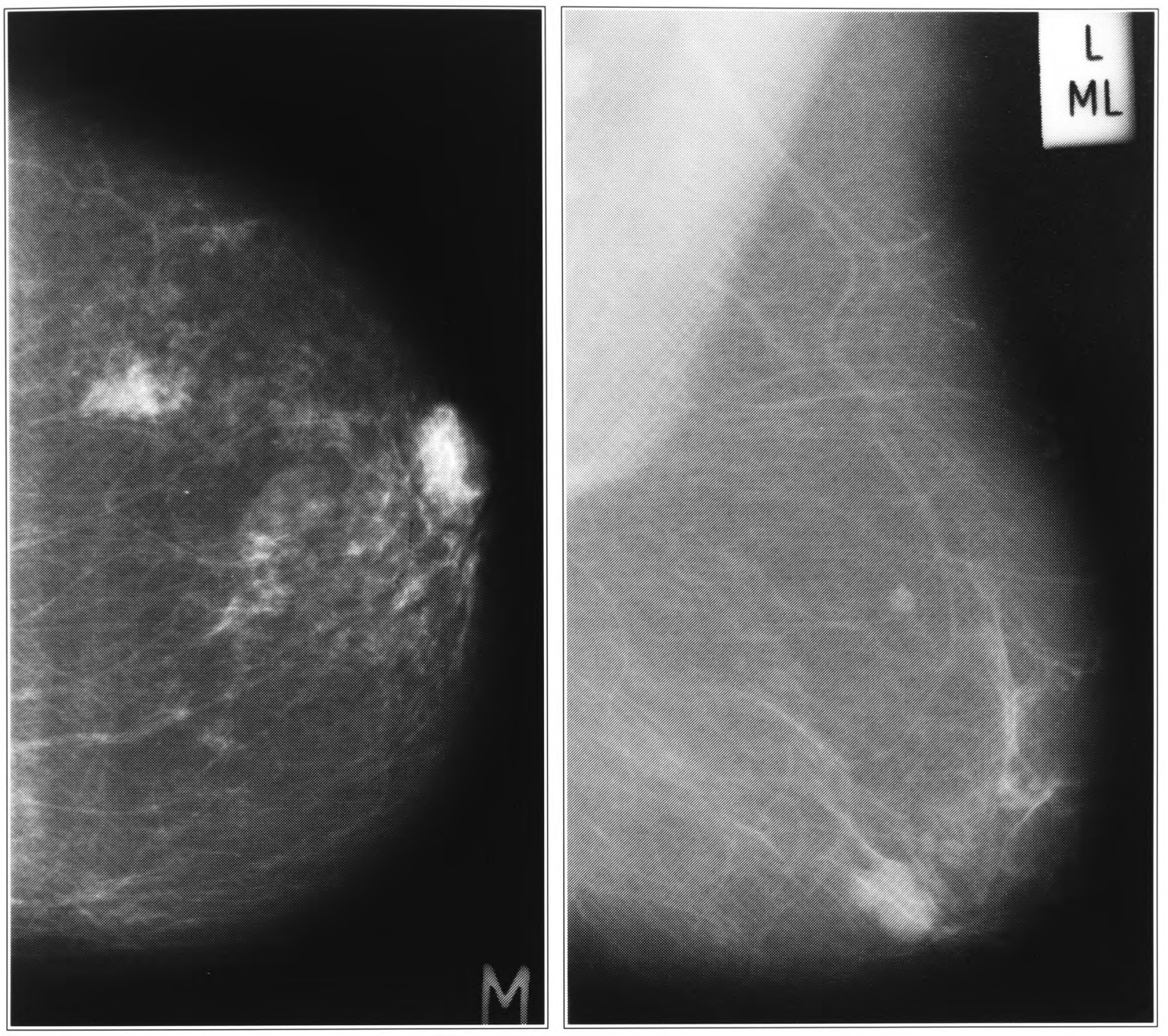


Figure 4.30: Example mammogram images.

chapter and the implementation details given above.

Figures 4.29 through 4.33 display some typical sets of CLS pixels identified by this procedure for a range of images. Note that due to the scale of these images, it is impossible to show that the bones and CLS features have been identified as discussed above (e.g., figure 4.28), however the procedure is identical, and consequently so are the results.

Observe that the CLS pixels identified in each case include the long connected curvilinear features of the image. Although many of the smaller regions of CLS pixels might appear to clutter the image and thus complicate the CLS description, they are in fact representations of the less salient CLS features present within the image and therefore their identification is warranted. This viewpoint is sustained by the further processing that this feature description supports, including the examples developed in the following chapters.

Note that as expected from the formulation of the model, the CLS detector algorithm is insensitive to the absolute grey-level in a local region, and is capable of extracting the CLS features at all local image intensities (observe the CLS features identified within regions of the

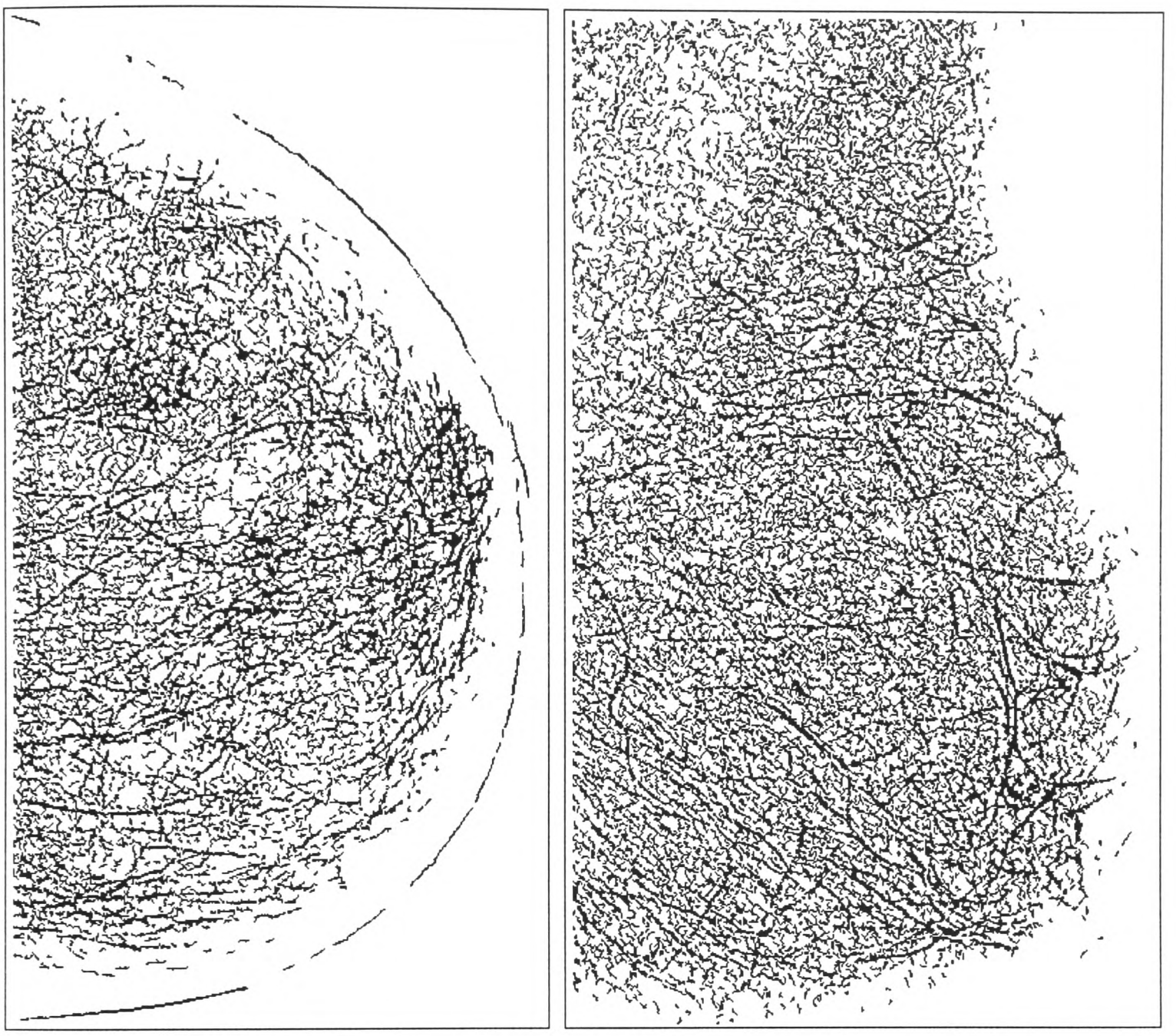


Figure 4.31: CLS pixels identified in the images of figure 4.30.

densities of the left image of figure 4.30 versus those from the fatty regions). By contrast, note that within the regions of densities and pectoral muscles of figures 4.29 and 4.33 the algorithm has correctly failed to locate any CLS features where indeed they do not exist.

Recall from section 4.4.9 that the implementation of the current CLS detector is based upon a *single* second-difference filter spatial scale of 5 pixels (rather than a multi-scale implementation). The effects of this limitation are illustrated in the right-hand image of figures 4.30 and 4.31. Note the large surface blood-vessel extending from the pectoral muscle towards the nipple, and the inadequate capture of that feature in the corresponding CLS pixel image. This feature is generally $\approx$ 8 pixels in width (at 300μm/pixel resolution) in those regions that have not been captured by the CLS detector. At this stage, this inadequacy is simply noted for completeness, although it is reiterated that this situation is simply a consequence of the current implementation of the CLS detector algorithm, rather than a fundamental constraint of its formulation. This point will be discussed further in section 7.2.

Note that the CLS detector has not only identified the CLS features but also the edges

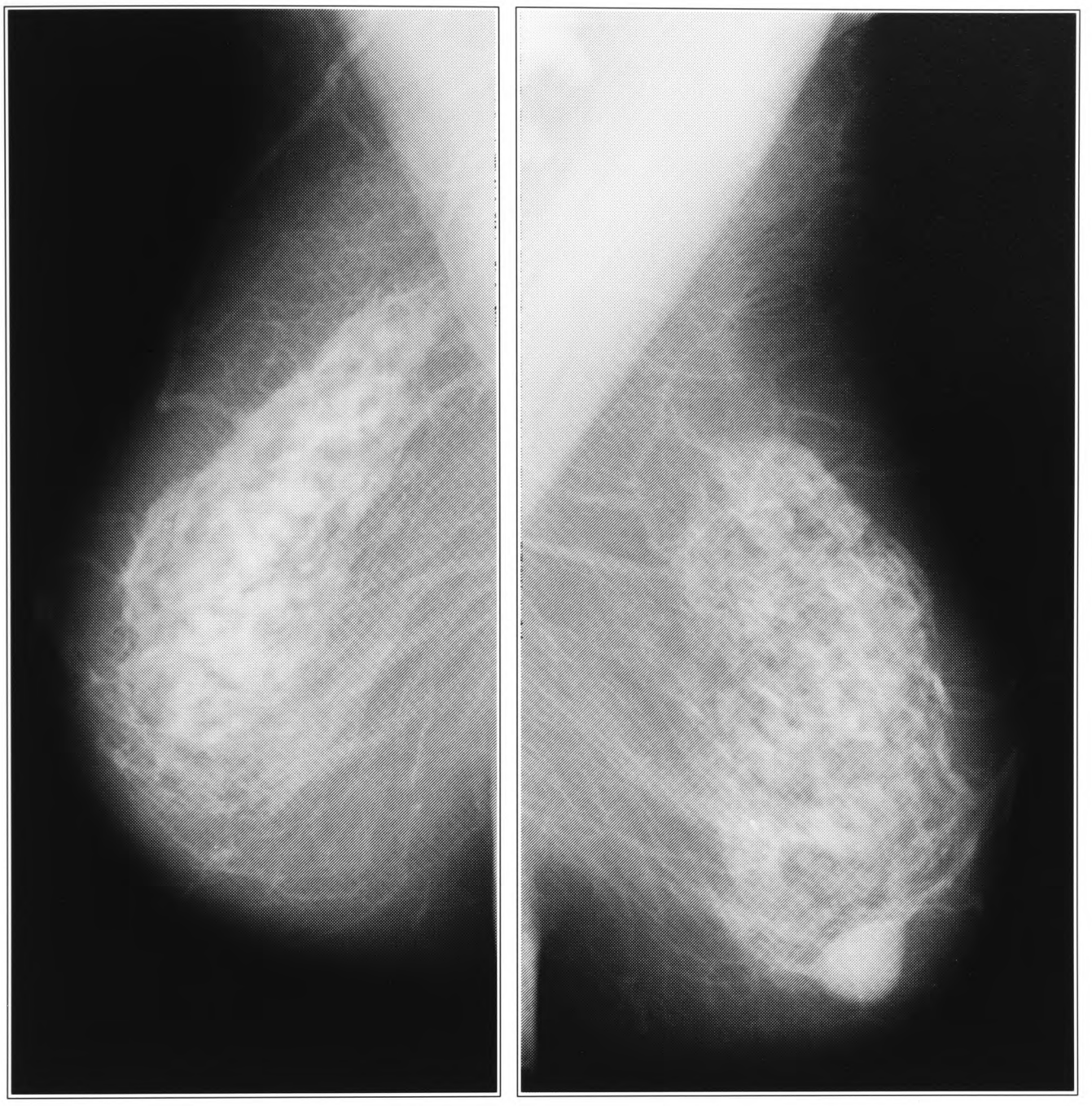


Figure 4.32: Example mammogram image pair.

of some larger features. This is a consequence of the fact that edges contain large negative second difference components as previously mentioned. Observe the 'CLS' features identified in figure 4.33 around the border of the large density in the bottom of the right image of figure 4.32. This error is simply noted in passing at this stage, however it will be addressed further in section 5.1.

4.9 Stability Analysis

With the results from the CLS detector algorithm now available, the obvious question that must be considered concerns the accuracy of that description. That is, are the results displayed in the last section correct?

The task of determining the correctness of the CLS identification determined by the CLS detector described in the preceding sections is a non-trivial problem that is chiefly constrained by the lack of a known correct answer against which the output of the CLS detector can be

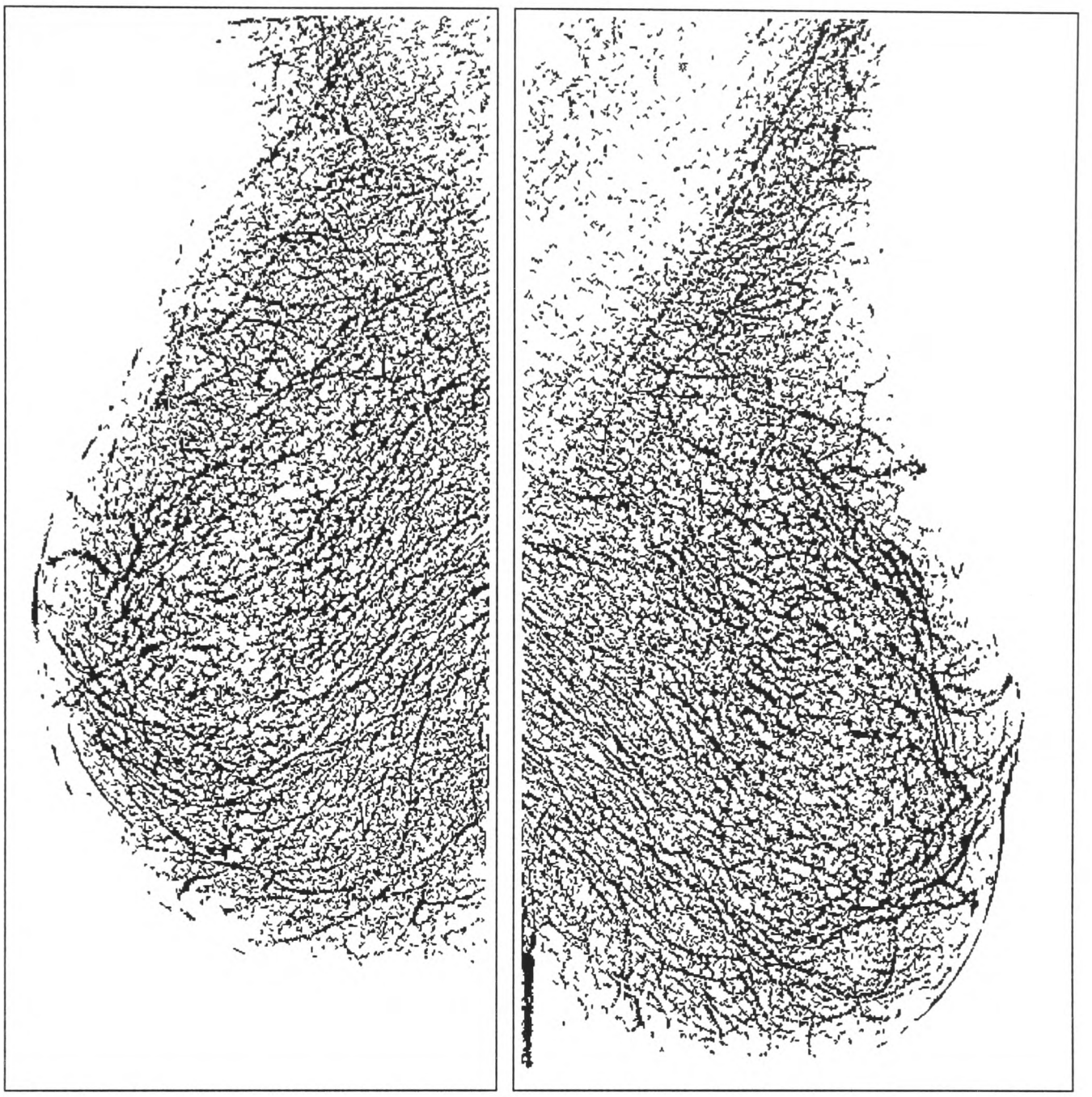


Figure 4.33: CLS pixels identified in the images of figure 4.32.

compared. Such a situation is often tackled in the computer vision community by analysing synthetic data, where contrived images are manufactured such that the results of analysis can be compared against a known correct answer. The difficulty here is ensuring that the synthetic images are accurate simulations of real mammographic features, particularly the CLS features. One might try to simulate the CLS features on the basis of the model of the CLS profile developed in section 4.4, although care must be taken to ensure that the significance of this approach is not compromised (since the detector is designed to identify instances of the very model used to generate the simulated data).

Given the existence of an image with a priori known locations of CLS features, then clearly the assessment can proceed directly. The difficulty is acquiring such an image. Cowen et al. [43] describe a mammographic phantom, containing a region that aims to more realistically simulate the appearance of breast tissues than traditional phantoms. The authors note however, that they "do not claim that this pseudo-breast structure is tissue-equivalent either in the

radiological or anatomical sense". Whilst an image of this phantom may or may not be suitable for assessing the accuracy of the CLS detector, it is noted that in general such assessment may be possible. Within the scope of the present work however, such an image was not available, and thus we have evaluated the algorithm's performance by determining the *stability* of the results against variations in the factors that influence the output. Before proceeding to consider the stability of the CLS detector, it should be noted, that although this is an unacceptable answer to the question of accuracy, the results do accord well with the visual identification of CLS features as found by a resident Radiologist. Furthermore, the results have supported the subsequent development of analysis tools (chapters 5 and 6 for instance) that depend upon the accurate identification of the CLS features. The fact that these tools do actually work, provides an 'existence-proof'—of sorts—that the CLS features identified by this detector are accurate. With that said, we proceed to consider the stability of the detector.

As formulated in the sections above, the output of the CLS detector for a given image is a function of both the threshold T (of section 4.4.9) and image noise. Thus we consider the performance of the algorithm as a function of these parameters.

4.9.1 Second difference response threshold T

Recall from section 4.4.9 that the nominal default value of T was empirically found to be $T = -t\sigma$, where $t = 0.7$ and σ is the standard deviation of the second difference values calculated over the breast region for the corresponding filter kernel.

To assess the stability of the CLS features identified as a function of t, consider in the first instance the results obtained from application of the algorithm at a number of different values of t through the range $0.3 \leq t \leq 1.1$ as shown in figures 4.34 and 4.35.

Note that at small values of t, proportionately more regions of the image will be identified as having a strong second-difference response (see equation (4.16)), and consequently more pixels are identified as components of CLS features, and vice versa. This result is reflected in the images of figures 4.34 and 4.35. Observe the (qualitative) stability during the progression from low to higher values of t such that the CLS features that are no longer identified at a given stage were the least salient features identified during the previous stage. Conversely, those CLS features consistently identified over the range of values of t are visually the most salient. This is hardly surprising when the formulation of equations (4.16) and (4.17) is considered.

Figure 4.36 considers a more quantitative assessment of this stability, plotting the percentage of breast-area pixels identified as components of CLS features against the value of t. Each curve (there are 16) of this figure represents the variation for a separate image, selected semi-randomly from the MIAS Mammographic Database chosen to represent the variety of mammograms that present in practice[15]. Note that from the perspective of stability, a vertical displacement of these curves is irrelevant since that is simply a measure of the absolute CLS

[15]The 16 images consist of 6 'Glandular' images (MIAS image numbers 015, 016, 021, 022, 071 and 072), 6 'Fatty' images (025, 026, 075, 076, 077 and 078) and 4 'Dense' images (099, 100, 101, 102). The Glandular/Fatty/Dense distinctions are classification terms of the MIAS Database [204].

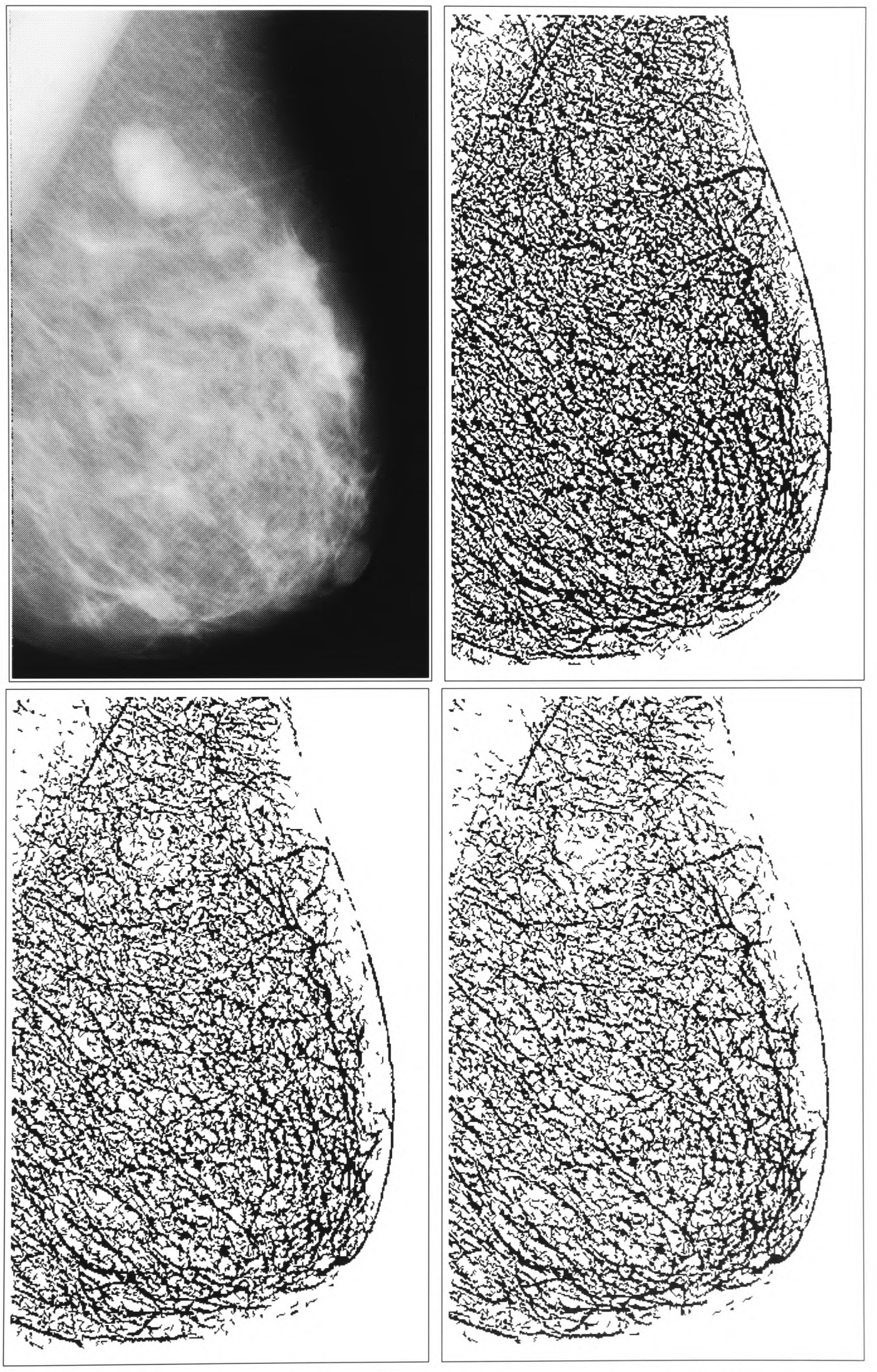


Figure 4.34: Stability of the CLS detector versus t. From top to bottom, left to right: original image, CLS pixels for $t = 0.3$, $t = 0.5$ and $t = 0.6$.

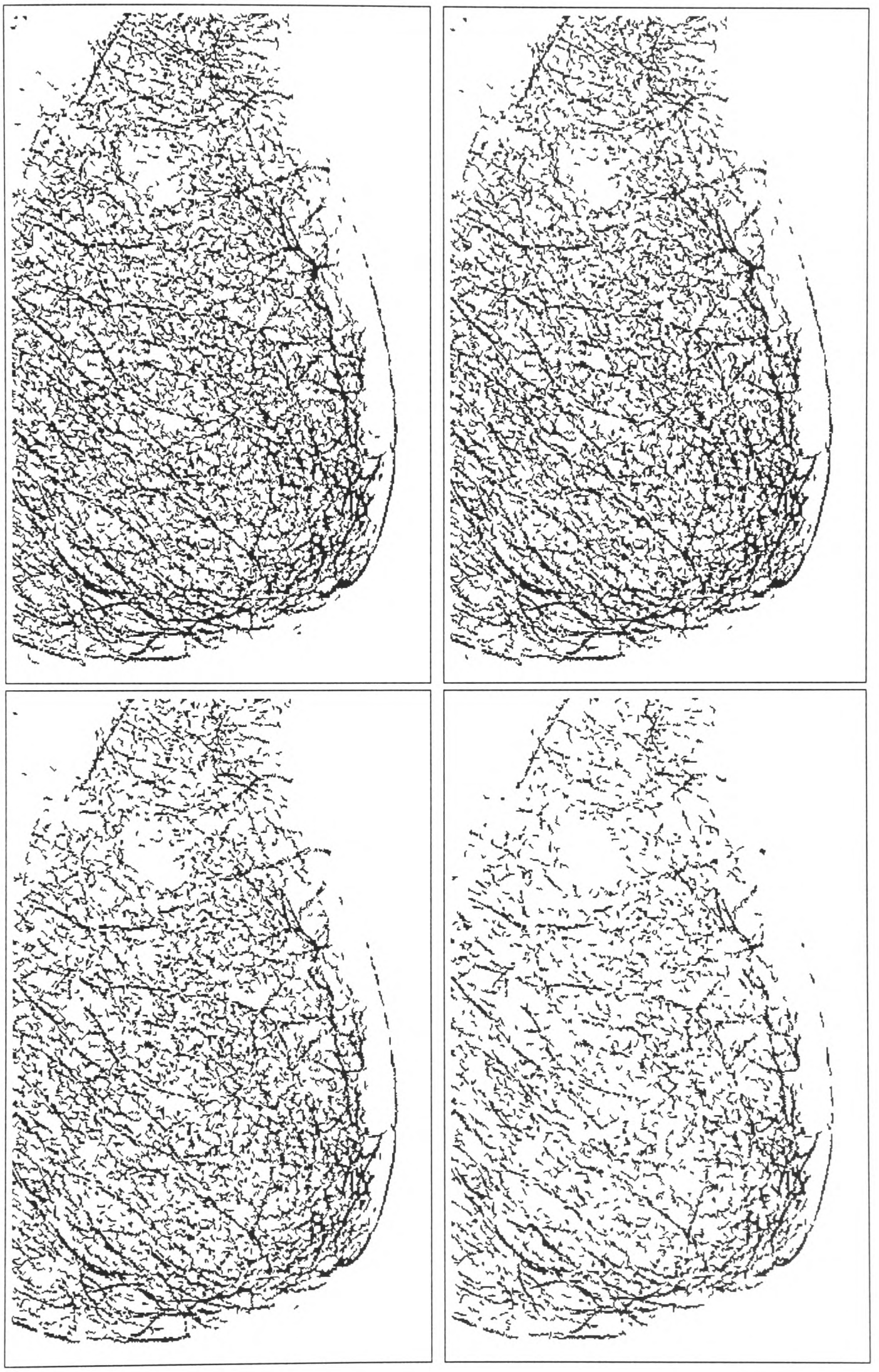


Figure 4.35: Stability of the CLS detector versus t. From top to bottom, left to right: CLS pixels for $t = 0.7$ (the default), $t = 0.8$ and $t = 0.9$ and $t = 1.1$.

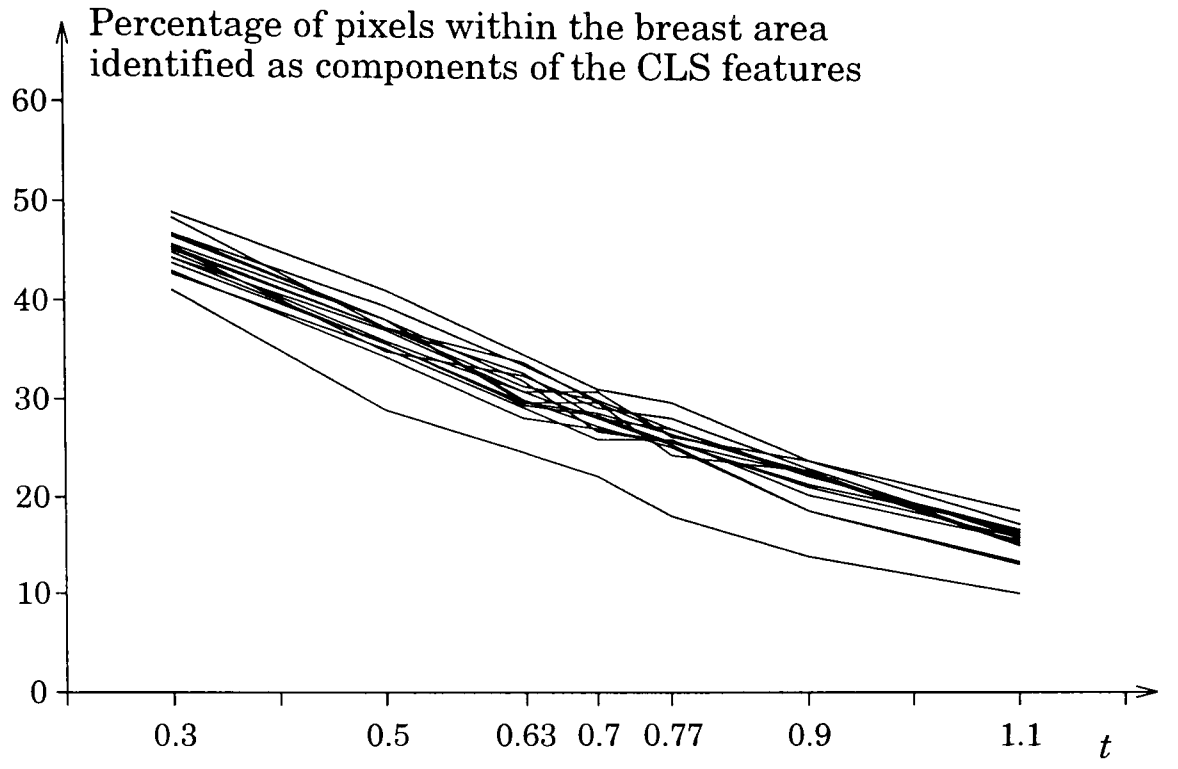


Figure 4.36: Variation of the percentage of a breast area identified as CLS features versus the second-difference threshold T for a range of mammogram images. Note that the domain of this figure shows the values of t, where $T = -t\sigma$, and σ is the standard deviation of all the second-difference response values.

activity in a given image. What is important is the variation of the curve with t, that is, the smoothness and gradient of the curves. A large (negative in this case) gradient would indicate that the output was highly dependent upon the value of the threshold, whilst discontinuities in the curve would indicate gross instabilities in the output. To quantify the gradient of these curves, consider their variation in the $\pm 10\%$ region surrounding the nominal default value of $t = 0.7$, that is $0.63 \leq t \leq 0.77$. For the 16 curves of figure 4.36 the variation of the number of pixels identified as members of the CLS features is $(4.9 \pm 1.3)\%$ over the 20% variation of t. Clearly then the CLS detector algorithm is stable to variations in t.

4.9.2 Noise

In this section we consider the stability of the CLS detector to the degrading influence of additional image noise. Before proceeding, it should be borne in mind that the formulation of the CLS detector did not attempt to suppress the noise per se, or to determine those second-difference responses due to noise as opposed to real underlying CLS features in the image. However, as noted previously, the use of the median filter to reduce the resolution of the images to $300\mu m$/pixel does naturally incorporate some suppression of noise. Although no systematic attempt is made to isolate any remaining noise responses, a simple procedure that minimises their effects is included in the algorithm formulation. Recall from section 4.4.10 that the identification of CLS pixel regions suppressed those free standing regions of area smaller than a given threshold under the premise that the CLS features are spatially more extensive than that, and consequently within the context of this problem those regions could be considered as

insignificant. Obviously, a high proportion of such features will be due to (positive[16]) image noise (or possibly calcifications, etc.). Furthermore, those true regions of a given CLS feature that were not identified as such due to a *negative* noise process (a drop-out) are accommodated during step 2 of the CLS detector algorithm (section 4.3) which fills the small holes in the CLS pixel set previously identified. Further to these techniques, there is no explicit identification of the second-difference responses due to noise (such as Fleck's third-differences analysis for noise suppression in the *Spectre* detector—see section 4.2). Consequently we simply state here for completeness the stability of the detector (as currently formulated) against the effects of additional noise artificially introduced into an image.

To evaluate the performance of the algorithm in the presence of image noise, consider the difference between the CLS features calculated for an original image, and variants of that image with known introduced Gaussian noise. Figures 4.37 and 4.38 show the pixels comprising the CLS features identified in the original image of figure 4.34 corrupted with white Gaussian noise of zero mean and variances: 0 (the uncorrupted/original image), 0.5, 1, 2, 3, 5, 8 and 12.

Figure 4.39 shows the variation of the total number of CLS pixels identified versus the variance of the introduced noise. Note that the relative stability of the number of pixels identified is due to the automatic adjustment of the threshold T, rather than a stability in the CLS pixels identified at each stage. Observe (figures 4.37 and 4.38) that with increasing noise variance it is the most salient CLS features that are extracted consistently, whilst the less salient features are lost within the noise response. It is unlikely that an image would ever contain noise with $\sigma^2 = 12$, particularly considering the images have been median filtered, and therefore we simply note these results without addressing the issue further. Clearly if noise of this magnitude was expected, some of the steps of the CLS detector algorithm would require additional procedures to directly address the noise.

4.10 Summary

This chapter has sought to systematically address the difficulties posed by the CLS features when assessing mammograms for disease processes. Previous attempts at automation have not addressed these features directly, often requiring a subsequent procedure to address the complexity introduced into the results of these methods by the CLS features. Without exception, these procedures have addressed this complexity inappropriately.

We have begun with a model of the CLS features based on the assumption that these anatomical features adopt (near) circular profiles in the uncompressed breast, and therefore, (near) elliptical cross-sections in the compressed breast. This assumption has lead to the development of an algorithm to identify the CLS features in an image, utilising the negative image surface curvature associated with their presence, and their attributes of nominally one

[16]Recall that the CLS pixel detector identifies regions of increased image intensity, thus a noise feature giving rise to a strong second-difference response as formulated in this situation must have *increased* the local image intensity and is therefore a *positive* noise perturbation of the image surface.

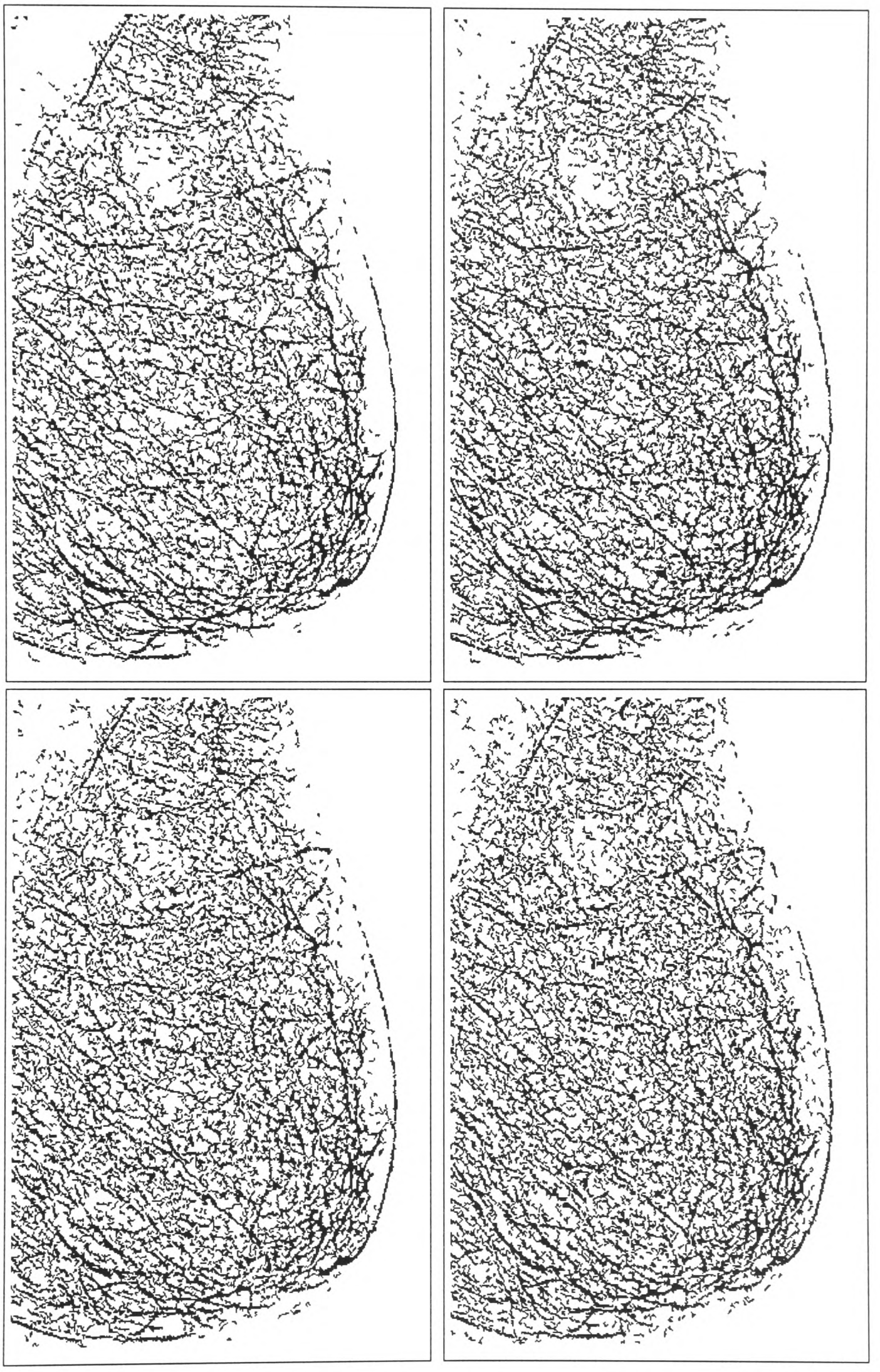


Figure 4.37: Stability of the CLS detector versus noise. From top to bottom, left to right: CLS pixels for $\sigma^2 = 0$ (original results), $\sigma^2 = 0.5$, $\sigma^2 = 1$ and $\sigma^2 = 2$.

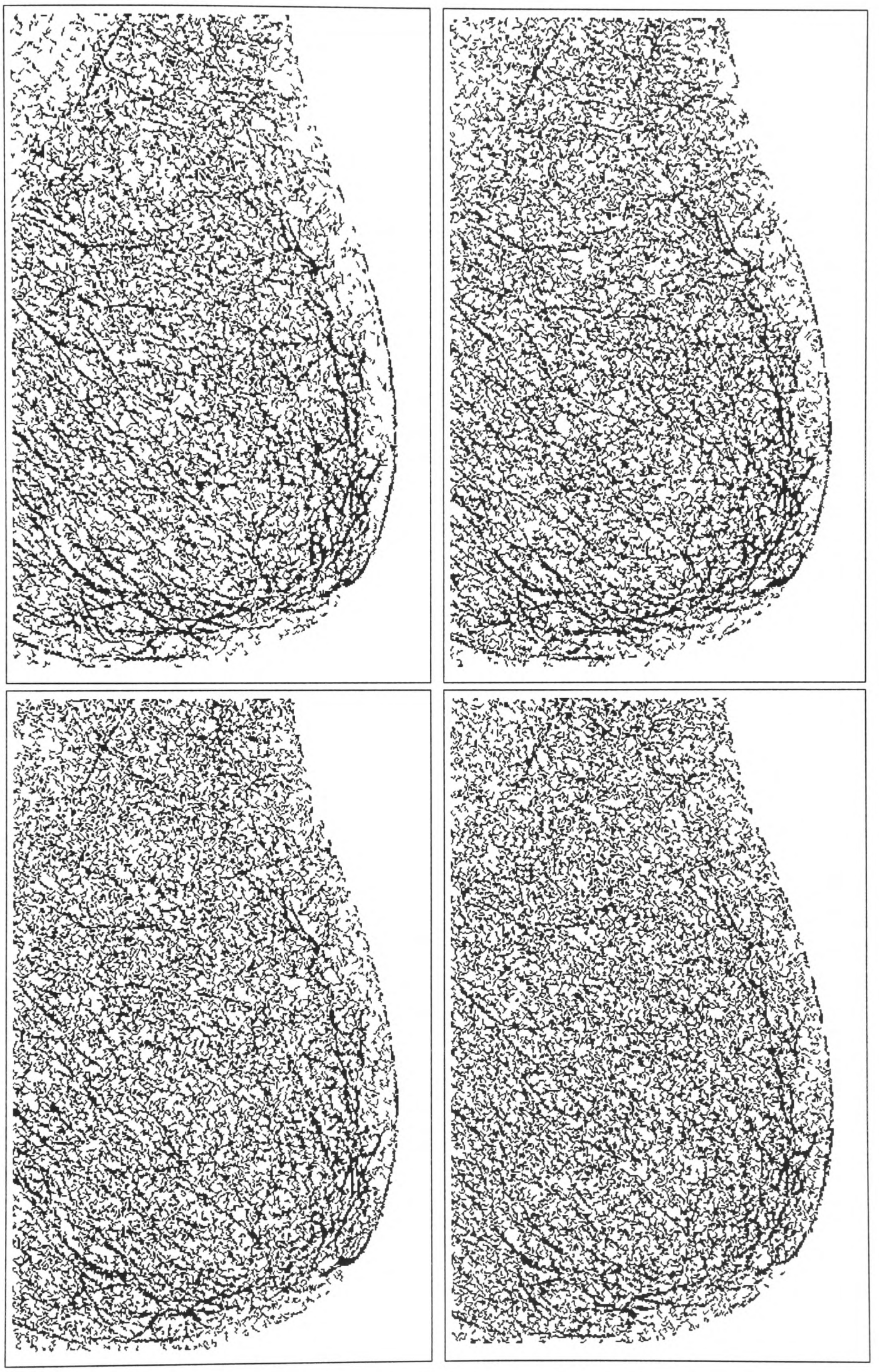


Figure 4.38: Stability of the CLS detector versus noise. From top to bottom, left to right: CLS pixels for $\sigma^2 = 3$, $\sigma^2 = 5$, $\sigma^2 = 8$ and $\sigma^2 = 12$.

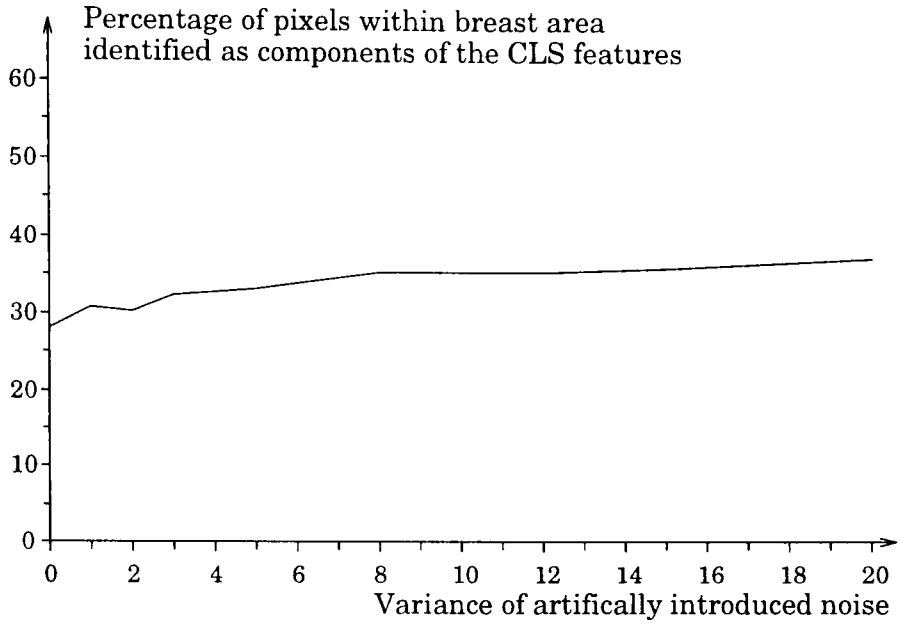


Figure 4.39: Variation of the CLS pixels detected in an image versus the variance of artificial noise introduced into the image.

dimensionality and near linearity.

We have established that CLS features less than 200–300μm cannot be identified in an image. The reduction of an image's spatial resolution from the typical digitisation values of 50 and 125μm/pixel using median filtering, reduces the computational expense of processing, simplifies the implementation of second-difference kernels used to identify the regions of negative image curvature, and reduces the noise present in the original image. Consequently the formulation of the CLS detector algorithm assumes the spatial resolution of the images has been reduced to 300μm/pixel.

A given mammogram film can be digitised in many ways, typically giving an image with intensities representing either the film-density or the luminance of the light transmitted through the film. Although the image surface curvature at the site of a CLS feature has mildly greater magnitude in the transmitted-light representation than in the film-density representation, we have concluded that the film-density image representation should be used for identifying the CLS features. This result is derived from the observation that the transmitted-light representation is coupled to many more (variable and uncontrollable) imaging parameters than the film-density representation. Therefore the results from analysing the film-density representation rather than the transmitted-light representation, will more accurately represent the structure of the breast, rather than the particular imaging conditions under which the mammogram was acquired.

From the CLS pixels, identified as those with high negative image surface curvature, a high-level feature description of the CLS features is extracted. The algorithm for extracting this feature description exploits the attributes of the CLS features known a priori (1-dimensionality and local linearity) to group the CLS pixels together into higher-level data structures describing the CLS features themselves.

The CLS feature description is a high-level feature description identifying given re-

gions/pixels of an image as belonging to a given CLS feature. The feature description is organised as a list of pixels comprising the CLS itself and another list identifying the skeletal pixels of each CLS. This data format has supported further image processing and interpretation, examples of which are to be developed in the following chapters.

5

Applications of the CLS feature description

Chapter 4 described a technique for extracting a description of the CLS features from a mammogram image. It is possible to apply this feature description to a number of applications, many of which were introduced in the introduction to that chapter. In this chapter we consider some of the *immediate* applications for this description, which in some cases, lay the foundations for further processing and image interpretation, such as the mammographic density detector developed in the next chapter.

5.1 Simulation of an image without CLS features

In chapter 3 it was noted that some authors had experienced difficulties analysing mammograms due to the undesirable response of the CLS features to their algorithms of analysis. Indeed, a major motivation for seeking a description of the CLS features was the hope that such a description might allow the explicit accommodation of their influence and effects within the formulation of any such algorithm. Given the description of the CLS features courtesy of the previous chapter, this is an application which we are now well placed to consider.

There are two main approaches that might be considered at this juncture; in a reactive approach, correlate the results of an algorithm with the known location of a CLS feature, or in a pro-active approach, pre-empt the undesirable algorithm responses by *removing* the image features that unfortunately give rise to those responses *prior* to the application of the algorithm. Although the first approach can, in theory, accomplish this task, it is a tedious approach that does not make full use of the information available from the feature description, since it is only used to correct for the mistakes of other algorithms, rather than to develop

new or improve existing ones. The second approach is the altogether more elegant solution, although its application must be accompanied by a healthy measure of caution. This is due to the fact that in altering an image to remove the features causing the difficulties, we must be careful not to remove those features forming the subject of our search/investigation.

Recall from section 4.1 that the main motivating factor for the identification of the CLS feature description was to account for the high-frequency information associated with those features when searching for mammographic densities (which occur with lower spatial-frequency). Recall also that one of the main considerations during the assessment of mammographic densities is the condition of the density boundary: for instance circumscribed or spiculated, blurred or sharp? Obviously these features can always be assessed in the original image irrespective of alterations made during this procedure, however it may be the case that indiscriminate alterations would render the transformed image unsuitable for analysis since some of the attributes defining the class of interest (the densities, say) may also have been removed. For these reasons we seek to remove from an image as many features giving rise to the complications as possible, whilst leaving as much of the image as intact as possible. Clearly within this context we seek to remove the influence of the high-frequency CLS, yet retain the remainder of the image intact for the purpose of simplifying the identification of mammographic densities.

Consider simulating an image that is in all respects the same as an original image, except that the CLS features have been 'removed' by resetting their pixel intensities to a value consistent with the 'background' image intensities in the neighbourhood surrounding each CLS pixel.

Consider the model of a CLS feature shown in figure 4.13 and the image surface profiles that consequently result, shown in figure 4.18. For the purpose of this exercise imagine the somewhat fanciful concept of a breast without CLS features, that is without the anatomical components that give rise to the CLS features in images. Obviously the corresponding surface profile of the model would be a constant intensity at the value of the region surrounding the CLS features in these figures, that is a straight (horizontal) line joining the left and right tails of the graphs of figure 4.18. Thus the new intensity to assign to a CLS pixel is calculated by linear interpolation between those intensities of the background on either side of the current CLS feature. An example of this concept is displayed in figure 5.1 which shows a projective surface elevation (image intensity represented by surface height) of a CLS feature traversing a small image patch. For added realism, the 'background' intensities on either side of the CLS (the 'ridge' of figure 5.1(a)) are at different values. By moving along the axis of the CLS feature and considering the intensities of each cross-section in the direction orthogonal to the CLS direction (the bold line in figure 5.1(a)) the intensities of the background can be measured and the new values for the CLS pixels interpolated between them as shown in figure 5.1(b).

In practice the situation is complicated by the fact that a given pixel can, particularly near junctions of CLS features, be a member of more than one CLS feature, and thus the local orthogonal direction is multiply defined. In this case the final pixel value is calculated by

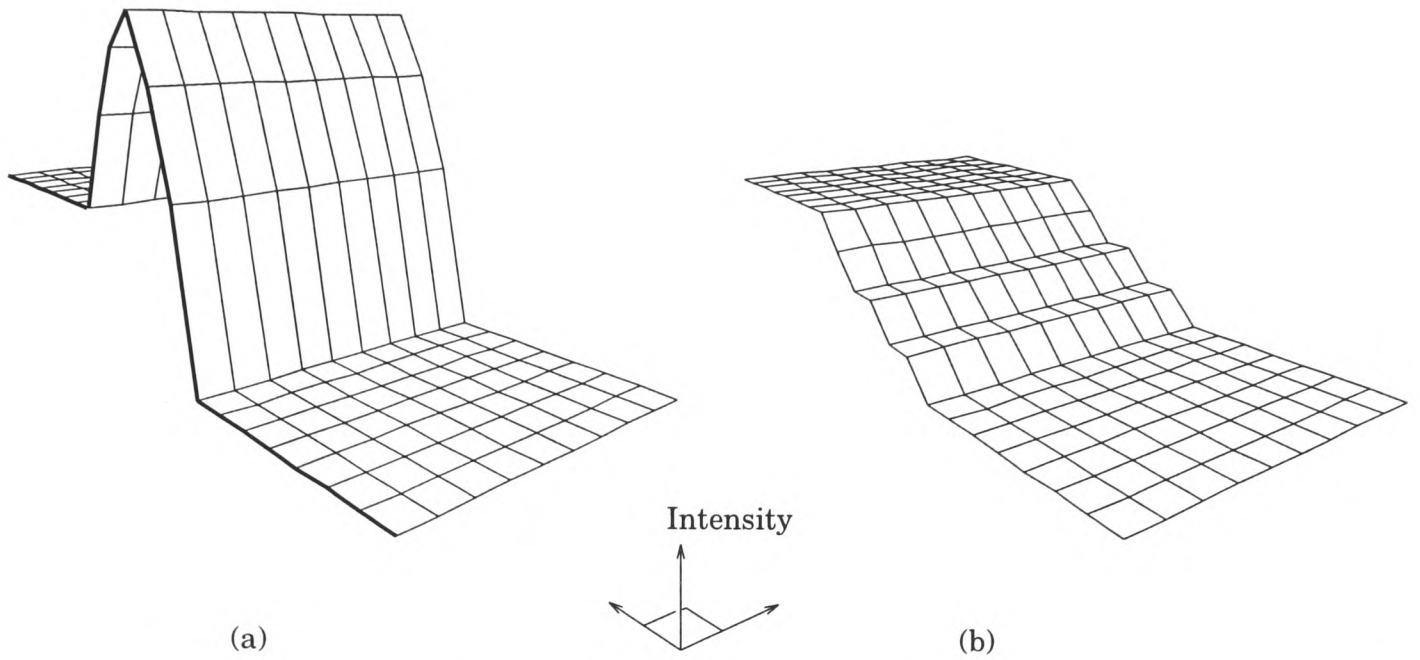


Figure 5.1: Removal of a CLS feature by linearly interpolating new intensity values between the background values at the CLS boundary. (a) a CLS feature and (b) without the feature.

considering all the values suggested for that pixel by each of the interpolations it is involved in.

Although it would appear a relatively simple task to determine the background intensities at either side of a CLS feature by simply searching the CLS pixel space in the orthogonal directions from the current pixel until the background is reached, due to a limitation of the CLS detector algorithm this is in fact not the ideal approach. Recall from chapter 4 that the CLS feature detector searches for regions of high negative surface curvature. Figure 4.19 shows the second-difference image surface for modelled CLS features of width $\pm 1mm$. Observe that the negative response does *not* extend to the full $\pm 1mm$ width of the feature, but rather underestimates it by a single pixel. Although on a much smaller scale, this is effectively the same result identified by Claridge [39] and Richter and Claridge [171] who note that zero-crossing (and thus second-difference based) density boundary detection schemes always underestimate the extent of mammographic densities. As predicted by the simulations of chapter 4, this effect was observed during the implementation of the CLS-removal algorithm and consequently it was necessary to search for the background a little beyond the CLS border returned by the CLS detector algorithm of the previous chapter. However to ensure the accuracy of this procedure it is naturally desirable to measure the background image intensities as close as possible to the CLS pixels to be adjusted so that their new values faithfully represent the *local* background level, and accordingly there is a trade-off that must be balanced. We have found in accordance with the simulations of chapter 4, that the background does lie within a single pixel of the boundary identified by the CLS detector, and thus we search for the background up to a pixel beyond this boundary.

In order to search for the background from the current CLS pixel, it is of course necessary to know in which direction to search. Recall from section 4.6 that the local tangential direction to the cantons of that section was measured by fitting a low order polynomial to the pixels of

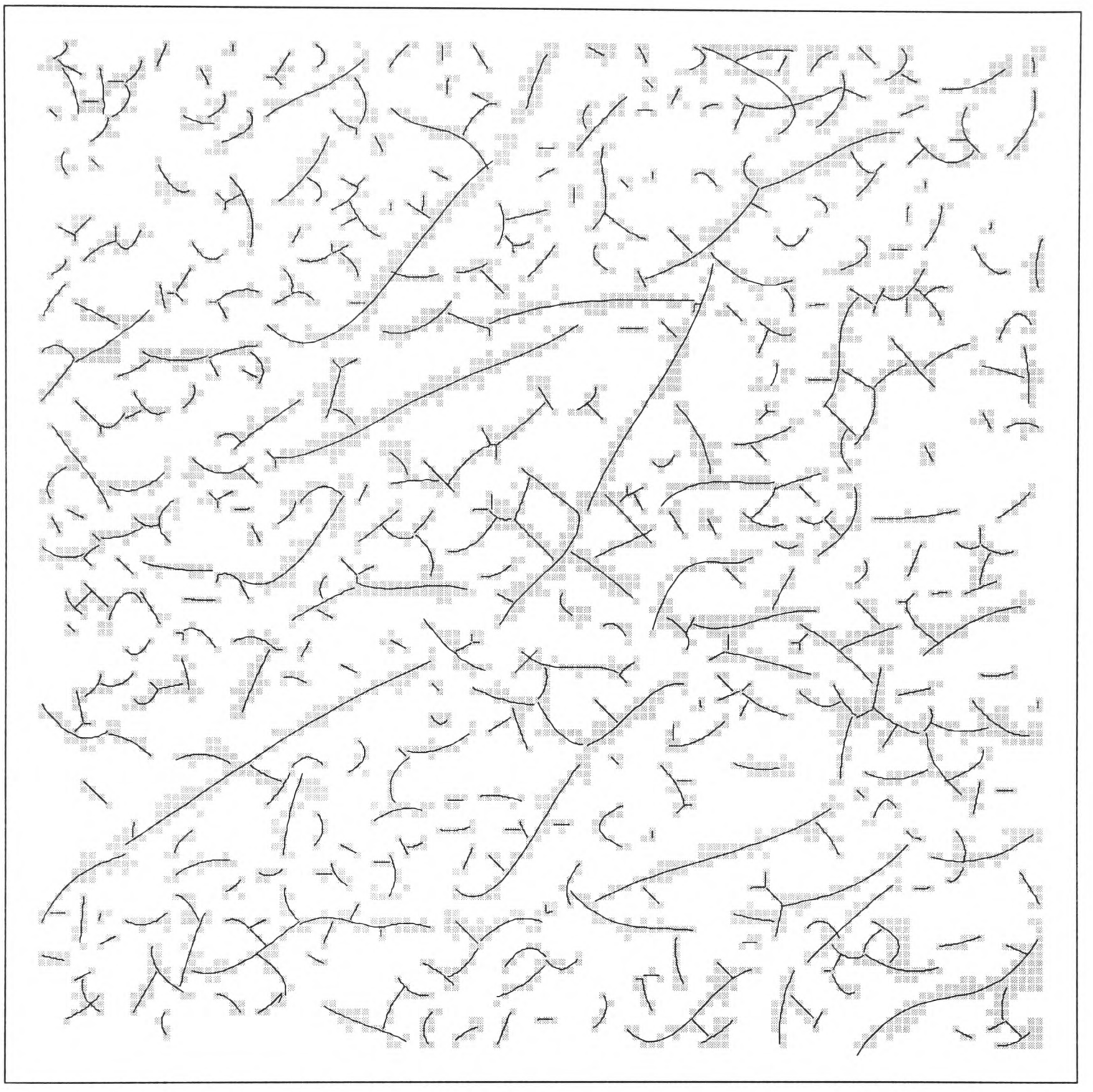


Figure 5.2: Polynomial curves fitted to the CLS features identified from the bones of figure 4.28. These curves allow the calculation of the local normal direction at each CLS pixel, and subsequently the identification of the background on either side of the CLS.

those features. Similarly, since we have a high-level description of the CLS features we can, in much the same fashion, fit a polynomial to the CLS features and measure the local normal of the polynomial to determine the search direction. Since we wish to measure the *local* normal direction, rather than maintain some global (over the CLS) influence as desired in section 4.6, we use polynomials of slightly higher order than those used in that section, typically fourth or fifth order. As an example of this procedure, consider the bones identified in figure 4.28 from the cantons of figure 4.27. Figure 5.2 displays the fourth order (cubic) polynomials fitted to the bones of figure 4.28 which are used to identify the local normal direction to each CLS feature.

There is another complication to consider prior to implementing the removal process as

described above. Recall that the CLS feature detector responds to regions of high negative curvature in the image surface. It is this fact that Fleck utilised in the development of the *Phantom* [65] and *Spectre* [66] *edge* detectors since a negative image surface curvature is a component of any edge, such as the edges of CLS features and the edges of densities. Consequently it is generally the case that the CLS detector may have identified the edges of some densities if the image surface displayed a strongly negative curvature at that point. Fortunately this is a relatively easy problem to correct since the identification of the current CLS as an edge of a larger feature rather than a true CLS feature (or perhaps as a component of a CLS feature not fully identified due to an insufficient spatial scale in the filter kernel as mentioned in chapter 4) is easily established from the relative intensities of the current pixel and those of the appropriate background pixels, which were identified in the preceding paragraphs. A current CLS is identified as a true CLS if the intensities of the CLS pixels are greater than both background intensities (i.e., on both sides of the identified CLS feature) and if the minimum difference between the CLS and background intensities is greater than some threshold. More specifically, we identify a CLS that has captured the full width of a true CLS as one that meets the following conditions:

$$\begin{aligned} I(p) &> I(b_1) \\ I(p) &> I(b_2) \\ \min\big(I(p)-I(b_1), I(p)-I(b_2)\big) &> 0.2\max\big(I(p)-I(b_1), I(p)-I(b_2)\big) \end{aligned} \tag{5.1}$$

where $I(p)$, $I(b_1)$ and $I(b_2)$ are the image intensities at the CLS pixel p and the corresponding background pixels b_1 and b_2 respectively.

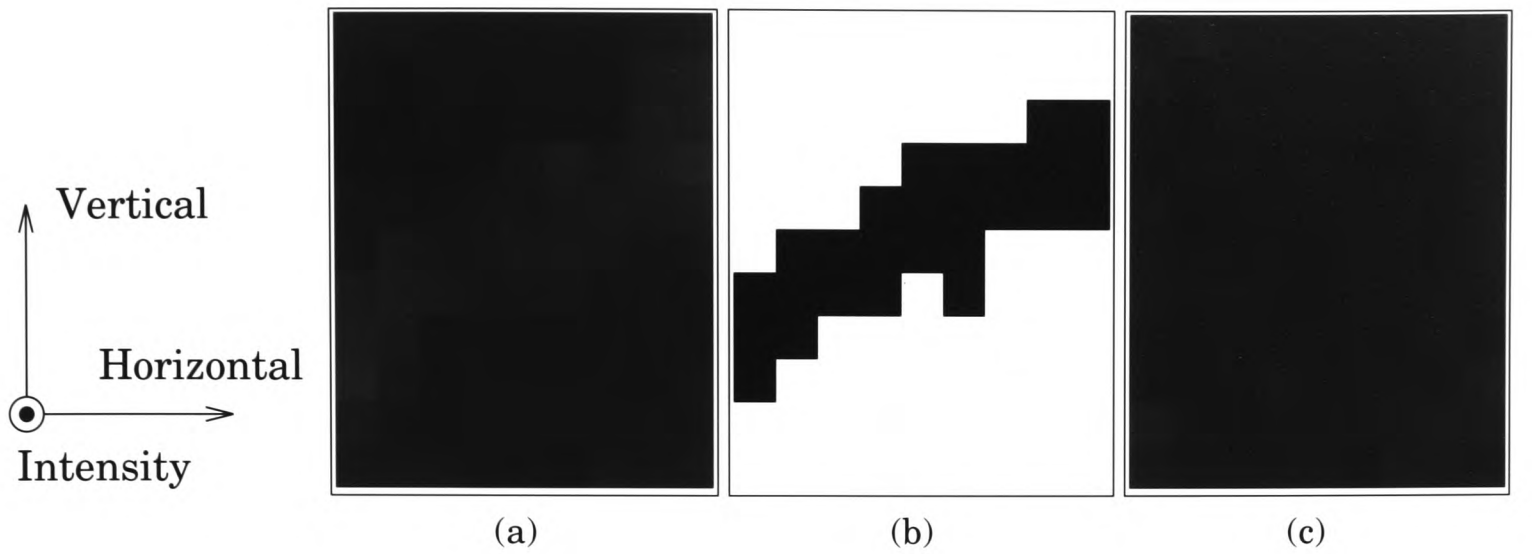
In summary then, the algorithm for removing the CLS features from an image can be described as follows. For each CLS feature:

1. Fit a low order polynomial to the pixels of the bone $\mathbf{b}_i$ corresponding to the current CLS. Since polynomials are single-valued functions, in practice the bone pixels are least-squares fit to two polynomials, using respectively the x-axis and the y-axis as the polynomial domain. The fit with the smallest residual fitting error is selected as the best-fit of the data. Typically we use fourth or fifth order polynomials.

2. At each pixel of the CLS, look in both orthogonal directions for the edge of the CLS and subsequently the pixels defining the local 'background' intensities as discussed above.

3. If the current CLS pixel and background intensities comply with equation (5.1), then linearly interpolate new intensities for *every* pixel on the linear path joining the current pixel and the two 'background' pixels. Keep a record of *each* new value calculated for each pixel by this process.

4. Assign each pixel the median value of all the new values calculated for it during the previous steps.



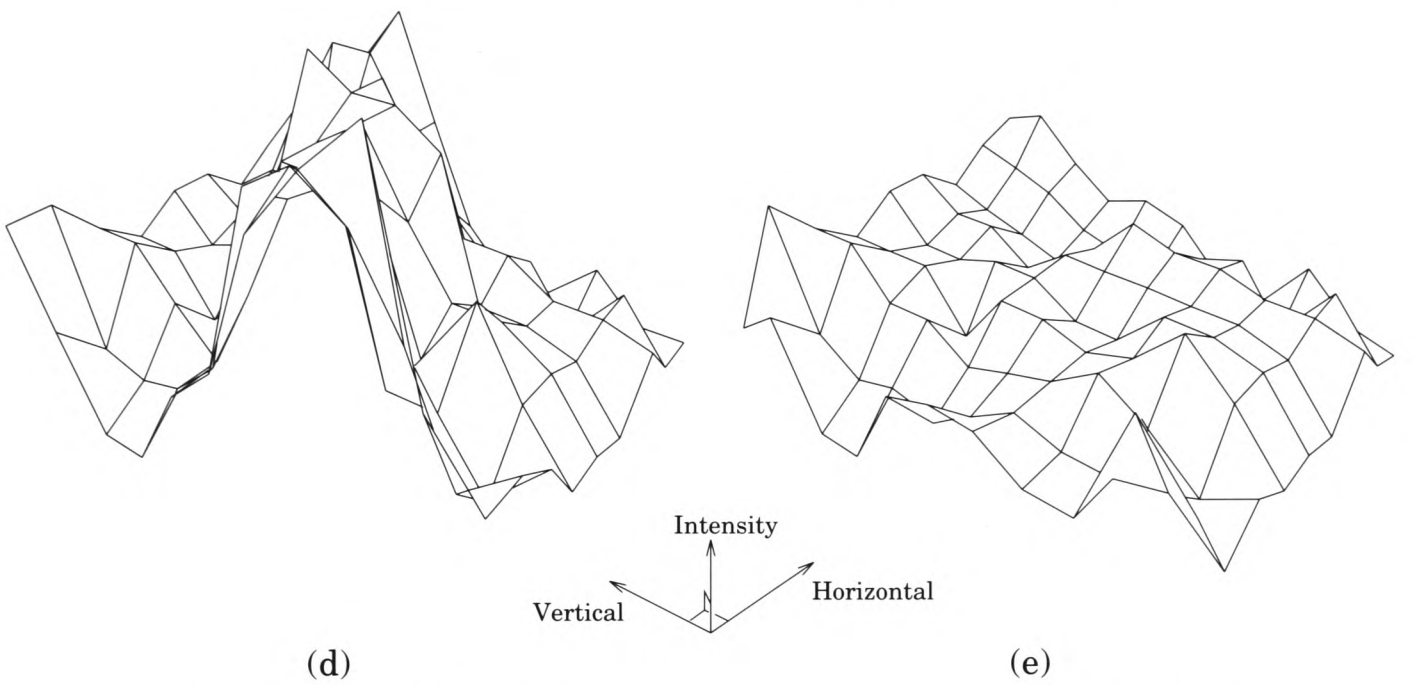


Figure 5.3: Removal of a CLS feature from real image data. (a) the original image, a 9×11 pixel patch extracted from a full image, (b) the CLS pixels found by the CLS detector, (c) the CLS 'removed' image, (d) projective surface plot of (a) where image intensity is represented by surface height. Projection is from the bottom left corner of (a), looking end on at the CLS feature, and finally, (e) projective view of the CLS 'removed' surface of (c) from the same viewpoint as (d). Note the stability of the image surface to the sides of the CLS. Note also the differences between the surfaces at the extreme left and front/low-centre ends of the 'removed' surface (the top left and bottom left corners of the image respectively), due to the effects of neighbouring CLS features in the full image that are not shown in this extracted patch.

Figure 5.3 shows the results of applying this procedure. A small image patch containing a CLS feature traversing the patch has been extracted from a whole image to illustrate the process.

Observe that the result of applying this procedure has been to remove the CLS feature as desired. More importantly, note the stability of the image surface in regions that have not been identified as CLS features. Although this result follows directly from the algorithm as it is presented (since the intensities of these regions are not adjusted) this is clearly the key to retaining the image features of importance whilst removing those giving rise to the difficulties. As an illustration of the stability of the image surface at density boundaries,

consider the images of figure 5.4 which show the CLS-removed image for the original image of figure 4.30(b). Also shown are enlarged views of the density of this image for both original and CLS-removed images. Note that although the CLS features have successfully been removed, at the location of the density boundary the CLS-removed image surface has retained the specific characteristics of the original image. This is illustrated by the graph at the base of the figure which compares the intensities of the original and CLS-removed images along the cross-sectional slice between the white markers shown in the top left (original) image.

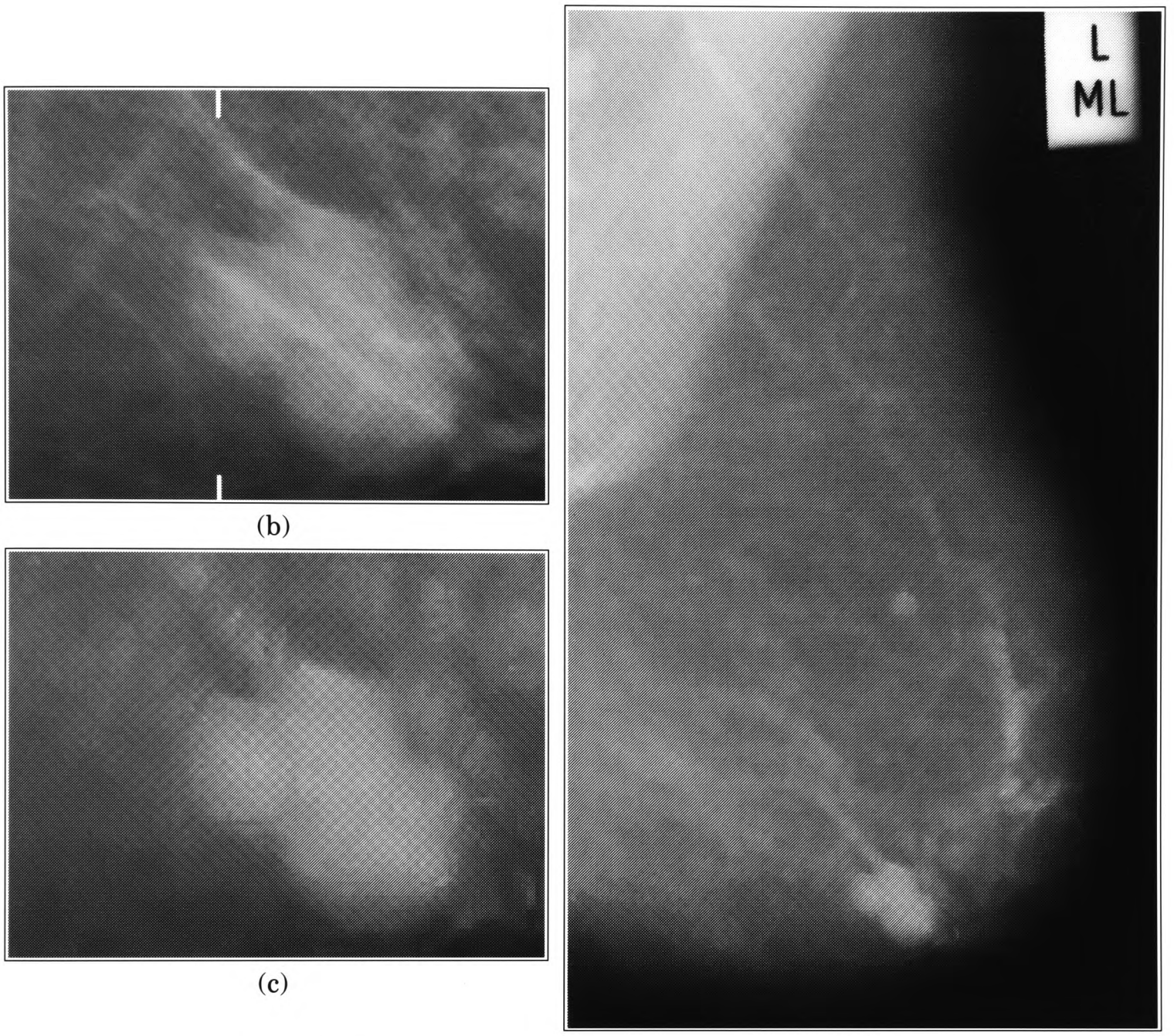
It was noted in chapter 4 that the current implementation of the CLS detector was limited by the consequences of adopting a single scale for the second-difference filter kernels, and that this would in particular exclude the identification of some portions of the wider CLS features. Such an example was identified in figure 4.31(b) which illustrates the CLS features found in the image of figure 4.30(b). We note here for completeness that as a result of failing to fully identify the large surface vein extending from the pectoral muscle of that image, we have consequently failed to fully remove that feature in the CLS-removed image of figure 5.4(a). Due to the relative infrequency of such instances, we have found in practice that the identification of masses is not consequently degraded, and accordingly we have not sought to improve the CLS detector to accommodate them. This is a point to which we shall return in section 7.2.

As another example of removing the CLS features, figure 5.5 illustrates the CLS-removed images for the original images of figure 2.7.

To illustrate the usefulness of the CLS-removed image, consider a preview of the mammographic density detector developed in the next chapter. The details of the detector are left until their development during the following chapter, however we present in figure 5.6 the bounding contours of the mammographic density identified for the original and CLS-removed images of the density of figure 5.4. Observe that the CLS feature extending towards the top-left corner of the images has been mistakingly identified as part of the density due to its high image intensity in that region. By contrast, the contour identified in the CLS-removed image correctly identifies the extents of the density. Observe that in regions other than near the CLS feature causing the mistake of figure 5.6(b), the contour identified in the two formats is essentially identical, a reflection of the similarity of the image surface in these regions, and hence the retention of the high-frequency information at the contour edge within the CLS-removed image.

Another important point to note is that the position of the boundary in figure 5.6(c) as it crosses the location corresponding to the CLS feature that has been removed is consistent with boundary of the underlying mammographic density at that location. This is a direct consequence of interpolating the new pixel intensities for the CLS pixels from the local neighbourhood of the CLS. Note that although in this particular example the CLS crosses the density border at approximately right-angles, this is not a requirement as the interpolation scheme will accommodate any scenario just as effectively.

It is worth making some general comments about the characteristics of these CLS-removed images. Note that although the CLS features have been removed from the images, the general



(b)

(c)

(a)

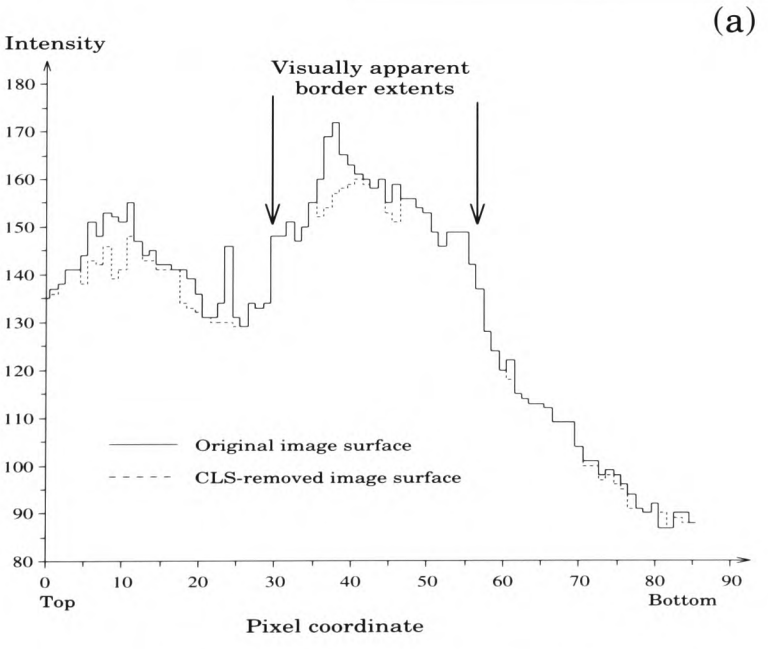


Figure 5.4: (a) The CLS-removed image for the original of figure 4.30. (b) Enlarged region of the density in the original image. (c) Enlarged region of the density in the CLS-removed image. The graph illustrates the image intensities of the vertical cross-sectional slice taken through images (b) and (c) at the location of the white markers of (b). Note in particular that at the boundaries of the density, the CLS-removed image surface has retained the characteristics of the original.

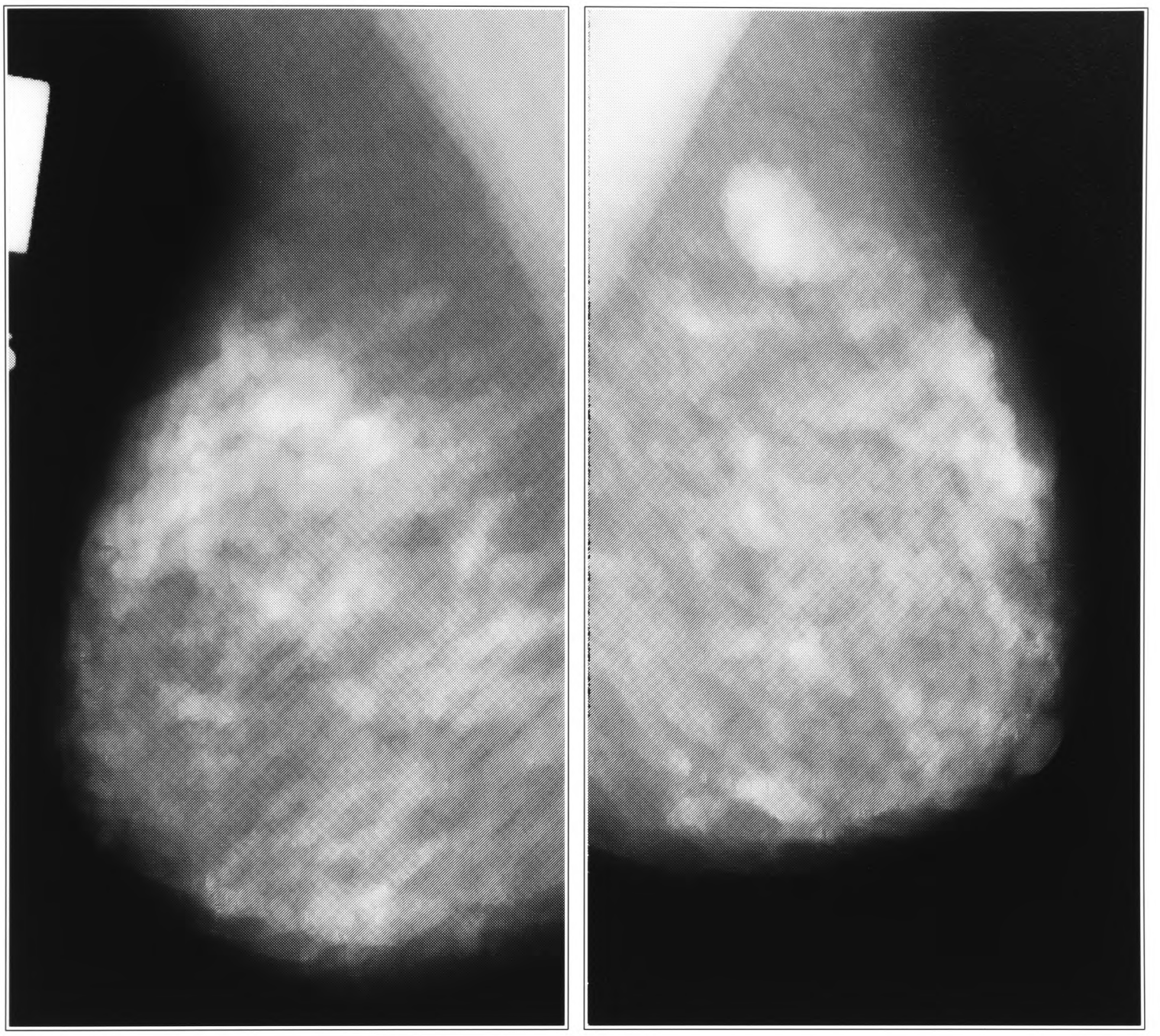
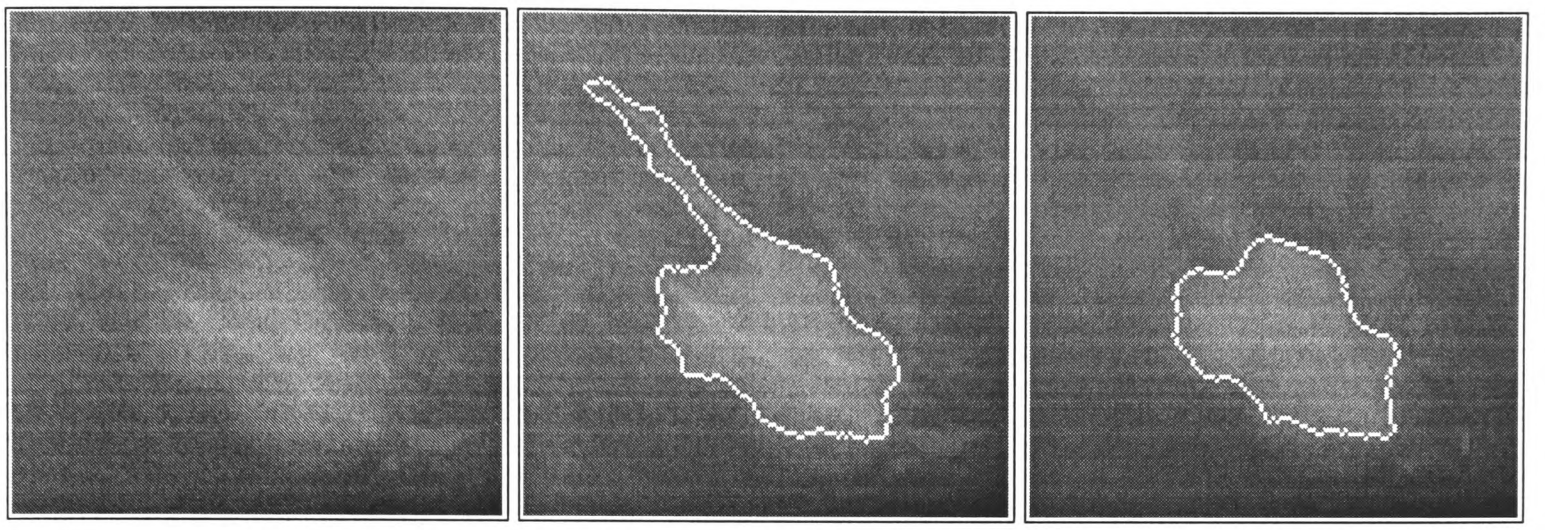


Figure 5.5: CLS-removed images for the images of figure 2.7

structure of the breast parenchymal tissues and pectoral muscle is retained. Observe with particular note the retention of all the small densities in the parenchymal tissues whilst the high frequency texture of the CLS features have been suppressed. Although the CLS-removed images take on the general appearance of traditionally smoothed images, achieved by Gaussian smoothing for instance, it should be patently clear from the example of figure 5.4 that we have *selectively* suppressed those high-frequency components of the image that cause us difficulty, whilst the very components that indicate the presence of the mammographically significant features have been retained. We can therefore use the CLS-removed image to ease the search for mammographic features such as densities, with confidence that the information indicating the presence of these features is still available. It is this confidence in the information content of the images that is not available if the images are simply smoothed to eliminate the high-frequency textural clutter of the CLS features.



(a) (b) (c)

Figure 5.6: Detail of the density of figure 5.4. (a) the original image, (b) the bounding contour identified in the original image, and (c) the bounding contour identified in the CLS-removed image.

5.2 Radial scar

Radial scar is a mammographic sign that often gives rise to nearly as many radiological opinions as there are radiologists available! Histologically it is not clear whether these fibrous tissues are premalignant or not, and consequently some radiologists are beginning to recommend surgical removal of the affected tissues as a matter of course [188]. As noted in section 2.2.2 radial scar can mammographically mimic carcinomas, mostly due to the mammographic resemblance of the sclerosing fibrous tissues to the spiculations of a spiculated mass. Since radial scar does not have the radiodense central mass associated with spiculated masses it is not possible to use a 'density detector' (such as the one developed in the following chapter) to search for these features.

Since the only mammographic features indicating the presence of radial scar is the radial pattern of CLS features, clearly the CLS feature description could be used to drive a search for them. Figure 5.7 shows an image with a relatively large radial scar process, whilst figure 5.8 shows the corresponding CLS features found in the image. Although the region surrounding the radial scar feature appears quite cluttered, observe that a large number of the CLS features radiate radially from its centre.

Whilst subsequent processing of the CLS feature domain to identify these features would not be trivial, recall from section 3.2 that it may indeed be possible. Karssemeijer [110] describes a process for locating spiculated masses using second order differential operators to determine the 'line' orientations of each *pixel*. The region surrounding each pixel is searched for other pixels with line orientations that intersect a small circle centred at the current pixel. Assembly of these pixels identifies *radial* spicules associated with the current centre. Clearly, substitution of the CLS feature description for the image pixels, would, in this context, allow a more precise assessment of the *spicules* surrounding the current centre. Since the CLS features already record information on their orientation this substitution and subsequent assessment

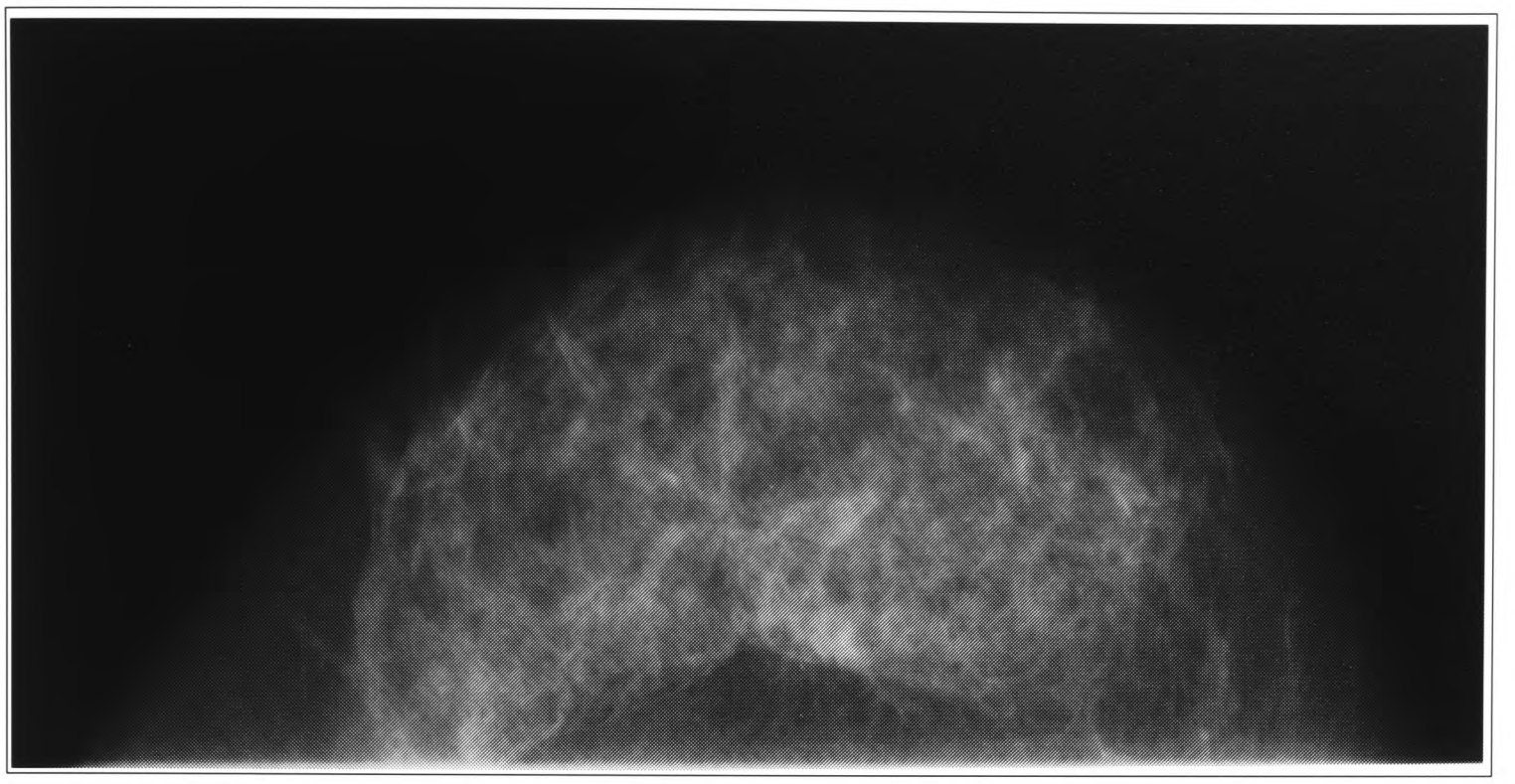


Figure 5.7: Original image with a radial-scar feature. Centre a normal Cartesian frame at the bottom left corner of the frame surrounding the figure. The radial-scar feature is centred approximately at coordinates $(74.25)mm$.

would be trivial. This potential application of the CLS feature description is simply noted for completeness.

5.3 Differential Compression Mammography

Highnam and colleagues [97, 99] describe a technique for analysing mammograms involving the acquisition of two mammograms of each breast, taken at different breast compression thicknesses. Due to the variation of material characteristics for the different tissues of the breast, and their various structural arrangements within the breast, these tissues deform due to the applied compression in both different modes and at different rates.

This information is of course, normally unavailable to the diagnostic radiologist from the single compression screening mammograms acquired today, and it therefore forms a new domain of analysis. Clearly one of the requirements of automating this analysis would involve automatically determining the differential tissue deformation between the two exposures. This is a classic example of a much broader field of computer vision, that of optic flow. Unfortunately, unlike the majority of situations considered in the general field of optic flow, the features and structures of the breast are not rigid bodies, and thus they do not exhibit the rigid body displacements that form the basis of many solution schemes. Clearly the only solution is to determine a dense correspondence mapping from features of one image to those of the other. Such a technique would require a dense feature map in each image, and obviously the CLS structures could fulfill this requirement, since they are suitable high-level tokens that could be compared. We make no claims about the relative ability of this procedure to detect breast

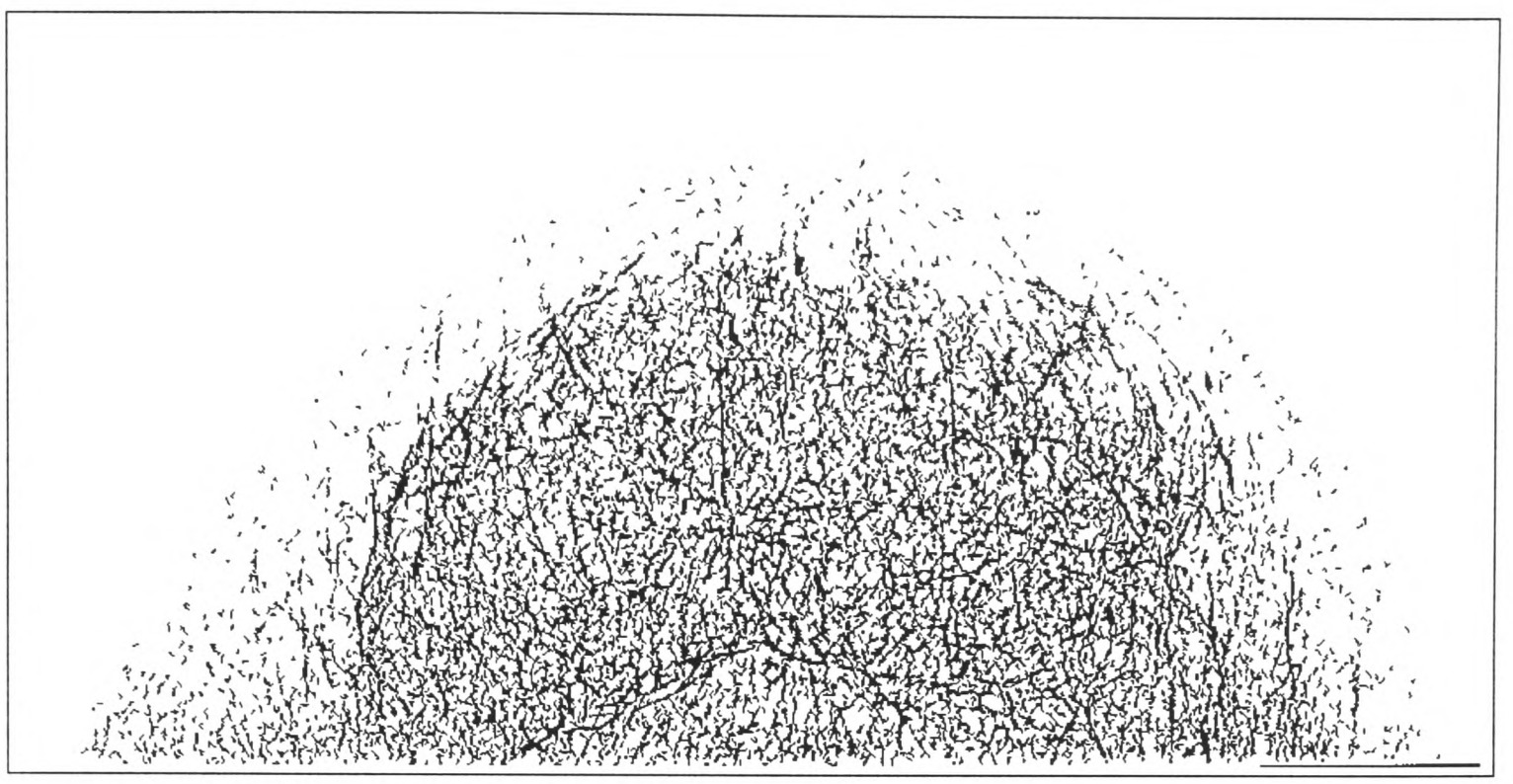


Figure 5.8: The CLS features found in the image of figure 5.7.

abnormalities, simply stating the fact that within the framework of differential compression mammography as forwarded by Highnam and colleagues, the CLS feature description would aid an automated implementation.

5.4 Asymmetry and temporal changes

Another application of the CLS feature description is the assessment of bilateral asymmetry and the search for temporal change in different mammograms of the same breast. Note that from the computer vision standpoint, these applications are closely related since they both involve the comparison of one image against another. This is of course in some ways equally related to the differential-compression mammography of the previous section, however the interpretation of differences between the images is fundamentally different.

Since no two breasts are alike, the automated assessment of bilateral asymmetry is a very difficult task due to the vague definition of what constitutes asymmetry. Section 3.6 reviewed some of the previous attempts to assess asymmetry automatically, noting that in general the task is nearly intractable from a simple low-level pixel-based description of a mammogram's constitution. Those attempts that have pursued this option have always given poor results. Clearly an assessment of asymmetry must be built upon a high-level feature based description of those features contained in a mammogram. It is argued that the CLS feature description, along with the mammographic density description developed in the next chapter, can be used as a domain for the assessment of bilateral asymmetry. This point is taken up again in section 7.2.

A simpler, though by no means simple, task is the assessment of temporal change within a given breast. This is an easier task to accomplish since each breast will generally display many of the same characteristics and features from image to image, particularly if the mammograms

have been acquired at the one clinic and under the direction of the same clinician. Clearly an attempt to identify temporal changes between the images must involve an initial stage of image registration, where the features of one image are aligned with those of the other. Typically the images will not exactly align and some means of image warping and/or image correspondence must be evaluated to achieve an alignment that can then allow the comparison of the image features. The concept of image warping is a research field unto its own with a great deal of activity currently under way. The details of the techniques are beyond the scope (though not the technical level) of this thesis (for further information the reader is referred to [81, 180, 206, 235]) and consequently we simply note that it is possible to achieve a reasonably good match between an image pair that is suitable for the comparison of high-level features. We demonstrate such a warping scheme in section 7.2. Note however, that these techniques are *not* capable of aligning images to an accuracy that allows direct pixel-based comparisons to be made (a fact overlooked by some researchers in the development of automated bilateral asymmetry techniques [73, 131, 231, 232, 235]). Consequently an assessment of local image alignment can only be achieved by refining the results of a general image warping technique with the evaluation of the correspondence between the local features of the images. In this context, the CLS-feature description forms a dense feature space that may be useful for determining the correspondence between regions of the images, and ultimately the determination of temporal change. This technique has not been pursued within the present work.

5.5 Other Applications

The CLS feature detector is formulated in such general terms that the task of finding *nominally thin and straight features with higher intensity that the local background* is common in many applications and thus the CLS algorithm can be applied to these tasks without alteration.

5.5.1 Synthetic Aperture Radar (SAR) images

One advantage of applying the CLS detector to other classes of images is that the features of the algorithm may be exposed in an alternative light, offering the reader greater insight into its operation. Such a situation exists when considering the analysis of Synthetic Aperture Radar (SAR) images, an application that we will consider in this section.

Figure 5.9 displays two SAR aerial images of rural/urban scenes. Observe that the images incorporate many features with attributes of locally linearity, nominal 1-dimensionality and increased image brightness compared their neighbouring regions. Such regions can be identified as the roads, streets of the urban areas, roads along paddock boundaries and also the furrow lines in the ploughed paddocks. Clearly these features are therefore, in the eyes of the CLS detector, nothing more than CLS features, and consequently we would expect the CLS detector to respond to them. For the purposes of this discussion then, we will refer to these features as the *CLS-like* features.

Figure 5.9: Original Synthetic Aperture Radar (SAR) aerial images of rural/urban scenes.

The use of SAR images was considered appropriate since due to the increased contrast of the CLS-like features relative to their cousins in mammogram images, this analysis not only identifies a new application of the CLS detector, but also allows the operation of the algorithms developed in the present work to be observed with greater clarity. Additionally the operation of the 'self-tuning' nature of section 4.4.9 can be observed as the algorithm compensates for the increased contrast of the CLS-like features.

Figure 5.10 illustrates the CLS-like features identified by the CLS detector as formulated in chapter 4.

Clearly the majority of the CLS-like features have been identified, although some of the less salient features have not. The 'failure' to identify some of the less salient features, despite the fact that locally they appear as salient as some of the CLS features identified in the images of section 4.8, is a reflection of the self-tuning property of the detector. The major roads of these images have higher contrast than the most salient CLS features of the mammograms, and since the measurement of image surface curvature will be high at the edges of these high contrast features, the detector naturally biases the detection of the CLS-like features a little in favour of the features of higher contrast. In defence of this self-tuning property however, note also the very large dynamic range of responses that *were* detected, from the most salient roads, to many of the faintest paddock furrows. In the context of mammography we have found this range to be a good balance between sensitivity of the algorithm and the information content of an image.

These images also display a good example of the CLS-detector's inability to capture the full width of a CLS-like feature due to the use of a single second-difference kernel of limited spatial extent. Observe the road intersection towards the lower right corner of the top image of figure 5.9 and the CLS pixels identified in the corresponding region of figure 5.10. Observe that a portion of the centre of the intersection has not been extracted by the CLS-detector. Although its existence is simply noted at this stage, we will return to this point shortly when considering the CLS-removal algorithm of section 5.1.

Figure 5.11 displays the CLS-like-removed images for the originals of figure 5.9. Although these CLS-like-removed images might find a host of applications in their own right, consider for a moment their ability to shed further light on the process of removing CLS features from mammogram images.

Firstly, observe that indeed the CLS-like features have been removed. Also note that the CLS-like features due to edges in the images rather than real CLS-like features have not been removed. This is clear from the retention of the sharp edges surrounding the lighter shaded paddocks in the lower left region of the bottom image. An extremely clear example of this is revealed at our previous example of the road intersection in the lower corner of the top image. Note that an isolated island of the intersection has been retained in the CLS-like-removed image, but more importantly, note that the size of this region is much larger than the corresponding hole in the identified CLS-like pixels of figure 5.10. This is a result of

Figure 5.10: The identified CLS-like features.

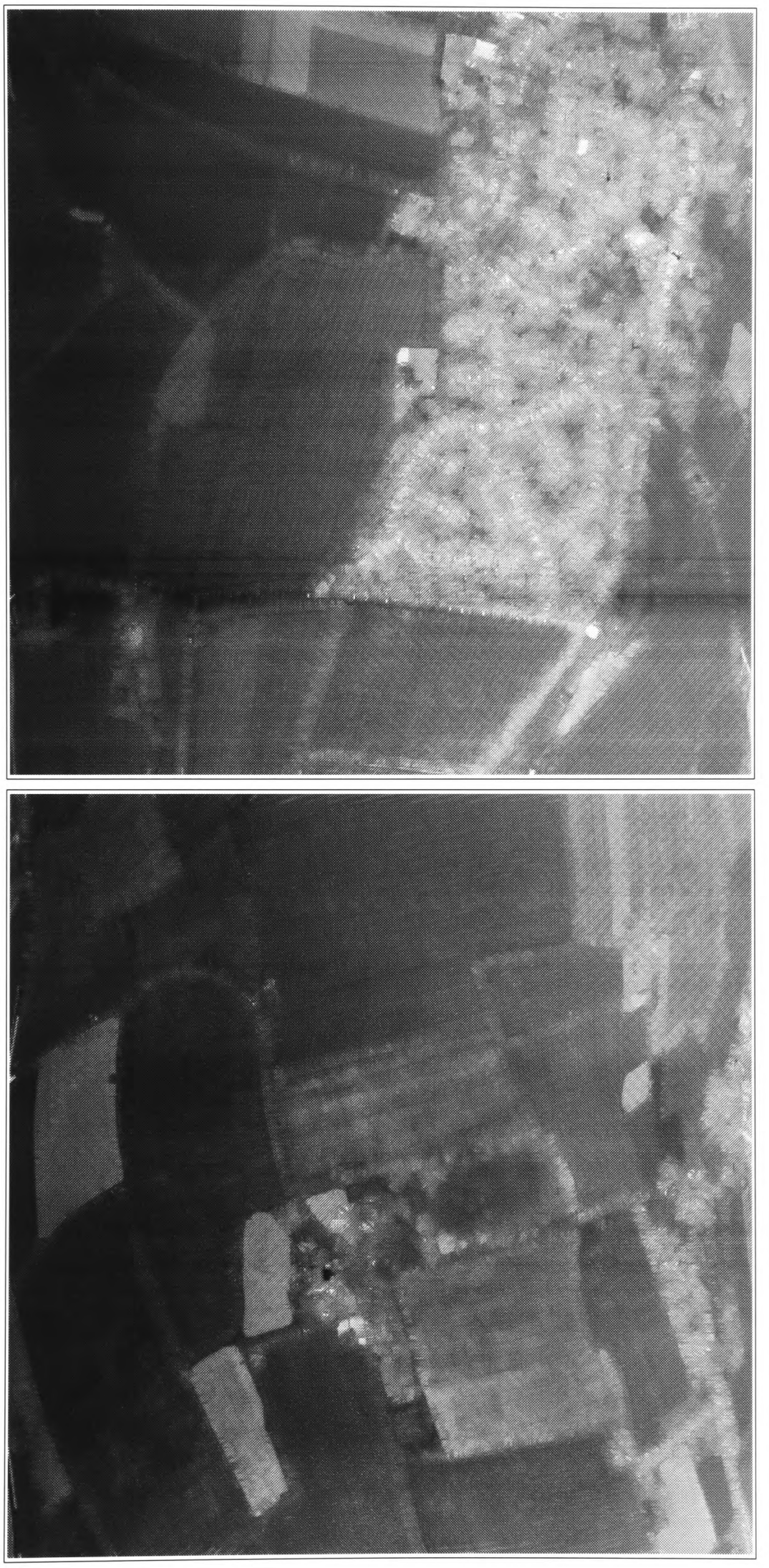


Figure 5.11: CLS-like-removed images.

equation (5.1) identifying that the CLS-like features of figure 5.10 surrounding the intersection are in fact at the edge of a larger feature and not the components of a CLS-like feature that successfully describes the image surface. Consequently the CLS-removal procedure has left these regions of the image intact, preserving the edge surrounding the region.

Within the context of analysing mammograms, it is clear that this procedure unfortunately has the potential to introduce artifacts into the CLS-removed image. However, we need not concern ourselves of this dilemma, as this potential is not realised in practice. The reason is quite simple, and centres around two facts; the most salient objects in a mammogram image are not the CLS features, but rather the larger glandular tissue structures, the pectoral muscle and the mammographic densities; and introduced structures such as the subject of this example, only occur at regions fully enclosed by identified CLS features, perhaps at the wider sections of wide CLS features (compared to the filter kernel spatial extent), or at the intersections of CLS features, as in this example. Thus where such an introduced feature is created, its relative saliency is small. In practice we do find these features in the CLS-removed images, but due to the low frequency of CLS features exceeding the capacity of the filter-kernel scale, and the low intensity difference of the region against its background, they do not prove to be problematic. Obviously an improvement in the detection of CLS features as previously mentioned would reduce considerably the occurrence of this situation.

Although the discussion above has centred on extracting a greater insight into the algorithms of the present work for the purpose of analysing mammograms, there are clearly alternative applications. The astute (or perhaps cynical) reader may have noted the potential use of the CLS feature description to extract the road network from these images for instance. This might be useful for any number of purposes including cartography, the automatic segmentation of rural regions from urban ones, or for the cynical reader, a host of military applications. Although the CLS-like features identified in figure 5.10 do include the road network, it is clearly cluttered by many other insignificant responses, such as the furrow lines in the paddocks. Without making any attempts to optimise the CLS-detector algorithm for the task of extracting a description of the road network, observe that of the CLS-like features in these images, the roads are of the highest contrast, and we would therefore expect them to also display the highest image surface curvatures. Thus in order to selectively identify these features, we could increase the value of the threshold T (see section 4.4.9). In no more than simply a statement of results, figure 5.12 illustrates the CLS-like features identified for the threshold set to $T = -1.5\sigma$ (as opposed to the default of $T = -0.7\sigma$ used for the mammography). For completeness, figure 5.13 illustrates the corresponding CLS-like-removed image.

5.5.2 General ridge and ravine finding

From the examples of the SAR images above, it is clear that the general formulation of the CLS detector means that it is not solely constrained to identify the CLS features from mammogram images, but can obviously be used as a general ridge-finder. The relative applicability of this

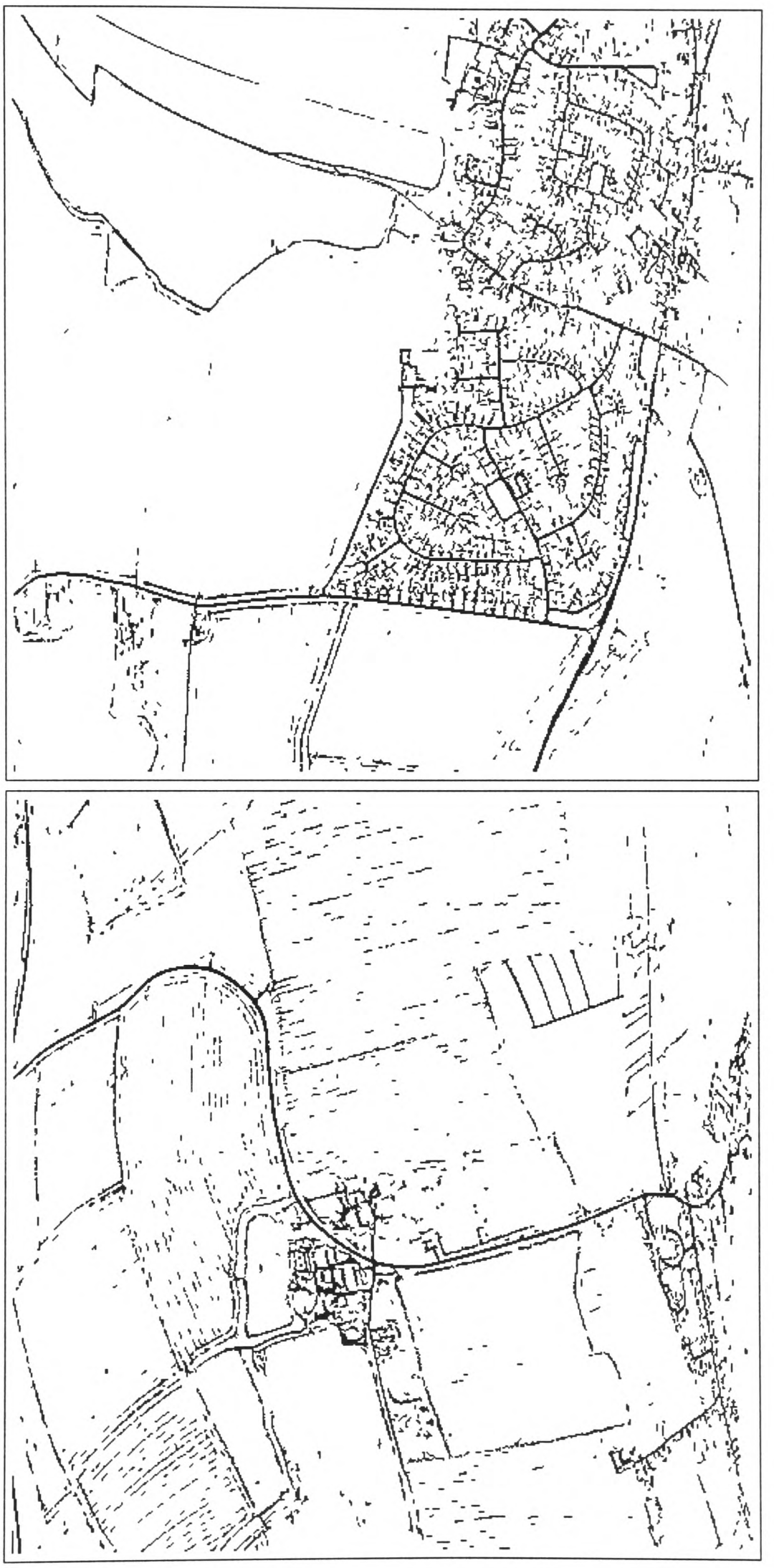


Figure 5.12: CLS-like features identified if the threshold is increased to $T = -1.5\sigma$.

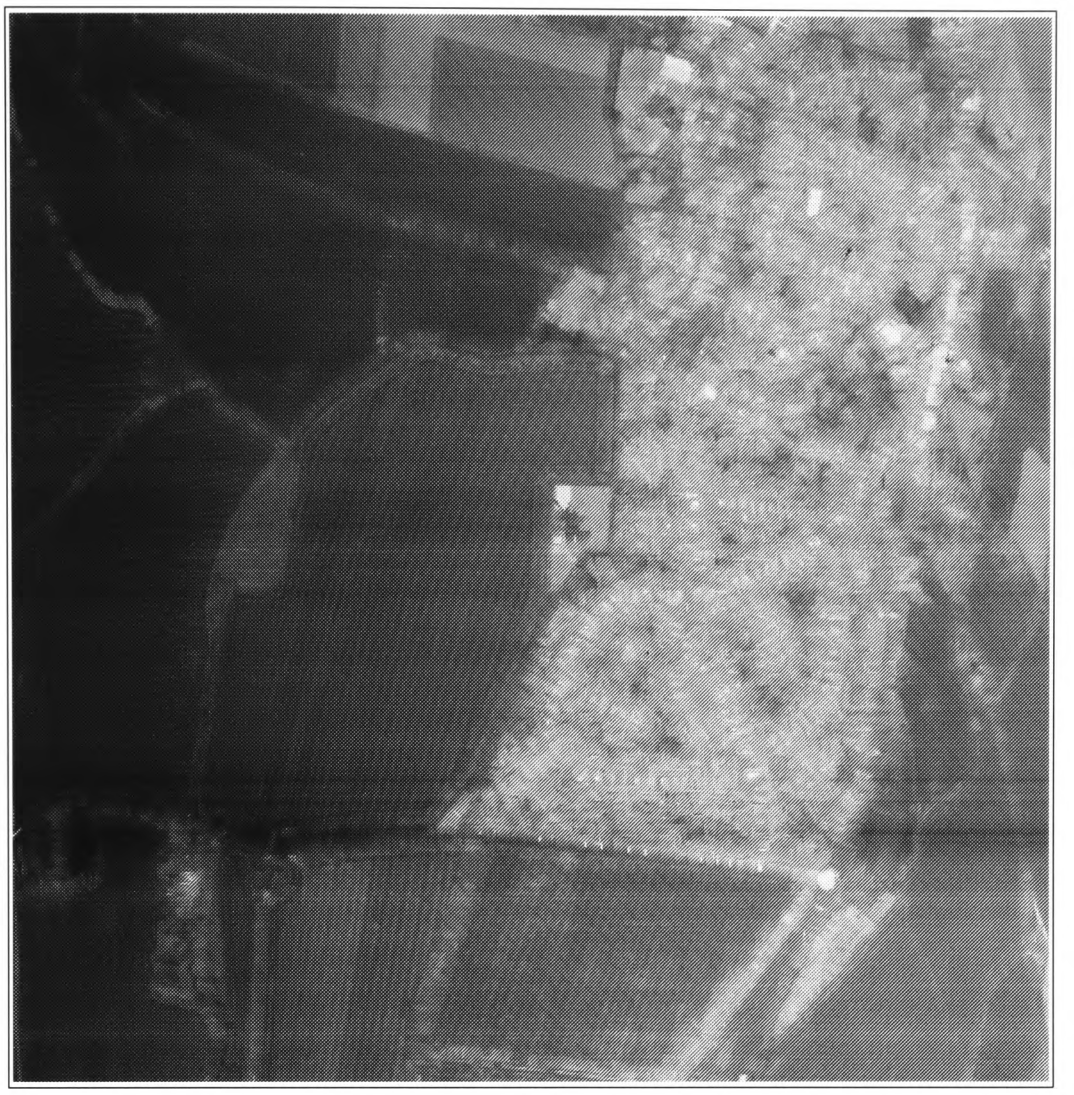


Figure 5.13: CLS-like-removed image for $T = -1.5\sigma$.

technique can only be assessed from the specifics of the task at hand, and thus the range of applications that results of the likes of sections 4.8 and 5.5.1 can be applied to is left to the imagination of the reader.

Also note that the CLS detector is not constrained to the identification of ridges only. By simply inverting either equations (4.16) and (4.17), or the signs of the filter kernel elements of figure 4.23 the CLS detector would respond to ravines, or features with *lower* image intensity than their neighbouring regions. Obviously there are many applications of this, and these are again left to the imagination of the reader.

5.6 Summary

This chapter has developed a technique for 'removing' the known CLS-features from an image. The key advantage of this approach is that the high frequency textural clutter associated with the otherwise insignificant CLS features can be removed from an image whilst those associated with more interesting processes, such as mammographic densities, can be retained. We have demonstrated the effectiveness of this approach for improving the detection of mammographic densities employing a technique to be developed in the next chapter.

The automatic detection of radial scar has received almost no attention in the literature, a reflection of the difficulty of identifying these features. As a mammographic indicator of disease, any attempt to fully automate the analysis of mammograms must be capable, in the first instance, of identifying these features. We have proposed a technique for automatically identifying radial scar utilising both the information contained in the CLS feature description and a technique reported in the literature for identifying the spicules of spiculated masses. Whilst this approach has not been implemented, and thus its effectiveness not assessed, it does indicate a possible solution to the problem of automatically identifying instances of radial scar in mammograms.

Similarly, it was postulated that the CLS feature description might provide suitable information to automate the assessment of differential compression mammography.

The model of the CLS features developed in chapter 4 is based on the (weak) constraints of negative surface curvature, 1-dimensionality and local linearity. As a demonstration of the versatility of this model, and thus the CLS detector algorithm, the algorithms were applied to SAR images of rural and urban scenes. The ability to detect the CLS-like features was demonstrated, illustrating an alternative application of these algorithms. This example also allowed a greater appreciation of the CLS detection and removal algorithms within the context of mammography.

6

Finding mammographic densities

6.1 Introduction

In chapter 2 it was noted that breast masses are normally disease processes and therefore their mammographic representations are highly indicative of breast disease. Consequently many researchers attempting to automate the analysis of mammograms have implemented strategies based upon the search for the mammographic signs of masses (see section 3.2). To complicate the issue however (and hence the 'highly indicative' of the opening sentence of this paragraph), it is possible within the framework of the projective nature of the mammographic imaging process for overlapping breast structures to resemble breast masses when imaged mammographically. This is due to the increased radiological density of the overlapping structures which further attenuates the X-ray beam, as it does in the case of a true mass. Often it is impossible for a radiologist to classify a mammographic density as a mass or as an unfortunate overlapping of glandular structure from the evidence available in a single view mammogram, and in such instances the radiologist would normally request a mammogram from an alternative view to assess the situation more completely. Clearly this technique is not available to automated assessment tools and therefore it is common practice to report *all* suspicious densities. It is this approach that is adopted in the present work.

Ultimately, we seek to extract from a mammogram image a description of not only all the salient mammographic densities, but furthermore a characterisation of malignancy for each density. This task can naturally be partitioned into two steps: initially locate all mammographic densities in an image and then attempt to further segment the densities into classes of benign, malignant and overlapping glandular structures. It could be argued that the differentiation of true masses from overlapping structures could be achieved by automated systems if they

were presented with four-view mammograms[1], however in the United Kingdom the standard screening procedure (medial-lateral oblique exposures—see figure 2.7) provides only a single view of each breast, and therefore such discrimination is impossible in this arrangement.

A consequence of the inability to distinguish overlapping structures from true densities is that it is unclear how an automated technique might accomplish the final stage of classifying densities according to their tissue types. For this reason it is the detection rather than the classification of mammographic densities that is considered in the present work. This constraint is hardly to be considered a limitation since a feature description of the mammographic densities in an image would be a very useful starting point for the subsequent classification into benign, malignant and overlapping structures. An additional advantage of a description of the mammographic densities is that it allows an assessment of the bilateral asymmetrical tissue structures. Recall from section 2.5 that "asymmetry is the radiologist's greatest aid in determining abnormalities both benign and malignant" (Andolina et al. [6]). Unlike the techniques of Giger et al. [73, 231, 232] (see section 3.2), and Lau and Bischof [131] (see section 3.6) who compare images bilaterally by simply subtracting the image intensities of one image from those of the bilateral image (after image warping and registration), a feature description of the mammographic densities allows a higher-level comparison of tissue structure itself. Given the variations in imaging and digitisation conditions (see section 2.4), there is nothing to guarantee that direct pixel intensity comparison can capture image structures, particularly if the technique was unfortunate enough to be presented with an image pair containing similar densities in the same location. Clearly the use of the density feature description allows a systematic inclusion of any a priori information describing asymmetry into a higher-level decision making process. For these reasons we propose to extract a feature description of the mammographic densities in an image.

In section 2.5 it was noted that the increased radiological density of breast masses results in mammographic densities being characterised by higher image intensity than their surrounding regions, for example consider the cyst in the upper right portion of figure 2.7. It is on this basis therefore that we seek to identify mammographic densities by searching for local regions of comparatively higher surface intensity than the surrounding tissues. As will be shown in the following section, the problem of finding locally salient regions of higher intensity can be partitioned into two distinct operations: determine any locally salient region, or *blob*, that could possibly be a mammographic density and then assign these candidates to the classes of 'mammographic density' or 'other', where the composition of 'other' is dependent upon what features the first stage detects.

Section 3.2, which reviews the techniques that have been reported in the literature for the detection of masses, shows that the approach adopted here of searching for regions of high intensity is clearly not novel [84, 131, 222, 228]. As noted in that section, a number of researchers have introduced measures to overcome the difficulties associated with the CLS

[1]This is a common screening format for some regions of the United States, though by no means all. Critics argue that the extra radiation dose involved in the double exposures for each breast are not warranted at the screening stage.

features [73, 231, 131] since their intensity (at the borders of spiculated masses for instance) can be comparable to the intensities of the mass itself, and this makes simple intensity based techniques ineffective. Furthermore it is also common for such techniques to employ a 'region-growing' stage to ensure that the extents of the density have been captured [228], however such techniques can erroneously include the CLS features alongside the boundary pixels of the densities. Overcoming this last difficulty by early termination of the expansion of the region is clearly unacceptable. In fact, Claridge and Richter [40] show that the image curvature zero-crossings, which are commonly accepted as the locations of boundaries in many computer vision applications, are not suitable for determination of density boundaries as they under-estimate the extent of a boundary, and this would only compound the errors in localising the extent of a region.

Recall from the previous chapter the technique developed for removing the CLS features from an image, giving an image that preserves the detail of both the regional image intensities and the local image intensities in regions without CLS features (see figure 5.4). Within the context of the present application, this representation of an image retains the image gradients across the boundaries of densities whilst removing the degrading effects of the CLS features. Thus, to resolve the difficulties experienced by previous authors, we propose searching for mammographic densities within the CLS-removed image, developing a feature description of the densities in such an image. This feature description of the densities can then be used in conjunction with the original image for a variety of applications.

As a consequence of removing the CLS features from the original image, the task of detecting the mammographic densities is considerably easier since the densities now appear as *localised* regions of relatively high intensity without the textural clutter and boundary distorting effects of the CLS features. This observation allows the development of a density detector based upon the expected features displayed by mammographic densities.

The detection of localised regions of relatively high (or conversely, low) image intensity, herein termed *blobs*, has been a topic of some concern to the computer vision community given the broad potential of such a generic feature description to describe image features. Initially this chapter reviews some of the most recent published work for detecting salient blobs in images, partitioning the problem into the detection of all possible blobs, and then the subsequent classification of the blobs. The following sections develop techniques to solve these problems.

As an aid to description of the blob detection problem and its solution, it is useful to consider an image as a surface elevation in intensity, rather like the terrain of the earth's surface. Regions of high intensity are like mountain peaks, regions between the high intensity patches (mountains) form valleys and regions of low intensity form depressions or wells in the image surface.

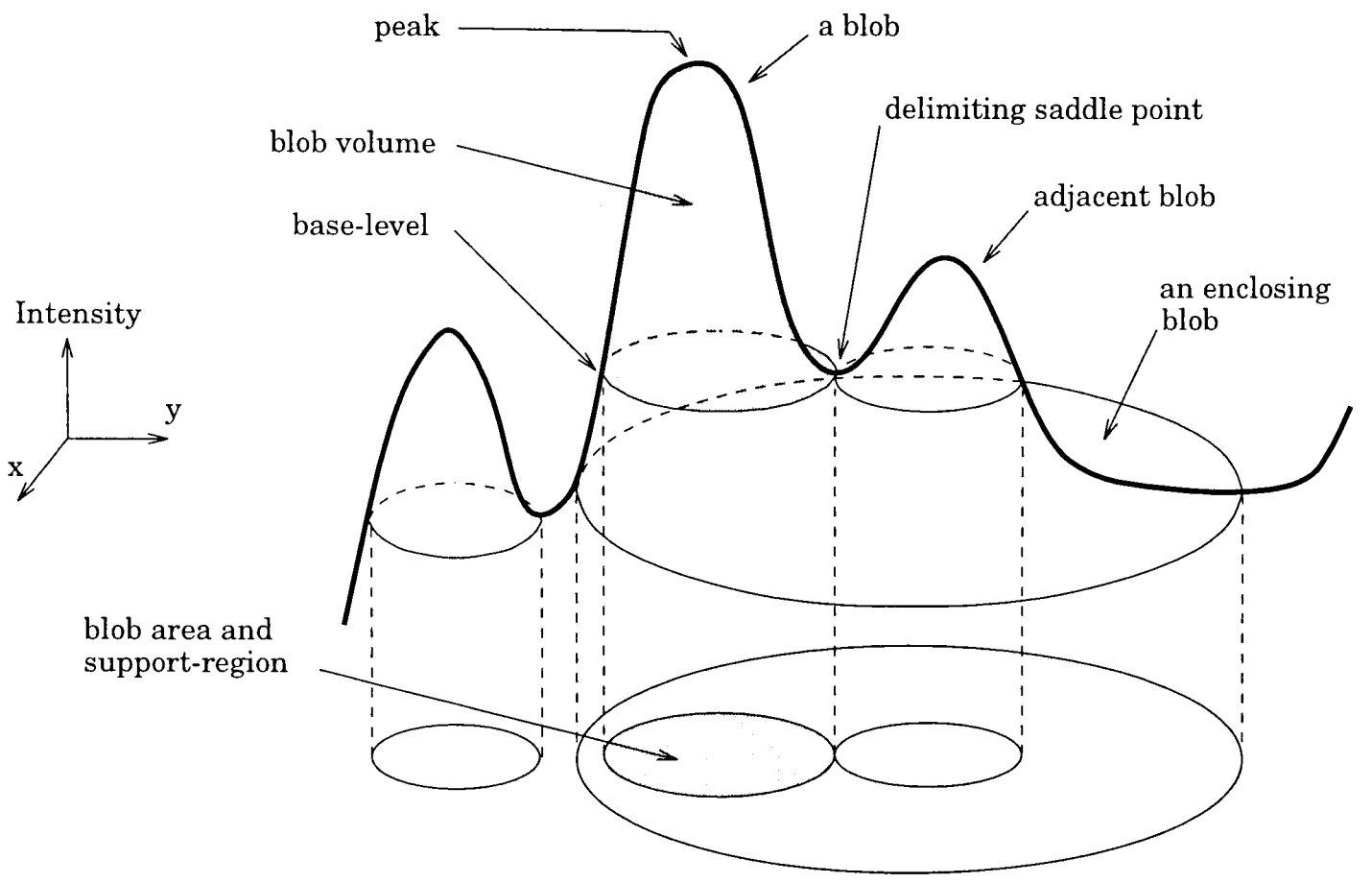


Figure 6.1: The model of a blob and its associated nomenclature.

6.2 Identifying blobs in images

The problem of locating salient blobs in images is a fairly general one that has been approached from many directions in the pattern-recognition/image-processing communities. The precise formulation of each approach is intimately related to the definition of a 'blob', and in particular the extent of the blob. For instance Ehrich and Lai [59] define the extent of a blob as reaching to the base of the valley in all directions surrounding the peak, that is to the lowest point (intensity) separating one peak from a surrounding one (where the river would flow in the valley in the analogy). This clearly leads to nonintuitive results for, say, a small peak on the side of a larger slope. On the other hand Lindeberg [137] defines the extent of a blob as that region enclosed when allowing a blob to expand from a peak until it would merge with an adjacent blob. This situation is depicted in figure 6.1. The grey-level intensity at which the blob would first begin to merge with an adjacent blob is termed the *base-level* of the blob, and the image surface saddle point marking the initial merging of the adjacent peaks is termed the *delimiting saddle point*. The *support region* of the blob is defined to consist of those pixels that have a grey-level exceeding the base-level and can be reached from the peak without descending below the base-level of the blob. Thus the support region quantifies the blob area. The blob *volume* is defined to be the summation of the differences between the image surface and the base-level for the pixels within the blob support area.

The simplest way to consider an image segmentation based upon this definition of a blob is to immerse the image surface in a fluid. As the fluid drains away (or the image surface is lifted)

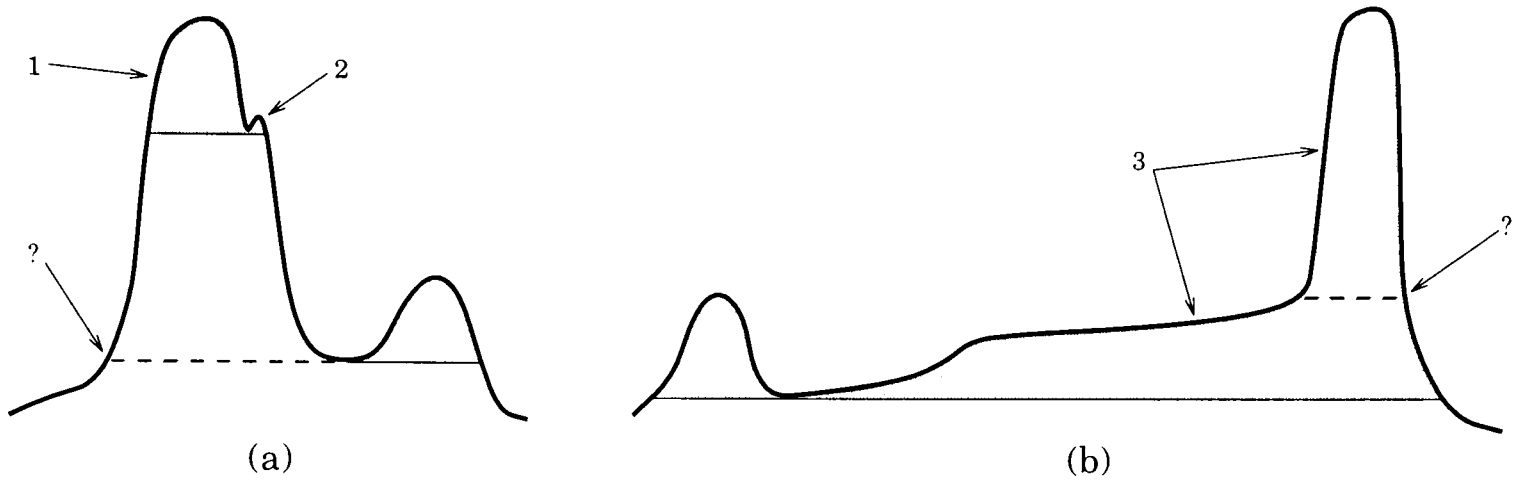


Figure 6.2: Some situations posing difficulties for segmenting the most salient local blob.

the image surface breaks the fluid surface at the local image peaks. The current shore-line of all the blobs breaking the fluid surface (imagine the Hawaiian Islands emerging from the Pacific Ocean) defines the *level-curves* of the image surface. The fluid is allowed to drain away, during which each blob expands its support region, until each blob would first merge with another, rather like Mt. St. Michel on the western coast of France joins with the mainland as the tide runs out. The fluid height just as the blobs merge gives the base-level of the blob thus defining the blob support-area. Allowing this process to continue by trapping each blob merging as the fluid continues to drain away eventually segments all the blobs from an image.

Despite the apparent intuitive results of Lindeberg's definition, there are however a number of limitations which compromise its effectiveness for the current application. Figure 6.2 illustrates some cases in which the segmentation does not accord well with the desired response. In figure 6.2(a) the existence of a small surface intensity increase (peak '2' in the figure), even as small as a single pixel with an intensity level only one grey-level greater than the current base-level, will terminate the expansion of the main blob (peak '1'). Clearly situations exist where the most salient blob for a given region is not at peak '1' or '2' (see figure 6.2(a)), but rather as defined by the dashed base-line marked with a '?'. Unfortunately an algorithm using the simple definition of a blob as above will only segment the image into blobs '1' and '2', and never find the '?' blob.

A potential solution to this difficulty is to initialise an expansion process at each and every pixel in an image, thus ensuring that blobs such as that identified by the dashed base-line of figure 6.2(a) are identified. This approach is however rejected for two reasons: (a) many blobs (such as blob '2') are simply insignificant in relation to the other, *more* salient blobs of which they are a component (the dashed base-line), and thus their extraction simply increases the complexity of the subsequent analysis, and (b) the excessive computational expense involved in initiating a process at *every* pixel.

Lindeberg solves this difficulty by approaching the entire segmentation task from a scale-space perspective. By convolving the original image with a Gaussian kernel at a number of different scales, a set of scale dependent representations of the image are obtained. By

successively smoothing out the little peaks such as '2' in figure 6.2(a), this allows features such as the larger blob of which peaks '1' and '2' are components to be successfully segmented.

Guissin and Brady [80] segment each image into a family of surfaces by thresholding the image at *every* intensity. The set of all so-called 'iso-intensity curves' which form the bounding contours of each segmented blob at each intensity, will naturally contain all the blobs in an image. To reduce the enormous set of level-curves to a manageable set of significant contours, they develop a saliency measure to assess each contour, retaining only those contours that exceed some adaptive threshold. Whilst this technique might be acceptable for relatively small images containing large features (and thus few net contours), the computation cost involved in finding small blobs in large mammogram images is prohibitively expensive. Accordingly this technique in its raw form is not suitable for the current application.

Before adopting the method of Lindeberg for the current situation, note that there is however a situation that no amount of scale-space effort can accommodate. Consider the surface profile displayed in figure 6.2(b). Furthermore consider that the surface is smooth in the sense that there are no small peaks such as peak '2' in figure 6.2(a). From an image intensity perspective, the most salient blob in the image, having good contrast and sharp edges is that shown with the dashed base-level and marked as '?'. However as defined above, the segmentation procedure cannot extract this blob as it will extract the entire shaded region of blob '3' instead. Although in theory this is quite a problem, in practice it seldom presents as such. This is due to the fact that a surface with the smoothness characteristics required to present this scenario is very rare. For once, and in some perverted sense, image noise can be considered an ally! The caveat to this is, of course, that whilst noise may force a segmentation at some base-level, there is no guarantee that it will do it at a base-level giving the most salient blob. It should also be noted that if a scale-space approach is used to overcome the difficulty posed by small peaks as described above, then the Gaussian smoothing operation will remove the noise peaks that might help the segmentation in this situation.

It is clear that the scale-space techniques of Lindeberg are not compatible with the two difficult cases presented above, yet the analysis of all the iso-intensity contours is too expensive. As a compromise between these extremes, consider the retention of Lindeberg's general idea of identifying a blob by expanding its support area until some terminating criterion is met. Suppose an expansion process were initiated at the peak of blob '1' in figure 6.2(a), terminating at the delimiting saddle point shared with blob '2'. Suppose the expansion process were then allowed to continue, eventually terminating with the blob identified by the dashed base-line. Thus for the given local region of this blob we would have two candidate segmentations from which to choose the *most* salient. It is at this point that a saliency metric, such as that of Guissin and Brady, would give a measure of each blob's saliency, allowing the selection and retention of the most salient blob, and the rejection of the other. It is this process that is adopted in the present work.

Before presenting the blob-detector developed in the present work, it will prove useful to

have reviewed the basis for the Guissin and Brady algorithm, and we present this in the first instance. They proceed by developing an edge detection scheme based upon the iso-intensity contours of an image surface. In the current application, the closed nature of the iso-intensity contours gives topologically closed edge contours that effectively segment an image as the edges can be used to describe the bounding contours of various regions. In the general formulation of their technique, they argue that although the task of spatial image analysis (in this case blob segmentation) is necessarily application dependent, and thus higher level in the sense it requires the application of a priori model-based knowledge, the first stage of any such process can be considered to be application independent. They postulate that a structured image representation in the form of the iso-intensity contours provides the necessary edge information for extracting the dominant regions of an image. This approach will extract contours at *every* intensity level in an image, naturally posing the question of which contours are significant, and how should that significance be quantified? Guissin and Brady resolve this issue by developing a single contour significance metric that quantifies the saliency of a blob relative to its local neighbourhood. Given such a measure of blob saliency it is possible to compare the relative saliency of associated blobs, and to reject some blobs if other *more* salient blobs adequately describe that region of the image. They note that the single most important feature identifying a blob within a region is the characteristics of the image surface at the blob's bounding contour, in particular the edge strength. They therefore base the definition of their saliency metric $\mathcal{M}$ on three aspects of the contour S: its length $L = \sum_S 1$, contrast μ, and stability σ, as:

$$\mathcal{M} = \frac{\mu}{\sigma} \sum_S |\nabla I| = \frac{\mu^2}{\sigma} L \tag{6.1}$$

where the contrast μ is a measure of the average image gradient along the contour, whilst the stability σ measures the standard deviation of the image gradient along the contour, as:

$$\begin{aligned} \mu &= \frac{\sum_S |\nabla I|}{\sum_S 1} = \frac{1}{L} \sum_S |\nabla I| \\ \sigma &= \left[\frac{\sum_S \left(|\nabla I| - \mu \right)^2}{\sum_S 1} \right]^{\frac{1}{2}}. \end{aligned}$$

Whilst this metric is suitable for their application, note that the magnitude of the measure is linearly related to the contour length, favouring larger blobs with longer perimeters. This clearly is at odds with our a priori knowledge that mammographic masses can be significant at any size, and although it is important to detect larger masses, it is the smaller masses that allow the *early* detection of disease processes within a breast, and thus there should not be an emphasis away from these features when searching for blobs in an image. Furthermore it is conceivable that a large bounding contour enclosing the entire parenchymal tissue of a mammogram image, with relatively low, but stable, contrast along the boundary between the fatty tissue background and the parenchymal tissues (in terms of equation (6.1), large L, intermediate μ and small σ) could have a significance measure equalling or exceeding a small

blob with high, but less consistent contrast (small L, large μ and large σ). Clearly this result is not desirable within the context of the present task. Whilst acceptable within the problem definition of Guissin and Brady, the implicit assertion within their significance measure that larger blobs (longer contours) with large and constant image gradients along the contour are more significant or 'important' is not appropriate to the present task, and we will therefore aim to remove such interpretative attributes from the low-level analysis of the significance measure.

It is within this framework that we present the blob detector developed in the present work as a mixture of the ideas of Lindeberg, and Guissin and Brady. Initially an image is scanned in search of the salient blobs. Rather than locate *every* blob (such as the segmentation provided by every iso-intensity contour), we seek to locate only the most salient blob from each location which adequately describes the given region of the image. By developing a saliency metric, comparative analysis of the saliency of the blobs in a given region is used to select the most appropriate blob for that region. A consequence of adopting this approach, is that the procedure returns many contours that are not mammographically significant, and thus the higher-level segmentation (into mammographically significant and insignificant blobs) must be more sophisticated. However the advantage of this approach is that at the higher-level second stage where a priori information (encoding the details of the current segmentation task) is introduced into the process, one can be sure that no segmentation based on inappropriate rules at a previous stage has omitted a blob from those presented to the higher-level stage for final consideration. Note that the second, higher-level stage is necessary as the saliency measure identifying the most salient blob in a local region may or may not (and as it turns out, it alone does not) provide sufficient information to segment the blobs into mammographically significant and insignificant classes.

In the next section we develop the low-level process that proposes all the locally salient blobs which are candidates for a higher-level consideration as real mammographic densities, a discrimination task to be considered in section 6.4.

6.3 Proposing potential densities

Consider the definition of a blob given by Lindeberg [137] as described in the previous section and illustrated in figure 6.1. The task at hand is to locate all such blobs in an image within the constraints imposed by the situations illustrated in figure 6.2 and discussed in the text above. As noted by Guissin and Brady the single most accessible feature defining the relative saliency of a given blob within its local area is a measure of the characteristics of the blob boundary, in particular the image surface gradient, or steepness of the image surface at the boundary (see figure 6.2(b) for instance). It is therefore proposed that a measure of a blob's saliency will involve a measurement of the image surface gradient at the boundary of the blob.

There are two distinct, though related problems involved here: (i) how to measure the image surface gradient ∇I, and (ii) the definition of a metric which quantifies a blob's boundary

gradient. Clearly the first task here is but a component of the second, for without it, calculation of a blob's boundary-gradient measure, which we denote as $\mathcal{B}$, would not be possible. Naturally these tasks are inter-dependent, and separation of their definition is not generally possible. As there are any number of definitions for $\mathcal{B}$ (with corresponding ∇I) which could be employed, we initially consider a number of simple, intuitively obvious metrics and then propose the definition of $\mathcal{B}$ used in the present work. Given such a metric $\mathcal{B}$ to quantify the boundary surface gradient, it is a relatively straightforward matter to define and implement a blob detector subject to the constraints mentioned above, and this is developed in the subsequent section.

6.3.1 Blob saliency metrics

An obvious starting point for developing a measure of the blob boundary surface gradient is the average surface gradient taken over the pixels of the blob boundary. Retaining the notation of Guissin and Brady we have:

$$\mathcal{B}_1 = \frac{\sum_S \nabla I(s)}{\sum_S 1} = \frac{1}{L}\sum_S \nabla I(s) \tag{6.2}$$

A common technique for estimating the image surface gradients is to filter the image with a rotational family of first-difference kernels based on the kernel $\begin{bmatrix} -1 & 1 \end{bmatrix}$ or its variants[2] (cf. the second-difference kernels of section 4.4.8). Recall from the opening discussions of this chapter the definition of a blob's base-level and delimiting saddle point (see figure 6.1). By terminating the expansion of a blob at the first saddle point on the border of the blob, one is guaranteed to have a positive surface gradient at all boundary pixels (except the one at the delimiting saddle point). In such a situation evaluation of equation (6.2) proceeds immediately.

Whilst this metric, which is simply the average edge gradient of the blob, is plausible enough, observe that in noisy images with noise corrupted image gradient estimates, this measure simply combines the apparent response from each contour pixel without regard for the quality of the data. Recall that we propose to extract the blobs from, and therefore calculate the image gradients in, the CLS-removed image. Recall from section 5.1 that the CLS-removed image is *unchanged* in regions that were not identified as CLS features by the CLS detector of chapter 4. Commonly the boundaries of the blobs lie within these regions, and the calculation of the image gradient at these locations is susceptible to the image noise present within the original image. Therefore compensation for the noise corrupted gradient measurements must be introduced into the calculation of the blob saliency metric $\mathcal{B}$.

Suppose that for each gradient measure $\nabla I(x, y)$ there was a corresponding confidence weighting $w(x, y)$ available, where $0 \leq w \leq 1$. In regions highly corrupted with noise, the estimated image gradient may poorly represent the underlying real image gradient, and in

[2]Often the kernels used in the computer vision field will be based on $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ to overcome the aliasing effects typical of video image capture techniques. As a consequence of the digitisation methods used to acquire the mammogram images these effects are not present, and consequently kernels of the form $\begin{bmatrix} -1 & 1 \end{bmatrix}$ can be used, exploiting the small spatial extent of this form. This has the benefit of localising, and thus improving the accuracy of the estimate of the local image gradient.

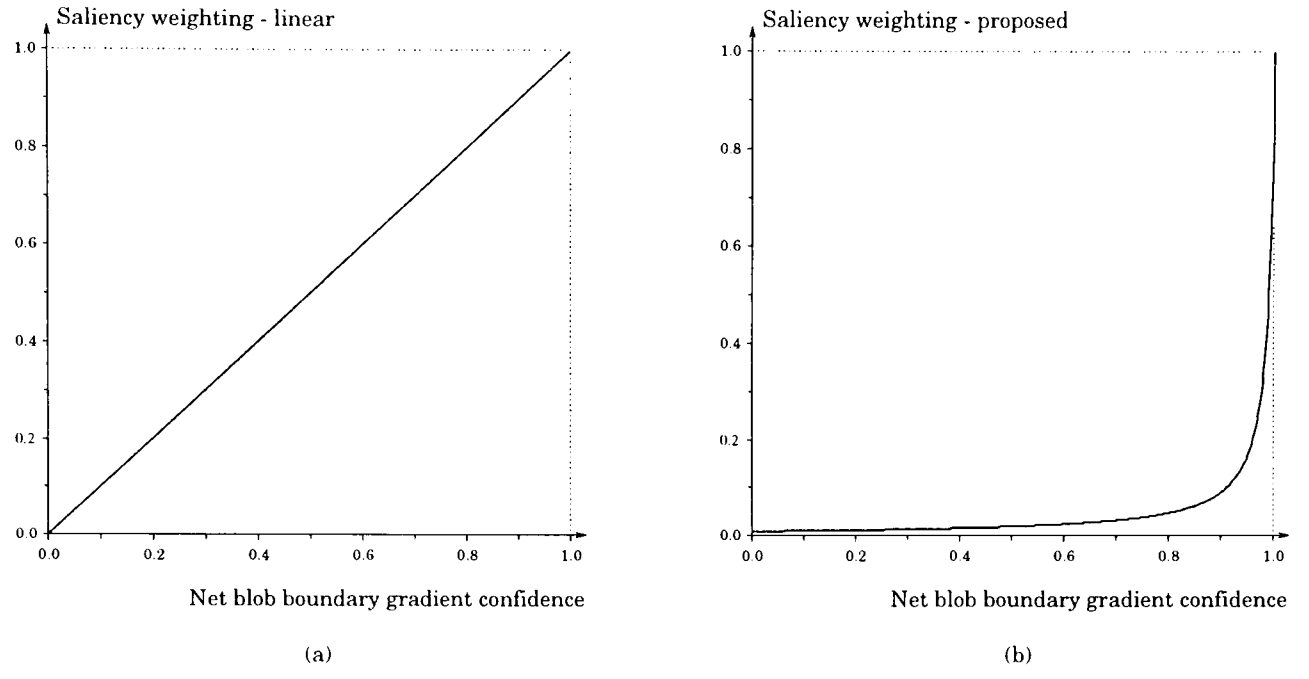


Figure 6.3: Saliency weighting as a function of the net confidence in the blob boundary gradient. A net confidence of 0.75 may be thought of as either 75% confidence of the gradient estimate for *all* pixels on the blob border, or alternatively complete confidence of the gradient estimate at only 75% of blob border pixels with zero confidence at the pixels of the remaining quarter.

these instances it would be desirable if w were small. Conversely, in regions with low noise and good gradient estimates, a value of w close to unity would be desirable. With such confidence measures available, equation (6.2) can be rewritten as the average weighted gradient measure over the blob boundary:

$$\mathcal{B}_2 = \frac{1}{L} \sum_S \nabla I(s)\, w(s).$$

This equation may by further rewritten as:

$$\mathcal{B}_2 = \frac{\sum_S \nabla I(s)\, w(s)}{\sum_S w(s)} \times \frac{\sum_S w(s)}{L}. \tag{6.3}$$

The first term of this equation is simply the weighted average image surface gradient over those pixels contributing to the contour saliency, whilst the second term weights the measure according to the percentage of the contour contributing to the saliency of the contour. That is, if the confidence in the image gradient estimates approaches 100% (i.e., 100% confidence of the gradient measurement for 100% of the contour pixels) then the first term reduces to equation (6.2), whilst the second term approaches unity. On the other hand at 0% confidence the numerator of the first term equates to zero, reducing the overall metric to a zero result[3].

The second term of equation (6.3) weights the blob saliency metric $\mathcal{B}_2$ in a *linear* relationship with the percentage contribution of high quality gradient measures. This weighting function is displayed in figure 6.3(a). Although such a weighting has the desired property of monotonicity from low values at low confidence through to high values at high confidence, there is no obvious reason why this 'middle-of-the-field' weighting should be preferred over any other. To consider this further recall the profile of the blob shown in figure 6.1. The more salient a blob is the

[3]Clearly the sequence of evaluation of the terms of equation (6.3) can effect the null result at 0% confidence and this case should be explicitly tested for in any implementation.

steeper will be the image surface at its bounding contour, and a relatively larger proportion of its contour will have the steep image surface, that is it will have fewer immediate neighbours (such as the 'adjacent blob' forming the delimiting saddle-point of figure 6.1). At such locations of steep image surfaces, noise makes very little difference to the local image gradient, and thus in spite of the noise the confidence weighting should be high since the measurement will still be relatively accurate. Conversely for less salient blobs with shallower image gradients, the noise is predominately more influential and the confidence weighting should reflect this. When these responses are summed over the contour, we therefore expect (and have observed experimentally) that truly salient blobs will have very few contour pixels at which the image gradient cannot be estimated with high confidence. Thus it is proposed to weight the blob saliency metric $\mathcal{B}$ such that contours with lower confidences are more heavily suppressed than the linear relationship of equation (6.3) allows. We have found through experimentation that the following weighting term

$$\left[1 + \sum_{S} \left(1 - w(s)\right)\right]^{-1}$$

which is illustrated in figure 6.3(b), is both easy to calculate, and implements a suitable weighting. Reorganising equation (6.3) by substituting this term yields the following:

$$\mathcal{B}_3 = \frac{\sum_S \nabla I(s)\, w(s)}{\sum_S w(s)} \times \frac{1}{1 + \sum_S (1 - w(s))}. \tag{6.4}$$

The definition of a blob given so far in this chapter, from which the metric $\mathcal{B}_3$ has been developed, does not reveal quite the entire picture. Consider the situation illustrated in figure 6.4 and recall that the expansion of a blob terminates upon reaching a delimiting saddle-point. For some salient blobs of relatively large area, and therefore relatively large blob height[4], it is possible that the image surface gradient at the contour boundary is either not at the maximum (left edge of large blob of figure 6.4) or the image surface gradient is not (relatively) high (right edge of large blob of figure 6.4). Consider a blob with much less support-area, as illustrated by the left blob of figure 6.4. Assuming this blob is of roughly equal saliency as the larger one[5], the value of the blob saliency metric $\mathcal{B}$ as given by equation (6.4) will be biased in its favour due to the relatively strong image gradients along its boundary. The solution to this problem is twofold: (i) calculate the image surface measures over a strip of pixels rather than the pixels of the border itself[6], and (ii) add an additional weight to the metric $\mathcal{B}$ by multiplying by the

[4] A blob of large area and small average height would not be a relatively *salient* blob.

[5] The issue of relative saliency has not been addressed formally since it cannot be considered without a rigorous definition of the relationship between the various parameters contributing to the definition a blob. That is, what is the relative importance of boundary gradient measures, blob height, blob area, and more importantly, at what point in the trade-off does each attribute dominate? We have so far identified that higher image gradients and more consistent image gradients are important, however we have made no assumptions regarding blob area, nor do we intend to do so (see section 6.2). Within this framework then, we establish a loose saliency ranking as follows: irrespective of blob area, blobs with low image gradient and low blob height have low saliency; whilst blobs with high image gradients and high blob height have high saliency. In between these two extremes lies a whole range of blobs that can be considered to be of intermediate saliency, and in this sense the saliency measures of the large and small blobs of figure 6.4 might be approximately equal.

[6] The careful reader would note that the concept of the blob 'border' has not been defined within the discrete domain. That is, does the blob border lie *between* pixels or at some other location given at sub-pixel resolution? The assumption

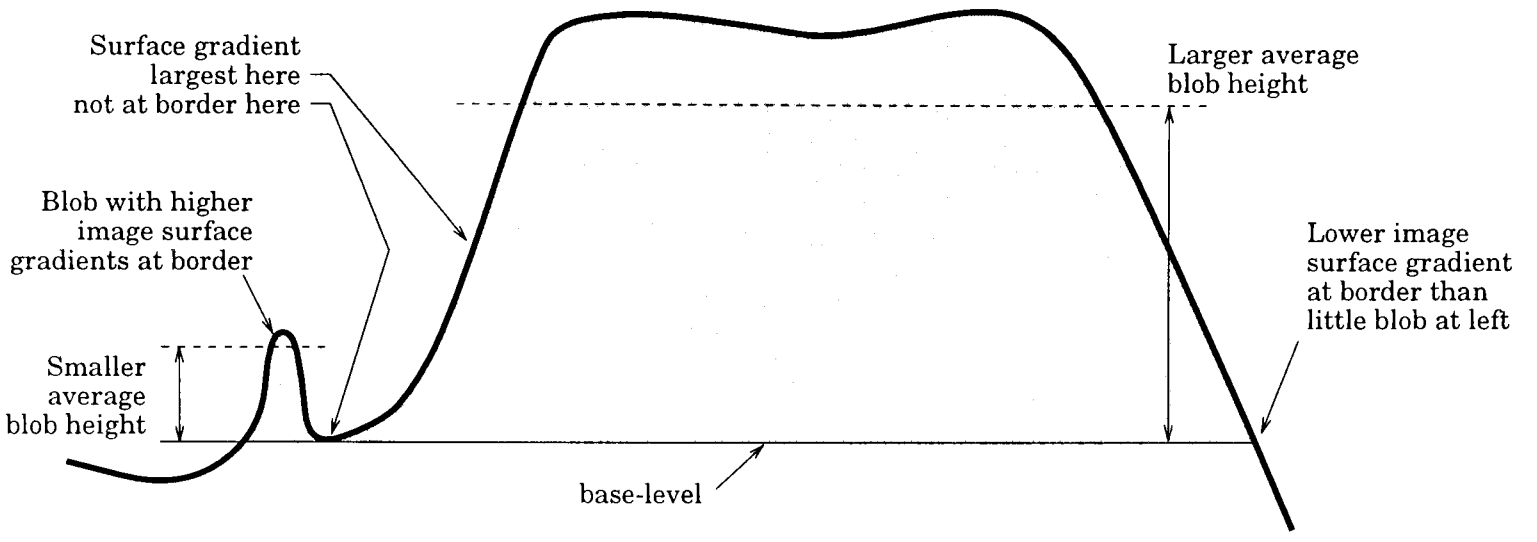


Figure 6.4: Some difficulties calculating the blob contour gradients.

blob height. The first correction is merely an implementation detail and is thus simply noted as such here. Assuming that the blob has area A, the second correction gives the final saliency metric as:

$$\mathcal{B} = \frac{\sum_S \nabla I(s)\, w(s)}{\sum_S w(s)} \times \frac{1}{1 + \sum_S (1 - w(s))} \times \frac{\sum_A I(a) - \text{base-level}}{\sum_A 1}. \tag{6.5}$$

Note that each component of this equation has a unique role identifying some characteristic attribute of a given blob. In combination then, they are able to accommodate each other's weaknesses and the net result is that the metric $\mathcal{B}$ given by equation (6.5) provides a single metric for a given blob's saliency.

In order to implement equation (6.5) we must be able to calculate the image surface gradient ∇I and the associated confidence weights $w(s)$, and it is this problem that we will now address.

6.3.2 Finding the image surface gradient

A standard consideration of any system attempting to measure the surface gradient of an image (e.g., gradient based edge detectors) is that of noise. The general computer vision technique employed to minimise the effects of noise and return a more stable gradient measurement is to smooth an image with a Gaussian filter (under the assumption of course, that the noise is white/Gaussian). During the development of the CLS feature detector of chapter 4 it was argued that this is a thoroughly inappropriate method to suppress noise since the features in that instance would be adversely effected by the smoothing operation. Although the features of interest in this application (the blobs) are generally much larger than the CLS features, and thus will most likely not be smoothed away, the pre-smoothing of the image is still discouraged for a number of reasons. The most pertinent of these reasons is that it is implicitly assumed that the smoothed image is free from noise and that the gradients calculated are accurate. That is, in terms of the previous section, subsequent processing adopts a confidence of 100% for all the gradient measurements. Clearly this is an inappropriate method for accommodating

we make here is that the border does indeed lie between pixels in much the same fashion as used in the CLS description of chapter 4. Accordingly we assume 8-connected blobs and a 4-connected background to preserve topology.

the noise given that the formulation of the blob saliency metric of the previous section requires a confidence associated with each gradient measure calculated. Furthermore, as will be shown in this section, the known characteristics of the CLS-removed image can be used to develop a confidence estimate of the gradient measurements. Gaussian smoothing of the image destroys this a priori information rendering the confidence estimate technique ineffective, and thus this method of noise suppression is to be considered wholly inappropriate in the current situation.

Recall that the image gradients are to be determined from the CLS-removed image, and also that the CLS-detector of chapter 4 identifies all CLS features down to some minimum area (number of 8-connected pixels). A region with a significant second difference response (see chapter 4), yet only a few connected pixels (read 'noise' here) will not be detected by the CLS-detector, and it will still be present in the CLS-removed image. Any regions of intermediate size (read 'CLS feature size' here) will have been detected and their effects removed from the image as illustrated in figure 5.3. Features of much larger size (read 'parenchymal structures/densities etc.' here) will not have been detected and will still be present in the image (e.g., figure 5.5). Thus we can consider small isolated peaks to be indicative of noise processes in the CLS-removed image, and it is this fact that will allow us to extract reliable image surface gradient measures from the image. Equally, since a mammogram image is a representation of the breast's net capacity for attenuation/absorption of the X-ray at each location, and that breast tissues are not randomly *absent* from otherwise uniform tissue sections, isolated wells, or dips, in an image surface are most likely to be the result of a noise processes.

On this basis then, consider the fact that we are searching for blobs that *always* exceed some minimum spatial extent, say a few pixels only if taken to the extreme. Thus, if a boundary truly exists at a given location, it is expected that the neighbourhood surrounding a given pixel will support that notion, unlike the neighbourhood surrounding a single-pixel peak for instance. By analysing all the surface-gradient information available at a given location, we can deduce the quality, or our confidence in, the image gradient calculated at that location as follows.

Since the gradient of a vector field is separable, that is:

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \dots$$

the standard technique employed for calculating the image surface gradient ∇I has been to determine the partial (or, in this application, directional) derivatives of an image by obtaining the first difference images (cf. the second difference approximations to the second derivative of chapter 4) in the x and y-directions and then combine them vectorially as shown above. Note that the gradient operator is symmetric in the sense that it is independent of the coordinate frame used to calculate the partial derivatives. In an approach similar to that taken in chapter 4, consider that the image surface gradient ∇I can easily be calculated at each location in either of two frames of reference, that of the usual system with axes aligned with the horizontal and vertical directions, and another at a rotation of $\frac{\pi}{4}$ radians relative to the first, where the axes are aligned with the diagonal directions of the image. These coordinate frames are respectively illustrated as frames **H** and **D** in figure 6.6(b). In effect we can obtain two independent

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$\Delta'_{\mathrm{N}} \qquad \Delta'_{\mathrm{NE}} \qquad \Delta'_{\mathrm{E}} \qquad \Delta'_{\mathrm{SE}}$$

Figure 6.5: The first difference directional convolution kernels for estimating the image surface gradient. Note that the labelling of these kernels appears contradictory to the current task (that is the Δ'_{N} measures the image gradient in the Easterly direction, and so on), however historically these kernels were used to find edges in images, since they lie at directions orthogonal to the gradient direction, and accordingly this is reflected in the labels given to the kernels.

estimates of the image surface gradient at each location. In light of the statements above, if a true boundary were present at the current position, one would expect the directions of the two estimates to approximate one another closely, whilst if a given region were corrupted by noise the two vectors would generally point in quite different directions.

To calculate the partial differences comprising the gradient vectors in the **H** and **D** frames, consider the rotational family of first-difference images given as follows. Adopting the notation of section 4.4.9, let Δ'_i be the n-first difference filters respectively with orientations α_i, where $i = \{\mathrm{N}, \mathrm{NE}, \mathrm{E}, \mathrm{SE}\}$, as displayed in figure 6.5. Let the convolution of these filters with the image I be represented as $I'_i = \Delta'_i * I$. Without loss of clarity, also allow I'_i to represent the first-difference value at pixel (x, y) in the i-direction as appropriate. Observe that between the four convolutions, each pixel is involved in eight difference calculations, one each with the immediate 8-connected neighbours of the current pixel[7], and therefore at each pixel there are eight estimates of the partial image gradients available from which to construct the local image gradient, as illustrated in figure 6.6(a).

Note that each partial difference pair formed by the application of each first-difference kernel maps directly onto one of the four axes of figure 6.6(b). By combining the four partial differences that map onto the **H** frame into a single vector, and equivalently combining the four that map onto the **D** frame into a single vector, one obtains two independent measures of the local image gradient, which might appear as illustrated in figure 6.6(c). The calculation of each vector component from the two partial differences mapped onto each axis is symmetric and we thus consider the one dimensional case only. It is at this point that our a priori information about the image surface characteristics can be directly introduced to improve the gradient estimates. Since small isolated peaks are noise, and only image surface gradients extending with some degree of uniformity over the two partial differences can be considered as representative of the features present in the CLS-removed image, we formulate the component fusion as follows. Let the component of the vector under consideration be C and the two first-differences which map onto the direction of C be labelled a and b, then

$$C = \begin{cases} 0 & \text{iff } \mathrm{Sign}(a) \neq \mathrm{Sign}(b) \text{ and } |a| \approx |b| \\ a & \text{iff } \mathrm{Sign}(a) \neq \mathrm{Sign}(b) \text{ and } |a| \gg |b| \\ b & \text{iff } \mathrm{Sign}(a) \neq \mathrm{Sign}(b) \text{ and } |a| \ll |b| \\ a + b & \text{otherwise} \end{cases} \tag{6.6}$$

[7]Obviously at the edges of an image the pixels are only involved in three or five difference calculations depending upon their locations. In practice one does not attempt to analyse an image at the extreme edges due to these effects.

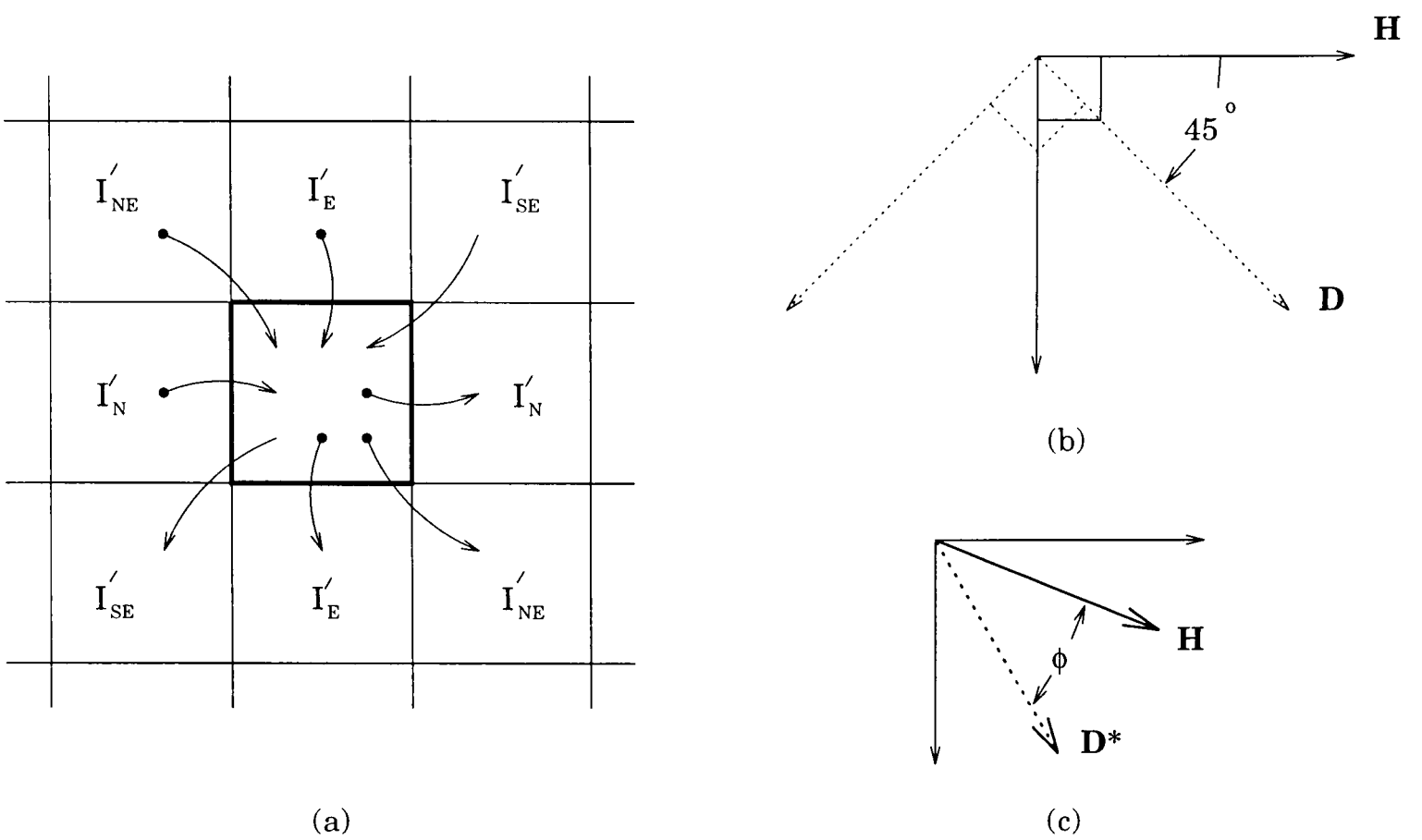


Figure 6.6: Calculating ∇I. (a) The neighbourhood surrounding a pixel and the first difference filter responses. The arrow shows the direction of a positive first difference response, whilst the dot shows the location of the top right corner of the filter kernel used to make the calculation. (b) The orthogonal coordinate systems of the vectors **H** and **D**. Note that the responses of the I'_{N} and I'_{E} of (a) are aligned with the axes of the **H** system whilst I'_{NE} and I'_{SE} are aligned with the **D** system. (c) The vectors **H** and $\mathbf{D}^*$ assembled from their components in the **H** and **D** coordinate frames of (b) respectively. Note that the vector assembled from the components of coordinate frame **D** of (b) has been rotated through $\frac{-\pi}{4}$ to be expressed in the coordinate frame of **H**. It has thus been labelled $\mathbf{D}^*$ to reflect this re-expression.

To facilitate an explanation of this equation consider figure 6.7 which pictorially depicts the scenarios accommodated by this equation. Figures 6.7(a) and (b) illustrate the image scenarios dealt by the first criterion of the equation. Whilst simple addition of the a and b components would often yield the zero response enforced by this criterion, case (b) illustrates one of our much vaunted 'noise' responses, and it is not certain that straight summation of the components would eliminate its effects. Accordingly we clamp the value of C to zero in these instances. Case (c) illustrates a situation that pits a dominant gradient against a relatively weak one. Since we know that the only strong gradients in the CLS-removed image are associated with either low spatial frequency structures (including the parenchymal tissues, pectoral muscle limits and densities etc.) or noise processes, although these have much lower magnitude than the responses to real structures, we choose to assign C the value of the dominant gradient, as detailed by the second and third criteria of equation (6.6). Naturally any situation in which the directions of a and b concur, as in cases (d) and (e), is indicative of a consistent gradient at that location and C is assigned the simple summation of the components, as detailed by the last criterion of this equation.

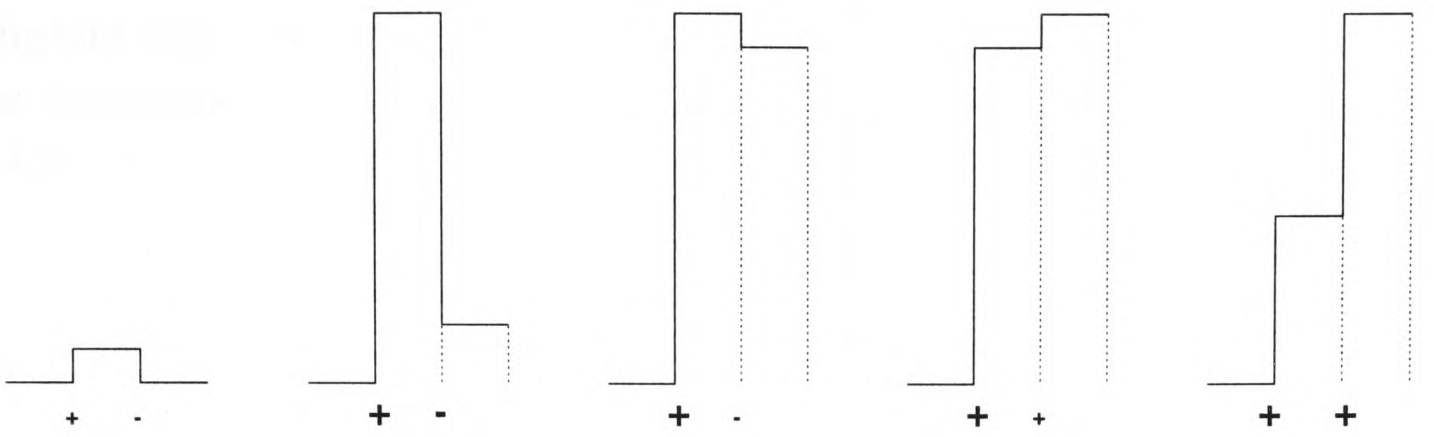


Figure 6.7: Combining the first difference information in one dimension. The solid line of each curve is a bar graph depicting image intensity for three adjacent pixels. Without loss of generality the first difference is calculated by subtracting from a given pixel intensity, that of the pixel immediately to the left. The relative results are displayed by the '+' and '–' signs under each pixel transition. A large **+** or **−** indicates a relatively large magnitude for the first difference, whilst a small + or – indicates a smaller magnitude.

Successive application of equation (6.6) to each pair of partial differences yields the components of the image surface gradient vector estimates in the **H** and **D** frames. Vectorial combination of these components yields two independent vectors, which in a slight abuse of notation we also label **H** and **D** respectively. Express **D** in terms of the reference frame of **H** (by rotation of the frame basis through $\frac{-\pi}{4}$ radians) and label this vector $\mathbf{D}^*$. The vectors **H** and $\mathbf{D}^*$ are independent measurements of the local image surface gradient at the current pixel, subtending an included angle ϕ as illustrated in figure 6.6(c). We are now in a position both to calculate an estimate of the gradient vector and to assess our confidence in that estimate.

Clearly the gradient estimate itself is given by the vector addition of **H** and $\mathbf{D}^*$, whilst the confidence in this assessment is directly related to the angle ϕ subtending the two vectors, that is:

$$\begin{aligned} \nabla I &= \mathbf{H} + \mathbf{D}^* \\ w &= w(\phi), \text{ where } 0 \leq \phi \leq \pi \end{aligned}$$

The task which befalls us now is to decide the precise form of the confidence weighting function $w(\phi)$. Clearly the domain boundary conditions are $w(0) = 1$ and $w(\pi) = 0$, and the function must be monotonically decreasing (or constant, but not increasing) over intermediate angles from 0 through π radians. Observe that due to the image quantisation (or put another way, the small spatial scale over which the vectors **H** and $\mathbf{D}^*$ are estimated) relatively minor perturbations in image intensity can result in considerable deviations in ϕ at values close to the domain limits, at say, the outer quartiles of the domain (i.e., $\phi < \frac{\pi}{4}$ or $\phi > \frac{3\pi}{4}$). Accordingly we seek to retain the boundary values of the weighting function for some distance into the domain, and we have empirically found the following simple rule to be an effective implementation:

$$w = \begin{cases} 1 & \phi < \frac{4\pi}{9} \quad (80 \text{ degrees}) \\ 0 & \text{otherwise.} \end{cases}$$

By calculating the surface gradient and confidence weighting given by these equations at every pixel in the CLS-removed image, the evaluation of the blob saliency metric $\mathcal{B}$ given by equation (6.5) is trivial for any blob under consideration. With this in mind we now describe the algorithm for locating all the salient blobs within an image.

6.3.3 The Blob-detector algorithm

With the necessary components now in place for the assessment of a blob's saliency, we implement the blob detector algorithm according to the concluding comments of section 6.2. The algorithm is completely described in the pseudo-code of figure 6.8, although the following brief elaboration of some of its finer details may aid the reader when considering the details of that figure.

Recall from section 6.2 that to overcome the difficulties of detecting the most salient blobs from a given region we initiate an expansion process which terminates upon either reaching a delimiting saddle point, or when the rate of increase of the blob support area exceeds some threshold, at which point the saliency of the blob is measured and recorded. To decide if the current segmentation is the *most* salient for the local region, we 'look-ahead' to see if there is another more salient segmentation near by. This is done by allowing the expansion process to continue beyond the current segmentation. Each time a new blob segmentation is reached we record the saliency of that blob, which may or may not be greater than the previous one. By looking ahead a number of times we can decide if the blob with the current maximum saliency is the *most* salient one for that region. Heuristically, we have found that if we look-ahead four (4) times beyond the current maximum, and do not find another blob of greater saliency, then the current blob is the most salient in the immediate region.

Note that a great many of the details of the algorithm are concerned with its execution speed and seek to reduce its computational expense by identifying those three-dimensional regions of the image space (two dimensional coordinates and a third representing intensity) that have already been searched during a previous search initiated at another location. Since the expanding process searching for the salient blobs surrounding a given pixel will expand to include the surrounding pixels, there is no point initiating another expansion process at a neighbouring, nearby or even distant pixel, if the blob that would be identified has already been identified. If it is established that the current location (x, y, I) (actually $(p, I(p))$) has already been assessed then the current search is terminated. To enable the identification of these situations, images L, H respectively record the lowest and highest intensities at which each pixel has been assessed for the existence of a blob, whilst image D assesses if the L intensity is still *high* enough to sustain the possibility of a blob at that location. Clearly we do not wish to identify a single blob describing the entire parenchymal tissue, or the entire breast area for instance, and therefore we implement a heuristic maximum blob area, above which the current expansion is terminated. The image D indicates whether the L intensity for a given location gives a segmentation with an area beyond the threshold or not. Thus, if an expanding process

Let p and q be variables representing a pixel, B the current base-level, $P(p, B)$ the set of all pixels that comprise the blob at pixel p with base-level B, $A(p, B)$ the support area of that blob, $V(p, B)$ the blob volume, $\mathcal{B}(P)$ the saliency metric for the blob P (given by equation (6.5)), I the CLS-removed image, $I(p)$ the CLS-removed image intensity at p, b_h and b_l variables recording the initiating (high) base-level and terminating (low) base-level for the *current* segmentation, $H(p)$ and $L(p)$ are data arrays used to store the highest and lowest base-level respectively at which the pixel p has been evaluated for the existence of a blob, $D(p)$ is a binary array recording if $A(p, L(p)) > T_2$, and a is a counter recording the number of times we have 'looked-ahead' of the most salient blob for the existence of another more salient blob. Given these definitions the algorithm proceeds as follows:

A $\forall p \in I$: set $H(p) = 0$, $L(p) = 255$ and $D(p) =$ False.

B $\forall p \in I$ such that $\big(I(p) < L(p)$ and $D(p) =$ False$\big)$, or $I(p) > H(p)$:

B.1 $B = I(p)$

B.2 $a = 0$

B.3 Find a blob as follows:

B.3.1 $b_h = B$

B.3.2 if $A(p, B) < T_1$, then reduce B until $A(p, B) \geq T_1$

B.3.3 if $A(p, B) < T_2$, then continue to decrement B (i.e., $B = B - 1$) until one or more of the following criteria are met:

- $A(p, B)\, V(p, B + 1) > A(p, B + 1)\, V(p, B)$
- A delimiting saddle-point is reached
- $A(p, B) \geq T_2$
- $\exists q \in P(p, B) : D(q) =$ True

B.3.4 if $A(p, B) \geq T_2$, or $\exists q \in P(p, B) : D(q) =$ True, then

- $\forall q \in P(p, B),\ \ D(q) =$ True
- there is no blob to be found here, so move on by jumping to step B

B.3.5 $b_l = B = B + 1$

B.3.6 $\forall q \in P(p, b_l)$:

- if $b_h > H(q)$, then $H(q) = b_h$
- if $b_l < L(q)$, then $L(q) = b_l$

B.4 if $\mathcal{B}\big(P(p, B)\big) >$ the previous maximum $\mathcal{B}$, then save the current blob parameters p and B and set $a = 0$, else increment a.

B.5 if $a < T_3$, decrement B and return to step B.3

B.6 declare the blob $P(p, B)$ to be the most salient for the region surrounding p and record it as a member of the set describing all the blobs of the image.

C Continue whilst $\forall p \in I$ has not expired

The thresholds are normally set as follows: $T_1 = 10$ pixels = minimum support area that a blob can take (used to ensure that boundary comprises sufficient pixels to make the saliency estimate meaningful); T_2 =5000–10000 pixels = the maximum blob area (used to avoid segmenting the *entire* breast area etc.); and $T_3 = 4$ = number of 'look-ahead' steps to check for a more salient blob in the local region.

Figure 6.8: The blob detector algorithm. Note that steps B.4 and B.5 address the problem identified in figure 6.2(a), whilst the first criterion of step B.3.3 addresses that of figure 6.2(b). The tests of step B and B.3.4 minimise the computational expense of the algorithm by ensuring that a segmentation is not initiated if it would return a blob that has already been detected, or if the initiating base-level intensity B would give $A(p, B) > T_2$. Steps B.3.4 and B.3.6 update the records for detailing this information.

wanders into a region at which it is known further expansion will give a blob area beyond the threshold, then the process is terminated. Tests for these conditions are shown in steps B and B.3.4 of the algorithm.

Although the algorithm as presented in figure 6.8 specifies that the search for a blob be initiated at each and every pixel of an image, this is clearly overkill since it is impossible for a blob not to extend to cover some region surrounding each pixel. Therefore, initiating a search at each pixel means that many will begin in regions already assessed, and although the details of the preceding paragraph would quickly identify such instances and terminate them directly, this is not a particularly satisfactory approach. We find that initiating the process at every third pixel does not fail to identify any blobs, yet as it initiates the process at only one-ninth of the pixels it further reduces the computational expense of the algorithm. Consequently the implementation of the algorithm of figure 6.8 within the present work includes a suitable alteration to step B to initiate the process at every third pixel (in both image coordinates) only.

6.3.4 Results

The results of this algorithm have already been displayed in a small preview of results given in figure 5.6 on page 130. The results presented in that figure illustrate the importance of using the CLS-removed image as the domain within which to conduct the search for the blobs, since the clutter of the CLS features in an image makes the accurate localisation of the salient blobs very difficult to otherwise achieve.

More complete examples of the algorithm's performance and results are displayed in figures 6.9 and 6.10 which illustrate the blobs identified in the images of figures 2.7 and 4.32 (pages 31 and 110) respectively.

Observe that the procedure has identified the *locally* salient blobs from each region of the images, even in those regions which to the naked eye, appear to be smooth, uniform and devoid of any blobs. This is a consequence of the algorithm's *local* assessment of relative saliency, and it is the very fact, as mentioned in the introduction to this chapter, that precipitates the need for the subsequent identification of the mammographically significant masses from this list of blobs. It is the task of identifying the masses from these blobs that we now direct our attention.

6.4 Identifying significant densities

With a description of the locally salient blobs in an image available, the task is to determine those blobs that are mammographically significant from those that are not. From the outset we should be under no illusions—this is a very difficult problem. Typically a radiologist will spend years developing and fine tuning his/her own decision criteria based upon exposure to thousands of mammograms in the clinical setting. Such a wealth of experience is indispensable in learning the characteristics of the mammographic signs indicative of disease.

From a perspective of automated analysis, this is a classic pattern recognition task where the objective is to identify a given class of widget (mammographic densities) from the domain

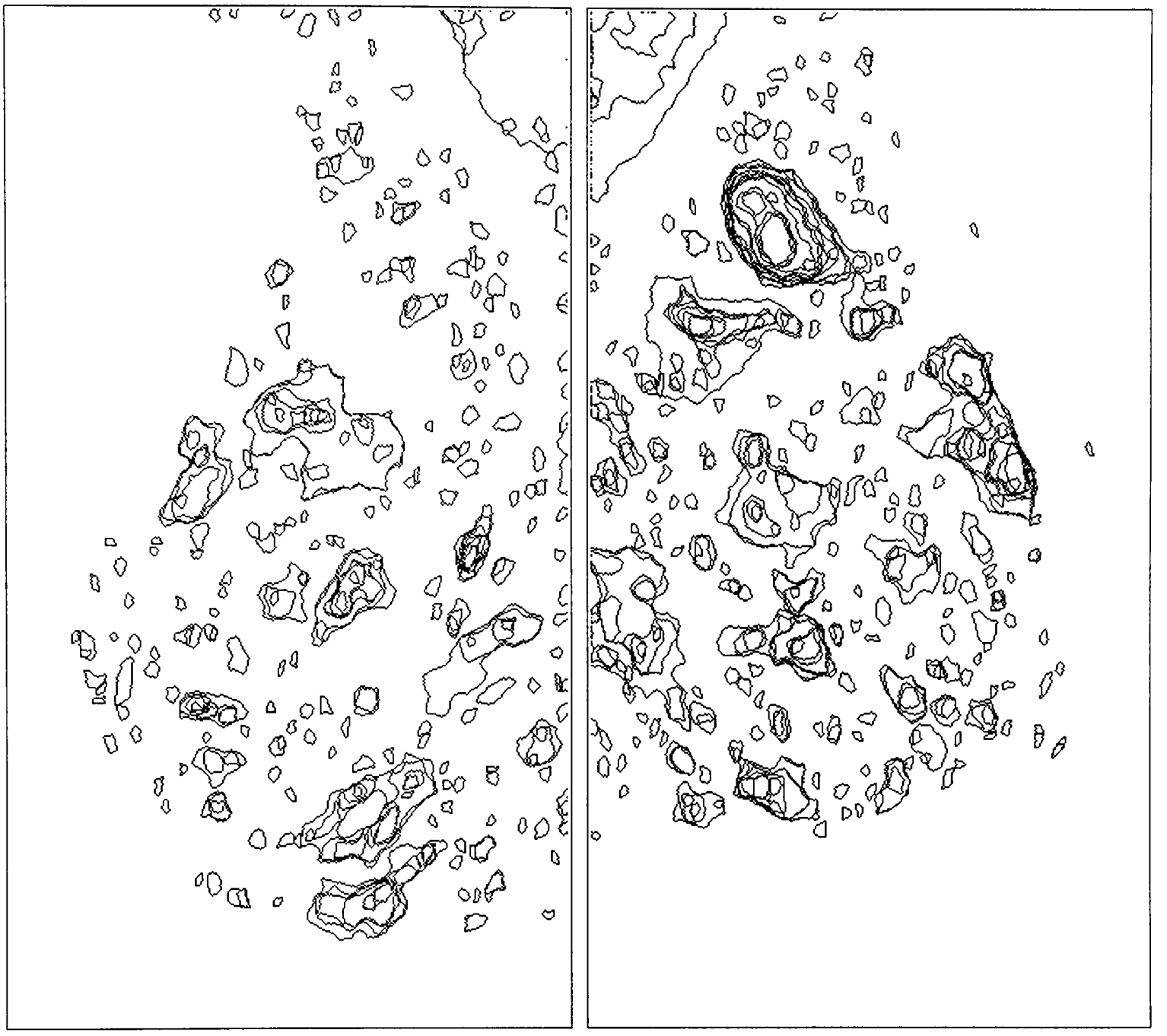


Figure 6.9: The locally salient blobs identified in the images of figure 2.7 on page 31.

of all widgets presented (the blobs). In situations where the characteristics which exactly define the widgets of interest are not completely known a priori (such as the current one), identification of those widgets often proceeds by statistical analysis, whereby an attempt is made to characterise the feature space describing all the widgets into regions that are most likely to be that of the class of interest. For the task at hand, such statistical analysis is an attractive option, provided due caution is exercised in its application to ensure statistical significance. So protean are the attributes of the mammographic densities that regardless of the implementation details chosen, the statistical assessment of their characteristics can only be made significant by analysis of a very large number of examples. It is to this point that we shall be forced to return on a number of occasions.

A favoured solution for implementing this type of philosophy has recently been that of the neural network approach, and it is little wonder then that some researchers have applied these techniques to the task of identifying significant mammographic features. It is however dis-

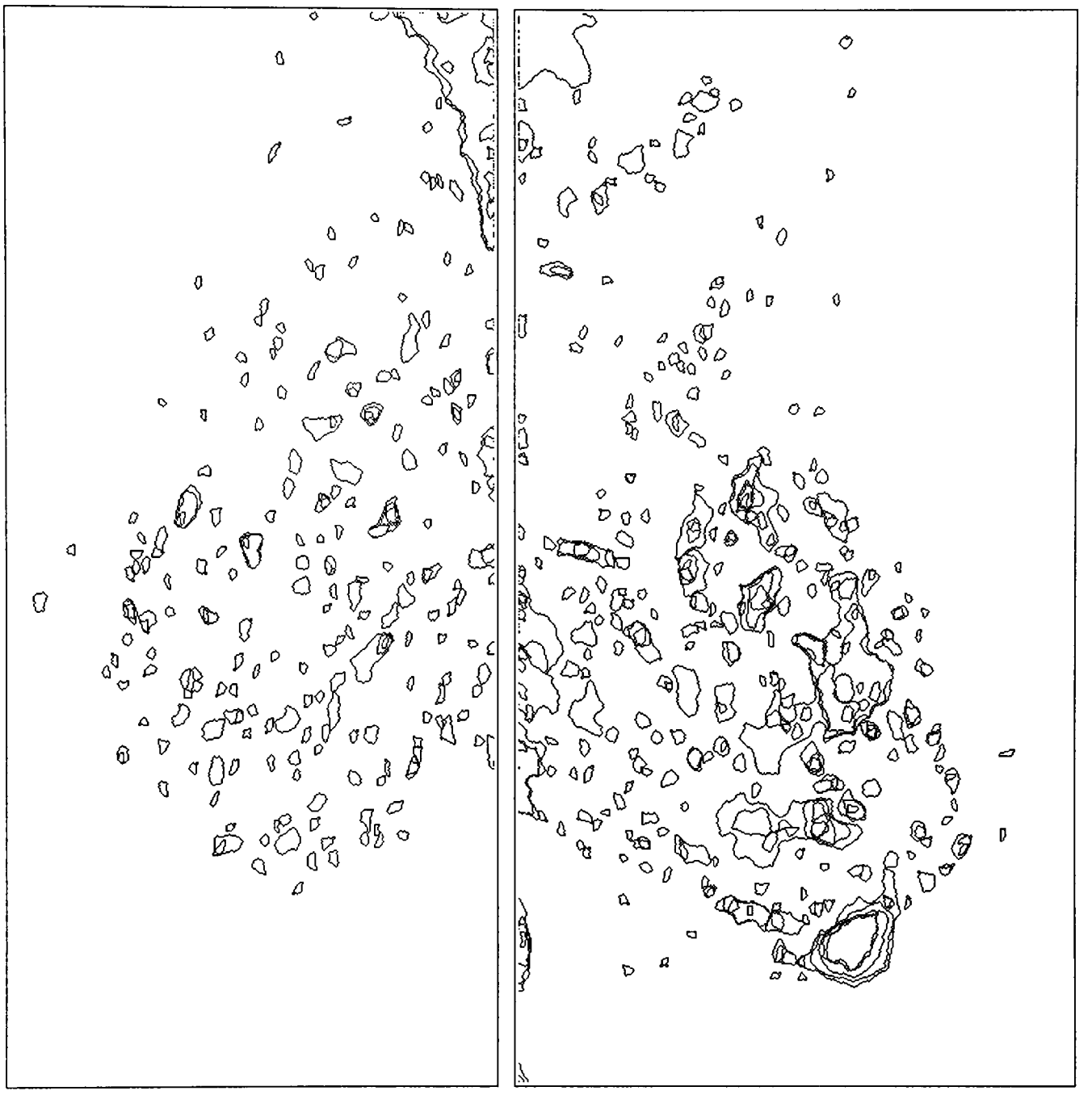


Figure 6.10: The locally salient blobs identified in the images of figure 4.32 on page 110.

appointing to note that, in general, these techniques have been inadequately, inappropriately, and even ineptly applied, giving commensurately disappointing results. Without exception, the approach adopted by these implementations has been to emulate the approach taken by the clinician. That is, by analysing examples of real mammographic densities in comparison with insignificant blobs, and thus assessing the statistics that separate each type, the method attempts to duplicate the 'experience' that a clinician develops with much the same approach over a much longer time period. By emulating the clinician's approach, these methods are also bound by its limitations, and the most important of these is experience, or more devastatingly, the lack of it. The wide variety in the appearance of mammographically significant densities means that a clinician will begin his/her career with a relatively modest success rate for the task of correctly identifying all mammographic signs of disease. Naturally, as the clinician gains experience, through exposure to more abnormal cases of varying diversity, his/her performance improves, and so it should be, in theory, with an automated analysis system. Without

sufficient a priori knowledge defining the limits of the class of interest, it is absolutely imperative that the analysis be presented with enough examples of the class, so that the limits can be captured. Inadequate exposure to cases of the class of interest gives an inadequate statistical description of the class and makes the chances of achieving a quality segmentation very slim indeed. By analogy, this situation is rather like asking a first-year medical student to analyse a mammogram. Without any a priori knowledge of what to look for in the film the student has little or no chance of correctly interpreting the signs of disease that may be present in the film.

To formalise this description, consider a given n-dimensional data vector quantifying the parameters chosen to describe each blob. The standard neural classifier attempts to define decision boundaries in n-space on either side of which one class (the significant densities, say) is more probable than another (the insignificant blobs), and vice versa. Observe that this probabilistic/statistical approach is built on Bayes' theorem:

$$P(C_k|\mathbf{x}) = \frac{P(\mathbf{x}|C_k)P(C_k)}{\sum_{i=1}^{N} P(\mathbf{x}|C_i)P(C_i)} \quad k = 1, 2, \ldots, N \tag{6.7}$$

which allows calculation of the posterior probability that a given data vector $\mathbf{x}$ is from class k. Assessment of this equation over all classes k allows determination of the most probable class describing the vector $\mathbf{x}$. In terms of the current task, this last result is the very one we seek for segmentation of the significant masses from the insignificant blobs.

Although the neural classifier, given a data vector $\mathbf{x}$ presented for classification, directly estimates the left-hand-side of this equation for each class $k = 1, 2, \ldots, N$, its ability to do so can be understood from an analysis of the right-hand-side. During the 'training' phase of the classifier, a large range of example data vectors spanning the limits of each class k is presented and the statistics assimilated. During classification then, one could determine the three components of the right-hand-side of equation (6.7), that is: the probability $P(\mathbf{x}|C_k)$ that a data vector $\mathbf{x}$ is part of that class k; the probability $P(C_k)$ of class k over all other classes; and finally the normalising factor $P(\mathbf{x}) = \sum_{i=1}^{N} P(\mathbf{x}|C_i)P(C_i)$ is merely the total probability of a data vector (of any class) at $\mathbf{x}$. Assessment of equation (6.7) for all $k = 1, 2, \ldots, N$ would reveal the probability of a vector $\mathbf{x}$ belonging to each class, and would thus determine the most probable class for $\mathbf{x}$.

Before simply rushing out to one's local supermarket, buying such a classifier and plugging in any data that can be found, there is a element of this formulation that must be considered. Observe that the probability of $\mathbf{x}$ belonging to class k, i.e., $P(C_k|\mathbf{x})$, is directly proportional to the probability of that class $P(C_k)$. In the intended application of this theorem (in a neural classifier), $P(C_k)$ is *not* a measure of the relative occurrence of each class k (densities verses blobs) in the population space, but is rather a measure of the relative occurrence of each class k in the sample space of the training data. It is desirable that a classifier should base its decisions on the particular *characteristics* of the data vectors rather than on the relative *number* of data vectors presented during training. The only way to achieve this desired result is to remove any bias that could be introduced in this fashion, and this is done by presenting the classifier with

equal numbers of data vectors describing each class.

A major difficulty in exposing both clinicians and automated systems to statistically significant numbers of the pathologic cases is that in the clinical screening environment the rate of presentation of mammograms displaying an abnormality is extremely low. In the United Kingdom alone, approximately three million screening mammograms are acquired each year. Of these, only 0.5% exhibit abnormalities. This is of course an extremely high rate when considered from the perspective of saving lives, however, it is a very low rate from the perspective of trying to characterise abnormalities. Given the disproportionate presentation of significant masses verses the number of films acquired in clinical practice it is a very difficult task to find sufficient examples of the real mammographic densities. Within the context of the current application, this situation is exacerbated by the vast numbers of insignificant blobs generated by the blob-detector algorithm of the previous sections. The algorithm will typically segment 150–600 blobs from each and every image, irrespective of the existence of a real density or not. Thus for every real mammographically significant density in the data-set, there are of the order of 50000 insignificant blobs that should be rejected. In such a situation any attempt to extract balanced data sets for presentation to the classifier is clearly a hopeless one. The obvious solution to some of these problems is to scour through the vast stores of mammogram films available in search of films with examples of abnormalities so as to boost the number of real mammographically significant densities available for analysis. Despite the (possibly temporary) technology limitations regarding image acquisition, storage and analysis, the fact that *every* image gives rise to 100–600 blobs means that even if every image in a database contained an abnormality, the data-sets would still be unbalanced by 2 to 3 orders of magnitude.

From these considerations it is obvious that the standard methods of neural classification are not suitable for the segmentation problem as formulated in this application. Sadly, it is these very techniques that have been implemented in the literature, giving indifferent results, as previously mentioned. With very few examples of real abnormalities, the requirement of balanced data sets has been met by exposing the classifiers during training to the relatively few vectors available, such that the statistical significance is at best questionable. Consequently it is little wonder that the results from these implementations have been disappointing. For completeness, we initially review the limited amount of published work in the literature, and then proceed to describe the techniques used in the present work to circumvent these difficulties.

6.4.1 Previous work

Brzakovic et al. [23] identify blobs by simple thresholding of an image, which is then followed by multi-resolution Gaussian pyramid approach to identify the salient regions. The blob boundaries are identified using the Marr-Hildreth zero-crossing edge detector. Classification of the blobs into significant masses and insignificant blobs is conducted by a linear decision tree. The identification of a possible tumour is assessed purely on the basis of its area. The

subsequent assessment of three metrics, a shape descriptor, an edge intensity variation metric and an edge-distance variation metric were done by a separate Bayesian classifier. The three classifiers were trained with the same 5 examples of malignant tumours, 5 benign tumours and 8 non-tumour blobs extracted from 10 mammogram images. Testing was conducted on 15 mammogram images of which 10 contained tumours. The authors report an 85% correct classification rate. Given our earlier introductory comments, there can be no question that this study is statistically insignificant, and whilst the acquisition of further data may prove the usefulness of this technique, these results must simply be considered as preliminary.

Tsai et al. [213] use a feed-forward multi-layer perceptron classifier with 1000 input units and 110 hidden units to classify 200×200 pixel blocks constituting 'regions-of-interest' into classes of benign or malignant. They train the network on a total of 10 benign and 10 malignant patterns extracted from 20 mammogram images. In essence they attempt to determine the values of 110 000 free parameters from 20 known data points. This is some orders of magnitude worse than trying to fit a straight line to a single data point! The authors note with enthusiasm that after training, "all 20 images used for learning were recognised correctly". Their network was further tested on another 20 images. Remarkably, the authors claim "that the neural network can correctly classify benign and malignant tumours at an average rate of 85%". They even have the audacity to claim "the recognition rate determined by the present method is considered to be comparable to that [of] the several-year experienced physician". Short of being even merely statistically insignificant, these results are simply meaningless as the network is completely under constrained.

Kegelmeyer et al. [120, 117] describe a binary decision tree (BDT) classifier designed to identify spiculated lesions directly from an image without the intermediate stage of identifying the blobs within an image. By assessing the orientations of edges in a large region, and some texture measures in a local region surrounding each pixel, they develop a 5 dimensional data vector describing each *pixel* as the input to the classifier. From the decision tree a probability map assessing the likelihood that a given pixel is a component of a spiculated mass is constructed, and then smoothed to impose a local consensus. Training and testing of the classifier was conducted on the *same* datasets of four-view mammograms comprising 36 known spiculated lesions and 49 negative cases. The authors claim a 97% sensitivity rate in the identification of those lesions. Yet again, whilst acknowledging that the acquisition of more training data may improve the situation, the ability of the present system to capture the diversity of the mammographic appearance of spiculated lesions must be questioned, and thus the statistical significance of these results must also be questioned.

Recently Woods and Boywer [228] used linear and quadratic classifiers to analyse the database used by Kegelmeyer et al. mentioned in the previous paragraph. Unlike the pixel based approach of that study, Woods and Boywer initially try to segment a selection of blobs from an image, adopting a similar philosophy in this respect to the present work, and then use the classifiers to identify the significant spiculated masses from the many blobs found by

their segmentation procedure. In addition to the 5 metrics used by Kegelmeyer et al. (except in this case they are calculated over the region of the blob rather than a window surrounding the pixel as above), Woods and Boywer use a further 12 metrics to describe each blob. The authors train and test the classifiers on a leave-one-out basis, achieving poorer results than the methods of Kegelmeyer et al. for the same database. Obviously the statements concerning statistical significance mentioned at the last paragraph apply equally well.

This last case highlights an important fact concerning statistical classification techniques. The inclusion of further descriptors into an analysis that do not provide any additional information can actually degrade the performance of a classifier based solely on the fewer descriptors. We will return to this point in the sections below when the problem is considered within the context of the present work.

Giger et al. [74] use a back-propagation feed forward classifier to segment known masses into classes of malignant and benign. Their classifier of 20 free parameters is trained using 90 examples of masses (it is unclear how many are malignant/benign, however it is safe to assume approximately 45/45), and evaluated on a further 57 malignant and 38 benign masses. Once again, despite the suitability of this method to analysing such problems, the ability of this implementation to characterise the diversity of mammographic appearance from 50 odd examples of malignant and benign masses is questionable.

Neural classifiers have been applied to other areas of mammogram analysis [38, 119, 156, 165, 203, 150] yet they have always suffered from training based on small datasets for the reasons previously outlined. Additionally, the testing and cross-validation that has been undertaken has either used the training data itself, or used very small separate datasets, adding to the concerns of statistical significance expressed earlier. In light of these comments, we now present the approach adopted within the present work to overcome these difficulties.

6.4.2 Novelty detection

The current task befalling us is to segment the handful of real mammographically significant masses from the plethora of insignificant blobs identified by the blob-detector of section 6.3. To facilitate our attempts to overcome the limitations of the standard neural classifiers as highlighted above, we initially consider them in a little more detail to gain an appreciation of why they perform as they do.

Difficulties with standard neural classifiers

As noted in the introduction, the standard neural classifier seeks to identify probabilistic decision boundaries between two or more classes. To illustrate some of the limitations of this approach consider an example of a two-class problem, such as the segmentation of people by gender. Suppose in our example, that the data used to describe each widget (person) is a two-dimensional vector (say the weight w, and height h of each person). If a large sample of typical vectors representative of the two classes were obtained one might expect a distribution

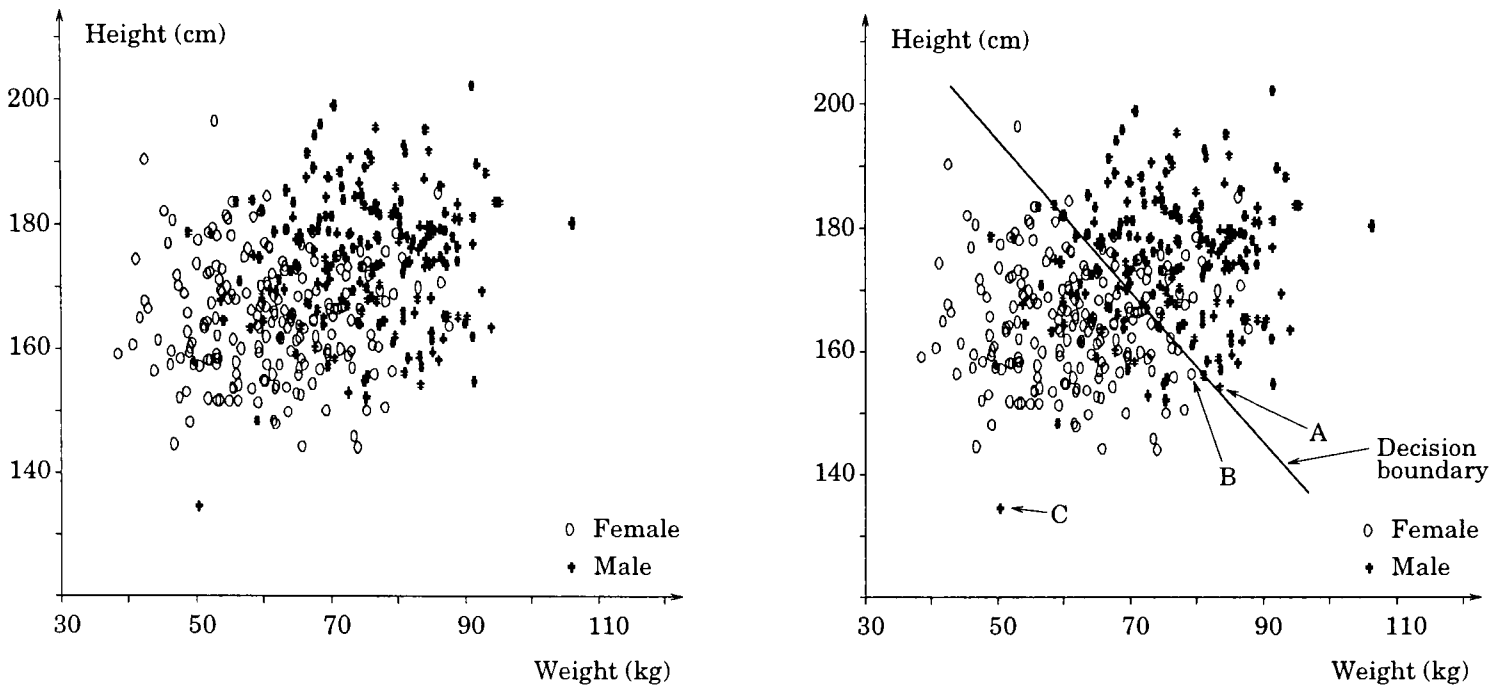


Figure 6.11: Segmenting data encoding the weight and height of people into classes of Male and Female. (a) A typical distribution of such data. (b) The decision boundary that a traditional neural classifier might construct with the Male class being more probable above the boundary whilst the Female class would be more probable below it.

of the form shown in figure 6.11(a). Although it is the case that on average males are taller and heavier than females, and thus the differentiation of the distributions might be possible, there will clearly be a significant region of overlap. Analysis of the distribution of points in figure 6.11(a) might reveal the location of the boundary most *likely* to separate the Male classification from the Female as the line shown in figure 6.11(b). At the location of the decision boundary the probability of a data vector belonging to the Male class is equal with that of Female (that is $\frac{1}{2}$) and accordingly no firm classification can be given. On either side of the boundary, the classifier gives the posterior probability that the data point belongs to one class or another (equation (6.7)). With this scheme then, vectors labelled 'A' and 'B' in figure 6.11(b) would be classed as 'Male' and 'Female' respectively with a probability slightly greater that $\frac{1}{2}$. But what of the point labelled 'C'? Assume for the moment that the correct answer is that this is a point describing a male dwarf. From the decision boundary of figure 6.11(b) the classifier might class this point as Female with high probability, an obviously incorrect result. Observe that the point labelled 'C' is a lone point in a region of input feature space unpopulated by other data points. Although the classifier might class this point as a 'Female' with high probability, it really does not have any evidence upon which to base this assertion. The lesson to be learnt here is that traditional classifiers are quite happy to pass judgements on *all* vectors presented for classification despite the fact that during the training phase the classifier may *never* have been presented with any data to describe the region of n-space within which the current vector lies. Furthermore they can dangerously give the impression that the current vector is well described by the training data by responding with a highly probable classification, as in the case of the male dwarf.

The 'take-home-messages' from this example and the introductory comments regarding pat-

tern classification in general, are that to achieve accurate decision boundaries such methods require training data that spans the *entire* basis of the feature space defining each class, and they need sufficient amounts of such data to accommodate the variability of the components of the data vectors. Each of these constraints pose a significant problem within the context of the present task: (a) there is no technique to determine *all* possible variations of significant mammographic densities; and (b) the acquisition of significant numbers of examples of mammographic densities is problematical, as previously discussed. The failure to appreciate these hard facts has led to the disappointing results reported in the literature mentioned in section 6.4.1.

Novelty detection

Given that the time and resource requirements for the acquisition of large balanced data-sets of mammographically significant and insignificant blobs cannot be met for the foreseeable future, how then are the limitations and constraints of the neural classifiers to be overcome? The solution is to turn the entire problem upside-down. Consider the 'novelty' analysis techniques introduced by Tarassenko [210, 211]. Novelty analysis is particularly suited to detecting rare pathologic events from amongst a vast array of 'normal' events. The current task is, quite obviously, an example of such a scenario in which the class of interest, the mammographically significant densities, occur very infrequently amongst the vast list of blobs identified by the blob detector described above. In the broadest generalisation, this identification of the interesting cases is achieved by learning a description of the normal or insignificant cases, of which there are many examples available for training, and then testing a given data vector for *novelty* against the description of normality. That is, by building a description of normality by identifying the regions of input vector space consistent with the description of 'normal' or 'background' data vectors, then any given vector can be assessed for 'novelty' against that description by determining if the region of vector space it occupies is, or is not, well described by normal/background data. Note that this approach is fundamentally different from that of the probabilistic neural classifier which seeks to determine the probability that a given location in the vector space is of one particular class or another, assigning 'decision boundaries' at locations of equal probability between different classes. To appreciate this difference, consider the example of the gender segmentation problem developed above. Suppose that the problem was formulated by considering the 'Female' class to be normal, and thus the classification of a given data vector as belonging to the 'Male' class would be achieved by observing that the data vector was not well described by the normal data (which represents the 'Female' class in our formulation) and thus it must be, in the logical sense, 'Not Female' which we therefore interpret to be 'Male'. It is important to observe that the reasoning here is intimately related to the choice of data representation, that is by choosing a novel response to represent 'Not Female' we make the assumption/interpretation that this is representative of the 'Male' class.

In this example we might decide to declare the region of feature space occupied by the

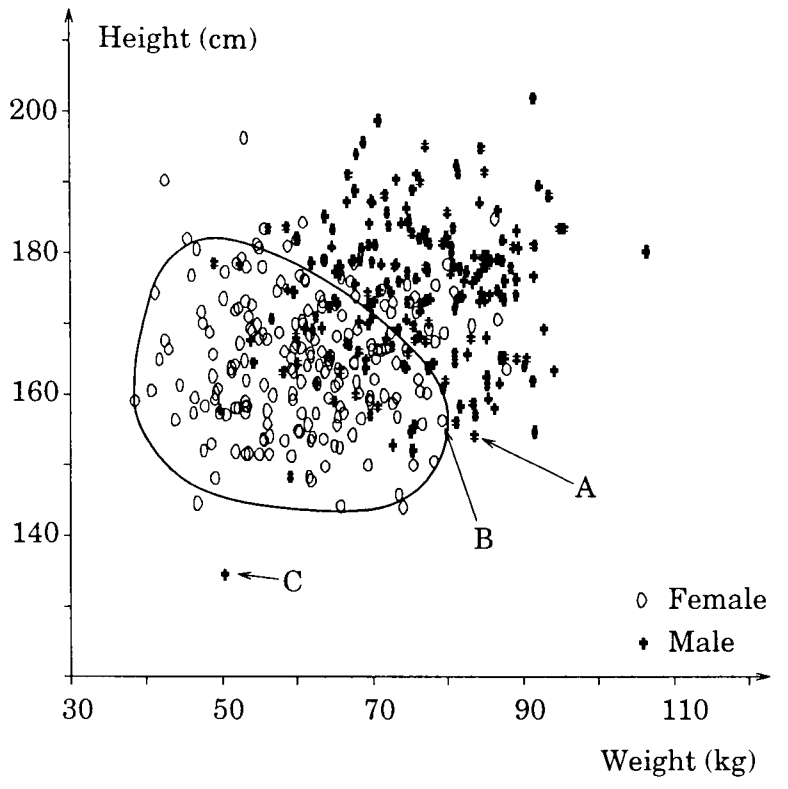


Figure 6.12: Defining a region of input vector space describing the Female class of data. Identification of a test vector as Male would proceed by noting its *novelty* with regard to the *normal* region describing the Female class identified during training of the system.

'normal' class as that enclosed by the boundary shown in figure 6.12. A test vector situated a long way outside this normal region would most likely be novel compared to the description of normality, whilst those still beyond the normal region, but closer to it would still be deemed novel, but with less certainty. Obviously the simple Euclidean distance from the centroid (say) of the normal region would not be a useful measure of novelty due to the possible variations in the size of the normal regions. On the other hand, a measure of novelty which accommodates the size, or *width*, of a normal region—such as the Mahalanobis distance of the test vector relative to the region centroid—would be a suitable measure.

Clearly in the context of the current task, these techniques can be used directly to identify the real mammographically significant densities as those that are novel in comparison to a description of normality, given by a description of the insignificant blobs that form the large proportion of the data sets available to us. By approaching the problem from this perspective we can learn a statistically significant description of the normal background by using *all* the data describing the insignificant blobs available to us, of which we have a vast amount, without the limitations that Bayes' theorem imposes by demanding balanced data-sets. Provided a judicious choice of the parameters of the data vector describing each blob is made such that the regions of the space occupied by the normal and the novel widgets are separable, and given enough data spanning the variety of normal/background cases that are likely to appear, such that the region of input vector space that is occupied by the description of normality (i.e., the insignificant blobs) can be accurately assessed in a statistical sense, then the identification of a novel data vector (i.e., a significant density) can proceed directly.

The questions yet to be addressed in order to implement this solution philosophy are: (a)

given a collection of data vectors known to describe examples of the normal/background (from which any vectors describing widgets from the 'novel' class have been selectively removed), how does one identify the regions of vector space that these training data vectors occupy; and (b) identification of a 'judicious' choice of descriptors with the properties mentioned above to encode the blob attributes. Clearly the first problem here is a generic one that occurs during any novelty analysis, whilst the second is particular to the task at hand and the feature descriptors available to describe each widget. Consequently we will tackle these problems in the sequence in which they were presented. Before proceeding to consider these questions, it should be noted that the final statement of the preceding paragraph summarises in total the philosophy upon which the solution to the current identification task is based. The details of the implementation, to be discussed below, could take many different forms, and the specific results obtained will reflect these choices. From the perspective of this thesis however, it is the solution philosophy embodied in that sentence that we present as the main result of this section, and although the details of the implementation discussed below have not been optimised, it is the results they provide in support of the notion that this technique is indeed suitable for the segmentation task at hand that is important.

Describing 'normality'

Suppose that we aim to learn a description of normality from a large data-set of n-dimensional training vectors $\mathbf{x}$, each describing an example instance of the 'normal' class. Clearly the best description of normality, in the statistical sense, is the unconditional joint probability density function $p(\mathrm{x})$ evaluated over the available training data-set. Thus if a test vector $\mathbf{x}$ belongs to a region of the n-dimensional feature space for which $p(\mathrm{x})$ is small, then that vector would be deemed to be novel. Obviously determination of $p(\mathrm{x})$ within the limitations of the current problem is strictly impossible, however there are a number of possible ways to obtain an estimate $\hat{p}(\mathrm{x})$ of the pdf.

Although computationally expensive, the Gaussian mixture model initially appears as the most logical option to pursue, modelling $\hat{p}(\mathrm{x})$ as:

$$\hat{p}(\mathrm{x}) = \sum_j p(\mathrm{x} \mid j)\, P(j)$$

$$\text{where } p(\mathrm{x} \mid j) = \frac{1}{(2\pi)^{d/2} \mid \boldsymbol{\Sigma}_j \mid^{1/2}} \exp\left\{-\frac{1}{2}(\mathrm{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathrm{x} - \boldsymbol{\mu}_j)\right\}$$

The problem with this method is that the normalising factor is inversely proportional to the magnitude of the determinant of the covariance matrix, that is, the 'spread' of the data in each of the input dimensions. Feature vectors belonging to regions of low data density will invariably be assessed as novel, irrespective of how 'similar' they might be to the training patterns in that region of input space. This is a reflection of the fact that the method naturally applies a *global* threshold at which the boundary is defined in input space between the identified normal regions and the novel regions beyond. In regions of relatively sparse data density (but possibly

equal net *numbers* of data vectors) with correspondingly low pdf response, the threshold (as biased by the regions of high data density) may simply exceed the pdf response in that region, and thus any test vector presented in this region (which is therefore normal), would be classed as novel.

The obvious solution to this dilemma is to incorporate a *local* measurement quantifying the extents of normality for that region of feature space alone. An alternative method for modelling $\hat{p}(\mathbf{x})$ that naturally allows the incorporation of such a local measurement is to use a robust K-means algorithm [211] to identify within the training database the *clusters* which exist in the feature space. If both the clusters in the feature space for a given set of training vectors and a local measure of the scope/extent/width of each cluster can be identified, then the novelty of a test vector $\mathbf{x}$ in relation to a given cluster can be assessed by comparison of its weighted distance to the cluster centre with the width of that cluster. Thus where the width σ_j of a cluster j is known, the novelty of a given vector relative to the cluster j is set to be the Mahalanobis distance

$$(\mathbf{x} - \boldsymbol{\mu}_j)^T \, \boldsymbol{\Sigma}_j^{-1} \, (\mathbf{x} - \boldsymbol{\mu}_j)$$

where the covariance matrix $\boldsymbol{\Sigma}_j^{-1}$ is simply $\sigma_j^2 \mathbf{I}$ ($\mathbf{I}$ is the identity matrix).

An important point to note here is the use of a *robust* K-means algorithm due to the sensitivity of the ordinary K-means algorithm to any *outliers* which may remain within the training database. If the procedure was unfortunate enough to initiate one of the original cluster centre estimates on (or near) an outlier, the lack of nearby training data to perturb the location away from the outlier ultimately results in a trained cluster situated at the outlier. By labelling the outlier as a cluster centre, the assessment of novelty will be severely distorted at this location. The solution to this problem is to use a *robust* clustering procedure, based on the adaptive K-means algorithm, which does not suffer from this problem. Using a training set of 'normal' data, and another 'normal' set for cross-validation the implementation of the procedure is summarised as follows:

- Use zero-mean, unit variance transform on input data so that each feature is given equal importance.
- Present data in random order but each pattern *once* per iteration.
- Since outliers are characterised by very few updates ('visits'), keep only the most-frequently visited centres, i.e. those which together contribute to more than say, 95% of the visits.
- For each remaining cluster determine the width σ_j that best describes the training data associated with that cluster centre.
- Assess the novelty of each vector in the cross-validation set of 'normal' data by determining the minimum novelty of a given vector relative to each cluster.
- Choose the value of K which gives the minimum number of novel patterns in the cross-validation set.

Once again, there are two elements of the formulation to be considered at this juncture: (a) how to determine the width of each cluster; and (b) at what percentage of the cluster visits (95% in the above algorithm) do we consider the cluster centres to be valid? Clearly the question of outliers is intimately related to the details of the database available, and this question can only be considered by careful evaluation of each case on an individual basis. If one believes that the database describing the 'normal' class is free from outliers (known as a *clean* database) then indeed the threshold should be set to 100% (reducing the *robust* K-means algorithm to the *ordinary* one). As this question can only be answered by consideration of the data at hand, we set this aside, and proceed to consider the question of determining the width of a cluster.

Recall that the use to which the width of a cluster is to be put, is to gauge the *relative* novelty of a test vector with regard to training data associated with the cluster. When determining the novelty of test vectors, it is important that the method used to set each cluster width should be consistent, however its exact details do not matter. This observation allows us a great deal of flexibility in determining the width of each cluster. Under a further assumption that the normal clusters are likely to be 'connected' through the feature space, Tarassenko [211] set the width of each cluster to be half the Euclidean distance to the nearest neighbouring cluster. To avoid this assumption, and in light of the fact that the clustering of training data vectors within the current application were observed to be somewhat dispersed through feature space, we adopt an alternative standard method [105] of setting the width of each cluster by considering the distribution about the cluster centre of *all* the training vectors associated with that cluster. We have investigated two such methods, such that the width is set to (a) the average radius in n-dimensional space of the training vectors associated with that centre, and (b) a radius that includes some percentage (typically 95%) of the training vectors associated with that centre within the radius. Due to the *relative* assessment of novelty mentioned above, it is no surprise that the results for either method have been essentially identical, and thus we present only those results for the width of the cluster set to the average radius of its associated vectors.

Choice of feature descriptors

With the elements of the novelty analysis techniques now in place, the question posed, is what information should we use to describe each blob, and how should we encode that information such that the separation of novelty from normality produces a separation in the data of significant density from insignificant blob? Clearly the choice of information to encode can have a significant influence on the results. Observe that the novelty analysis techniques as described above will always give an assessment of the novelty of a given test vector, however the responsibility for ensuring that the segmentation of the test data into normal and novel classes is a useful one, i.e., the 'novel' class truly is novel in the interpretation of the physical situation giving rise to the statistical analysis, falls squarely on the shoulders of the user. It is through the judicious choice of the feature descriptors that comprise the data vector describing each blob that a useful segmentation is ensured.

At this point the selection of good descriptors which each provide some information that can aid in the discrimination of density from blob might look a formidable task. Furthermore, a large amount of information may be available from which to choose a smaller number of descriptors to form the data vector representing each object. For instance the gender segmentation problem might have available the weight, height, eye colour, waist measurement, star-sign, favourite dessert and bank balance for each person. A naïve first attempt might be to simply use all the information that is available. This approach is however to be strongly discouraged. Within the gender segmentation analogy, whilst identification of a region of input space occupied by the training data (comprised of the 'Female' data points) might be possible by considering the weight and height only (leading to the normal region identified in figure 6.12 say), the addition of the star-sign to the data vector for instance, a random variable distributed evenly over the two classes, and therefore over *all* input space (both normal and novel), not only provides no additional information useful for discriminating between the classes, but actually degrades the performance achieved from the (w, h) vector alone. This is due to the fact that the addition of a purely random descriptor simply disperses in $(n + 1)$-space along the axis of the new descriptor the data clusters which have formed in the fewer n-dimensions. Since we are trying to find clusters of 'normal' data by identifying regions in which the unconditional probability density function $p(\mathbf{x})$ of the training data is high, such a dispersal of the cluster degrades the identification of the regions of normality, and results in lower overall system performance. Therefore it is desirable to select those descriptors that provide some (good) discrimination between the classes and to eliminate those that provide limited or no information.

There are two tasks to be undertaken here: firstly to determine a range of blob descriptors which provide some information about the blob; and secondly to determine which of these descriptors gives good discrimination between the blobs and the real mammographic densities. With a technique for evaluating the discriminatory power of a given descriptor, the overall task of identifying a set of descriptors with good discrimination to the exclusion of those that do not becomes an iterative process.

How can the ability of a descriptor to differentiate between masses and blobs be assessed? The solution is to evaluate the descriptor for a number of examples of each case and to compare the distributions of the values. If the distributions are exactly aligned then the descriptor cannot distinguish between the two types, whilst those that show good separation between the distributions may be useful, even if there is significant overlap[8]. Fortunately we have at our disposal a large number of images[9] from which the blobs identified by the blob-detector for a sample of real potential mammographic densities have been manually extracted. Thus for any proposed descriptor, its value can be evaluated for each example of this set of known real densities and a large set of known insignificant blobs and subsequently compared to assess the discriminatory power of the descriptor. Figure 6.13 illustrates a typical example of both a

[8]Clearly there will be some overlap in the distribution tails, for if there were not then that single descriptor would provide a perfect segmentation of the classes and there would not be any need for this involved statistical analysis.

[9]The 322 images of the *MIAS Mammographic Database* [204], containing 65 images with abnormal masses.

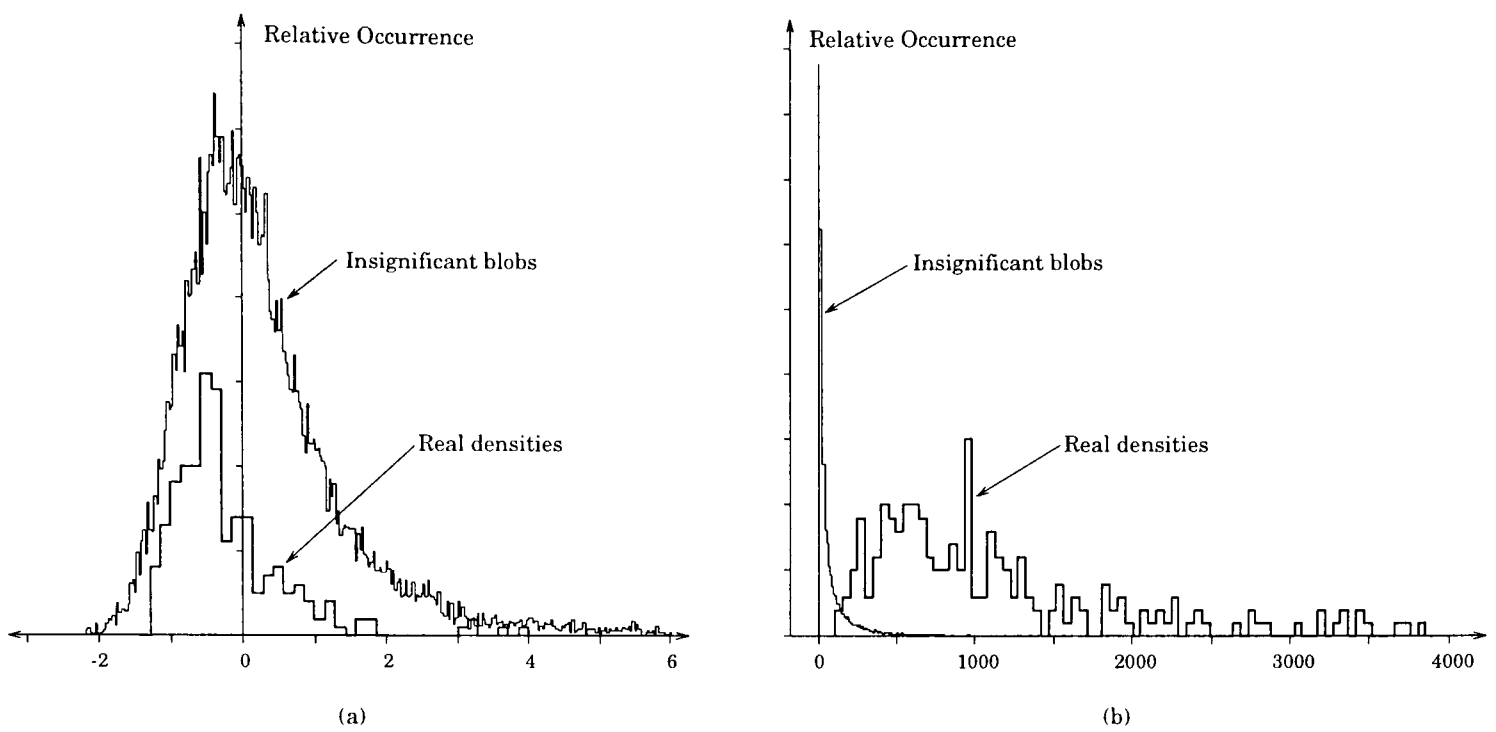


Figure 6.13: Distributions of the values of some descriptors for insignificant blobs (from 19281 data points) and real mammographic densities (from 232 data points) showing the relative discriminatory power of the descriptors. (a) The kurtosis of the grey-level intensities of the blobs, showing no ability to discriminate between the classes, whilst (b) the blob saliency metric $\mathcal{B}$ does show a significant difference between the classes.

descriptor displaying good separation[10] and one that does not.

Using this technique to assess the discriminatory power of a given descriptor allows us to select a data vector describing the blobs as the ensemble of the best descriptors. With this in mind, the process is undertaken by considering as many descriptors as possible that can be calculated and then selecting the most appropriate ones. Through experimentation we have arrived at a vector of five descriptors chosen to represent each blob, as displayed in figure 6.14. These descriptors were chosen from a longer list, which also included the blob support area, blob volume, blob perimeter, ratio of the area to perimeter, grey level statistics of the blob pixels in the CLS-removed image (average, skew, kurtosis, local and global contrast of the absolute image intensity), and the contrast between the blob and its surrounding region. These remaining descriptors did not show any ability on their own to discriminate between the set of known masses and known blobs, and were therefore not included in the novelty analysis. It should be noted that some of the descriptors displayed in figure 6.14 are closely correlated, a point to which we shall return in section 7.2.

Although the list of possible descriptors clearly does not exhaust the range of alternatives that might be considered, and thus there is scope for improvement, it is the demonstration (see next section) that even with this non-optimised data vector, the significant mammographic masses can be identified that is important. It is to this point of optimisation that we shall return in section 7.2.

[10] In fact this descriptor is the blob saliency metric $\mathcal{B}$ from section 6.3.1.

Descriptor	Explanation
σ^2	variance in the intensity of the blob's pixels
$\sum_S w(s)$	sum of the weights
$\mathcal{B}$	saliency metric, equation (6.5)
V/A	average blob height
$\mathcal{B}_3$	equation (6.4)

Figure 6.14: The descriptors chosen to encode the attributes and therefore represent each blob in the novelty analysis.

6.4.3 Implementation and results

With all the analysis techniques in place, there is nothing left to do but implement them. The major difficulty at this point is acquiring a suitable set of data upon which to conduct the analysis. We require in fact, a set for training, another for cross-validation and then a selection for testing.

Recently the *MIAS Mammographic Database* [204], containing 322 medial-lateral oblique images digitised at $50\mu m$ spatial resolution, became available, and it is mainly upon the images of this database that we have tested the techniques developed in the present work.

In order to extract suitable data sets for the novelty analysis, we initially reduced the images to a spatial resolution of $300\mu m$[11]. From the first 100 images only of the database (due to resource limitations) we obtained a CLS removed image in each case (see section 5.1) and then extracted all the blobs from each CLS-removed image (see section 6.3). With a description of the blobs from each image available, the instances of significant mammographic densities were manually removed to leave a remainder comprised of only insignificant blobs. In this fashion 19 281 known insignificant blobs were identified[12]. To reduce the computational overheads, but retain the statistical spread of data over the 100 images, the training data set was constructed by extracting every fourth blob from the 20 000 odd known insignificant blobs (i.e., the 1st, 5th, 9th, ... blobs). Similarly, the cross-validation data set was constructed by extracting every fourth blob, offset from the training data set by two (i.e., the 3rd, 7th, 11th, ...). Consequently the training and cross-validation data sets each comprised 4820 blobs, whilst 9640 were discarded. The segmentation of the significant potential mammographic densities from the images was performed manually by a resident radiologist. It should be noted that the 232 densities identified from these 100 images in this fashion include some actual real mammographic lesions and some other regions identified as potential candidates deemed interesting by the radiologist.

For each blob to be used in training, cross-validation or testing, the data vector $\mathbf{x}$ as shown in figure 6.14 must be calculated. Following assembly of all of this information, the implementation of the novelty analysis can proceed as described above.

[11] See chapter 4 for a discussion of the justification for this reduction in spatial image resolution.

[12] In fact a further 45 000 blobs have been identified from the remainder of the *MIAS Mammographic Database* however due to time constraints they have not been manually segmented into known classes of significant and insignificant blobs. Note however, that whilst these additional blobs cannot be used for training and cross-validation of the novelty system, they certainly can be used to determine if this novelty analysis approach is feasible.

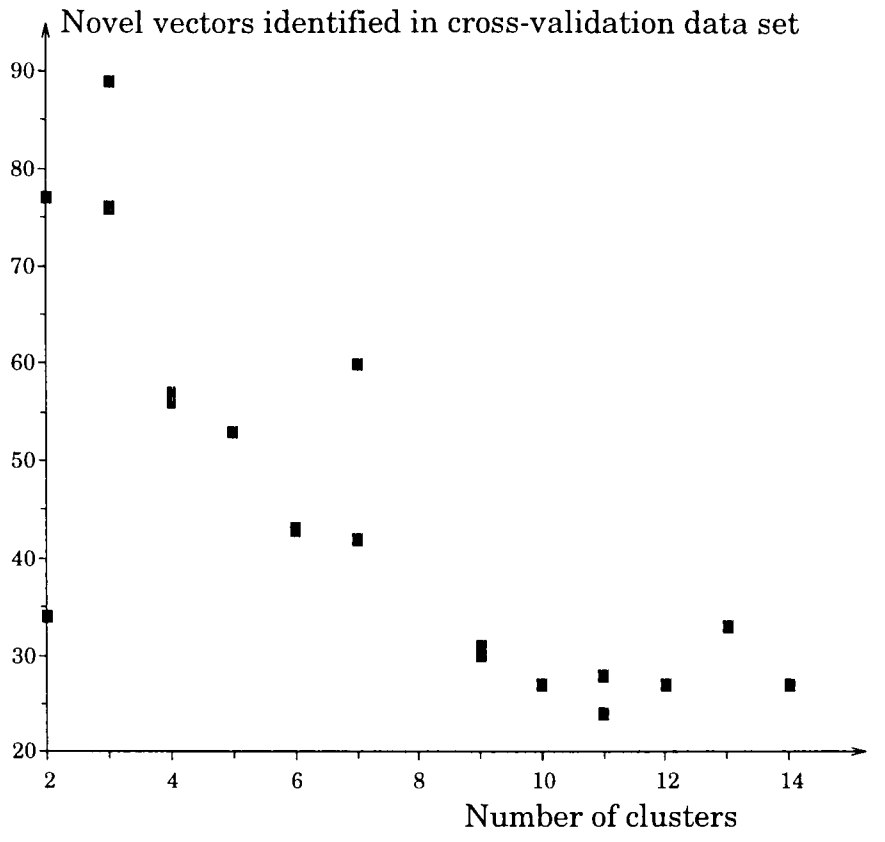


Figure 6.15: Novelty detection in the cross-validation data set as a function of the number of clusters. The optimal number of clusters describing the input feature space is given by the minimum rate of classification error in the cross-validation data set, which is known to be normal. In this example 11 clusters would be chosen.

In order to determine the optimal number of clusters to describe normality, the K-means algorithm is implemented for different values of K. The description of normality given by each is then tested by evaluating the novelty of each vector of the cross-validation data set. Since the vectors of this data set are known to be 'normal', we select the optimal description of normality from those available as the one that returns the *fewest* number of novel patterns from the cross-validation data set. For the purposes of this task, a threshold on normality for each cluster is set to be twice the cluster width (cf. setting the 'width' of a Gaussian distribution at twice the standard deviation of the distribution, or 2σ), and the novelty of each cross-validation vector is set as to be its minimum novelty relative to each cluster centre (determined as previously described). Figure 6.15 shows the number of incorrectly labelled cross-validation vectors versus the value of K. Too few cluster numbers are incapable of characterising the complexity of the pdf $p(\mathbf{x})$, leading to high numbers of errors, whilst too many clusters begin to characterise what outliers may be in the training data set. Somewhere between these two extremes lies the optimal value, corresponding to the minimum of figure 6.15 at 11 clusters.

With the clusters quantifying the regions of normality within the feature space identified, the evaluation of test vectors can proceed directly. Figure 6.16 shows the distribution of relative novelty values for the known significant densities against the distribution for the known insignificant blobs. Observe that whilst the separation between the distributions for significant and insignificant blobs is very good, it is not perfect. This is a reflection of two points: (a) some of the data grouped into the significant data set is borderline, in the sense that

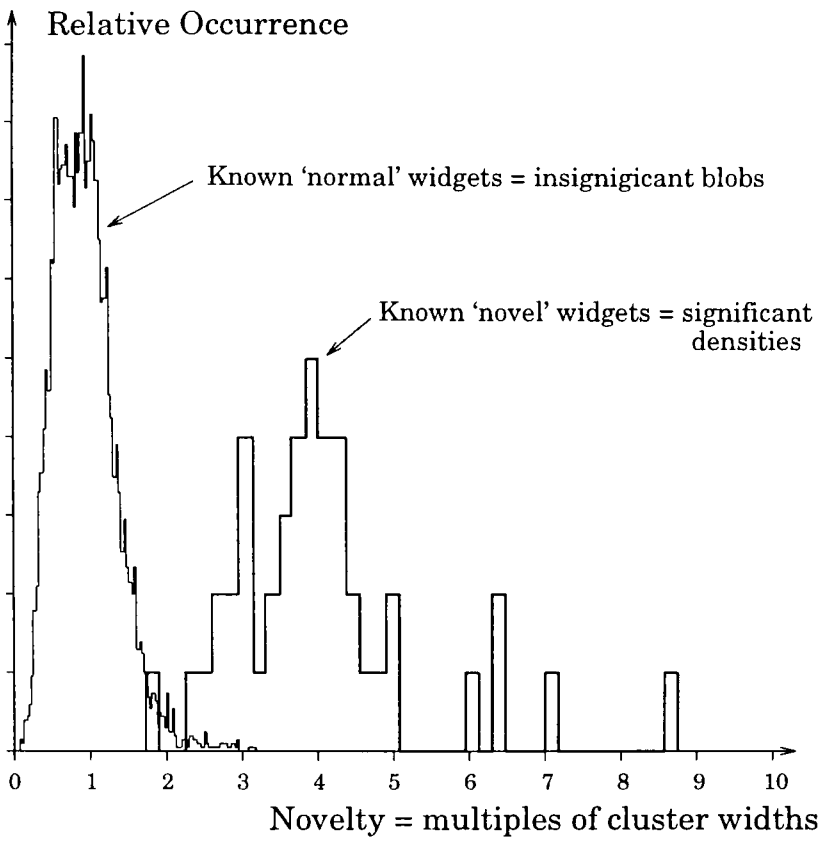


Figure 6.16: Distributions of novelty values for known significant densities, and known insignificant blobs.

even the human operator would be unsure as to the accurate labelling of such a blob; and (b) as has been previously mentioned, the selection of the *best* descriptors for encoding the attributes of a blob has not been addressed. Such optimisation has the potential to improve these results.

As a further vivid demonstration of the novelty analysis technique, consider the analysis of all the blobs—both significant and insignificant—of an image in search of those that are significant. Figure 6.17 displays the novelty values calculated for each blob as the colour scheme overlayed on the image, and the scale given below the images. Compare this figure with the original images of figure 2.7. Clearly the large cyst in the upper quadrant of the left breast has been well identified. Figure 6.18 similarly displays the novelty values calculated for the blobs identified in the original images of figure 4.32.

6.5 Conclusions

The identification of mammographically significant densities is an important component of any attempt to automate the analysis of mammograms, yet the task is complicated by the vague definition of the attributes of such densities. In fact there is no uniform definition as such, and clinicians and automated systems alike have been forced to develop a description of "significance" from the experience of exposure to many examples that span the range of possibilities.

In recent times, the standard statistical tools available for encoding such 'experience' into an automated system have often been those of the neural classifier. Such classifiers depend upon the analysis of *equal* numbers of the various object classes to be identified so that the

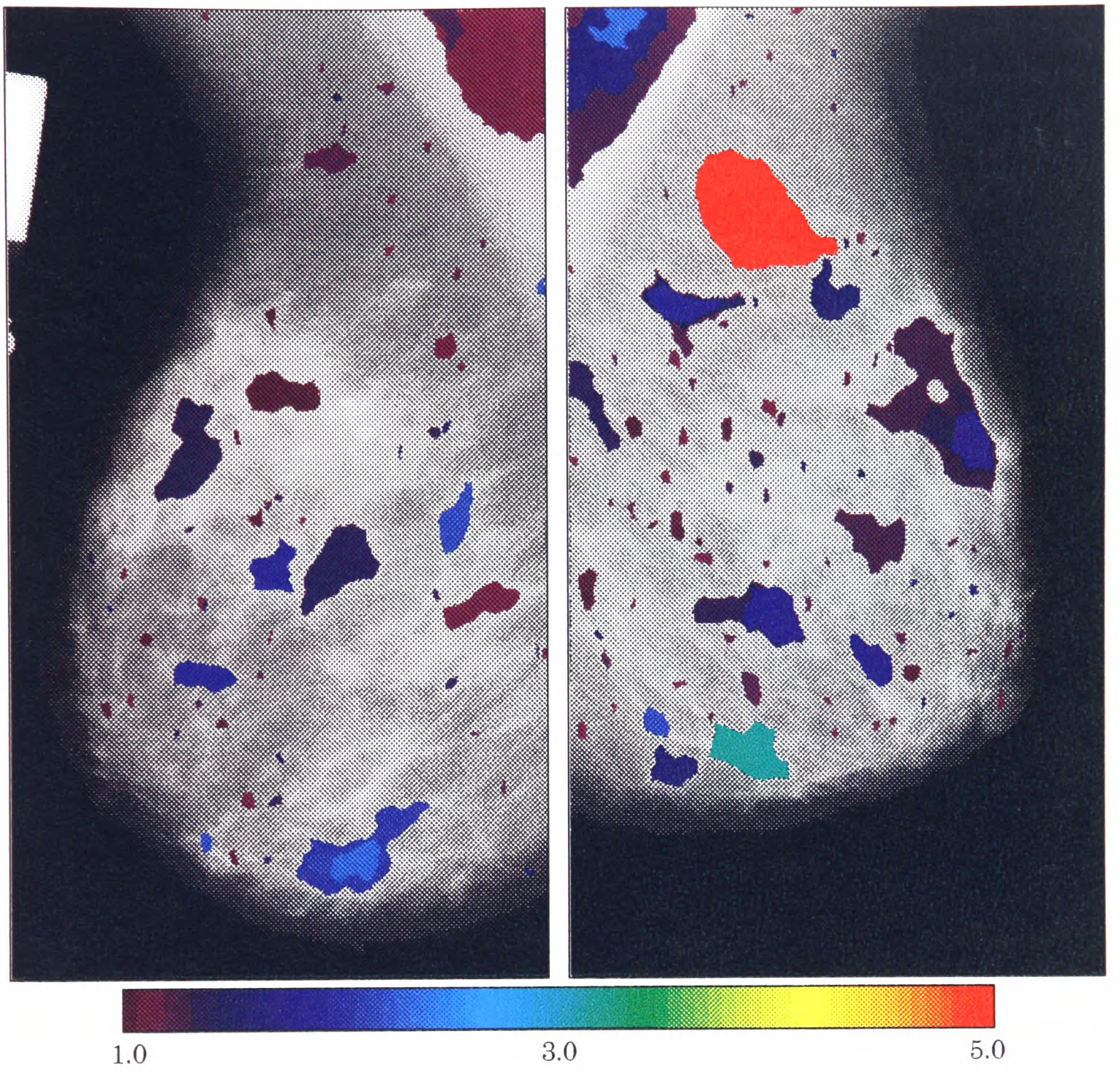


Figure 6.17: Significant densities identified by novelty analysis in the images of figure 2.7 from the blobs of figure 6.9. The colour scale indicates the relative novelty of each density, where the threshold of 1.0 is set to the average radius of each cluster's associated data vectors. Only densities with novelties exceeding the threshold are displayed, and figure 6.9 should be consulted for identification of those blobs not displayed here. Observe that a number of insignificant blobs have been shaded with relatively minor novelty values, a reflection of the fact that a novelty of 1.0 is set as the *average* of the training data patterns associated with a cluster, and thus a distribution of normal patterns with novelties in excess of 1.0 is to be expected. Alternatively, a threshold of 2.0, as discussed in the text, could have been used.

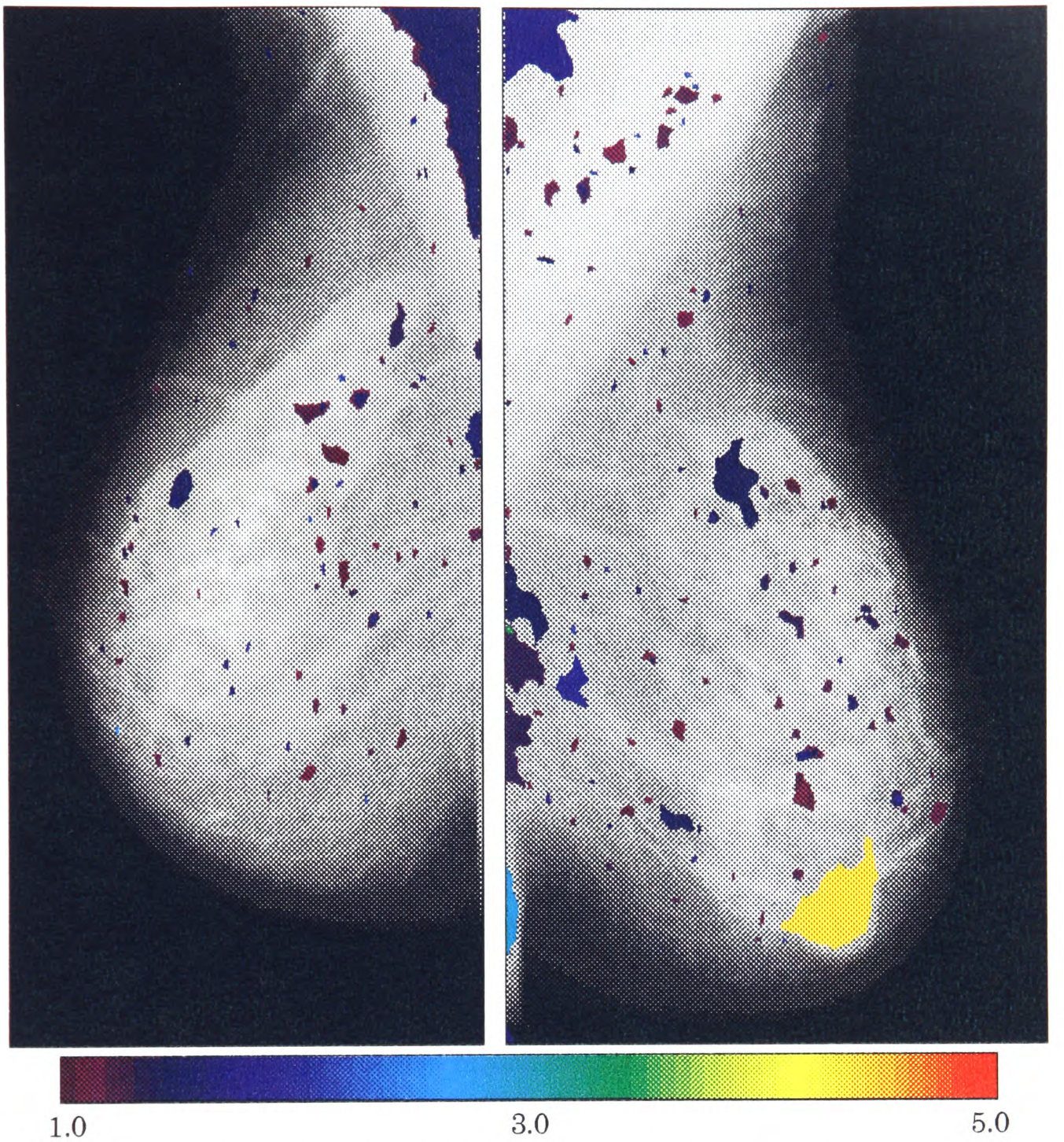


Figure 6.18: Significant densities identified by novelty analysis in the images of figure 4.32 from the blobs of figure 6.10. See the caption of figure 6.17 for more information.

decision boundaries so calculated are not biased by differences in a priori probabilities. The exposure of both clinicians and automated methods to enough cases of abnormalities spanning the range of possibilities is problematic in the sense that there is no known bound to the range of possibilities. Furthermore, so protean are the attributes of the known abnormalities that large numbers of examples are necessary to characterise them adequately. In combination these constraints reduce the effectiveness and quality of the training of a classical neural classifier to a state of worthlessness, and the results obtained from these methods reflect that state.

In order to identify the significant densities in an image we developed a 'blob-detector' capable of identifying regions in a mammogram image fitting the model of a blob given in figure 6.1. Building upon the results of chapters 4 and 5 we can transform a given mammogram image into a CLS-removed image. This transformed image possesses image intensity attributes at locations of mammographic densities that in general match the definition of a blob used in the formulation of the blob-detector of this chapter. Thus the blobs of an original image are identified by analysing the CLS-removed image with the blob-detector. A consequence of this approach is that the blob-detector identifies the most *locally* salient blobs over all regions of the image. In this fashion many blobs in excess of the mammographically significant densities are identified, requiring their subsequent segmentation. Such a segmentation falls into the statistical classification problems mentioned above, and it is therefore bound by those limitations also.

To circumvent these difficulties we employed recent developments in novelty analysis. By learning a description of the mammographically insignificant blobs, of which we have many tens of thousands of examples available, we can identify a significant mammographic density by testing for its *novelty* against that description of *normality*. Unconstrained by the limitations demanding equal numbers of insignificant blobs and significant densities, we are free to train on as many examples of insignificant blobs as are available, and can thus ensure a much more statistically significant result in comparison to the standard classification schemes.

At this juncture the analysis can diverge along any of a great many paths concerning the choices of attributes to quantify and describe each blob, and the specific details of the algorithm chosen to 'learn' the description of normality. Although we develop a particular implementation of the procedure in this chapter, it is the demonstration that at least *one* such implementation is capable of identifying the significant densities that is important, for it is the solution philosophy of novelty analysis that should be considered as one of the main results of this chapter. Further to this point it should be noted that of the descriptors chosen to quantify the attributes of each blob (see figure 6.14) in the current implementation, not one of them encodes a 'global' measure of the environment within which the blob resides, nor is there a measure of the blob shape or boundary roughness. Clearly such measures may have the capacity to identify other significant attributes and improve the segmentation, and in an exhaustive analysis of this technique these points would need to be considered. Nonetheless, given the data vector of figure 6.14 we have demonstrated the feasibility of novelty analysis techniques for the task of

identifying the significant mammographic densities.

It should be noted that the assessment of the novelty analysis, whilst trained on many thousands of insignificant blobs, is, like the Bayesian techniques reviewed previously, limited by the small quantity of data describing the significant potential densities (232 cases only in this study). As with the Bayesian techniques, acquisition of more data would allow a more detailed and significant analysis to be made. Observe however that a key component of the novelty analysis over the standard Bayesian classifiers has been the ability to include the enormous quantity of data available describing the normal/insignificant blobs, and therefore, this has improved the statistical significance of this component of the analysis.

7

Conclusions and Future Work

7.1 Conclusions

In this thesis we have developed a model of the curvi-linear structures (CLS) appearing in mammogram images, right from the anatomy of the breast, through the physics of mammography and digitisation (the image forming process) to the grey-level intensities of the image. This model has allowed us to undertake a range of subsequent image processing tasks with confidence that the results obtained can be interpreted in terms of the physical reality of breast anatomy. It is of course within this domain that breast disease occurs, and we can therefore proceed with confidence that the algorithms based on this model are capable of extracting information that is truly significant in the context of breast anatomy and more importantly, breast disease. To our knowledge, this thesis represents the first work to both seek and achieve a description of the CLS features per se, and to subsequently accommodate their degrading and complicating effects in a logical, systematic and quantitative fashion. Furthermore the feature description itself leads to new applications and opportunities for analysing mammograms that have never before been possible.

Except for the fibrous spiculations associated with spiculated lesions and radial scar, the CLS features are not themselves signs of breast disease, despite the fact that it is at the site of many of the tissues comprising these structures that many forms of breast disease do develop. Due to their narrow width and high spatial frequency, the textural clutter they directly produce complicates the search for mammographic densities and the assessment of bilateral asymmetry. This complication arises since although densities and asymmetric processes are generally of larger spatial extent and often occur at lower spatial frequencies, there is important diagnostic information in some of their high frequency components (for instance the edges of the densities), and consequently low-pass filtering cannot be used to remove the high-frequency

CLS features. By identifying the CLS features in an image we are able to selectively remove the high frequency image features due to their presence whilst retaining those due to other processes. Previous attempts to accommodate these difficulties have used quickly conceived ad hoc approaches often without direct or logical links to the physical realities of the breast anatomy they seek to describe.

We have developed a model of a potential mammographic density as a region of higher image intensity resulting from the increased attenuation of the X-ray beam as it traverses a physical breast mass. By removing the CLS features from an image, we have been able to develop, and realise an implementation of, an algorithm to identify the locally salient regions of higher image intensity, some of which therefore, correspond to mammographic densities. To segment the significant densities from these many regions, this thesis describes the first application of a new statistical classification technique, that of *novelty analysis*, to the automated analysis of mammograms (and the second application of the technique ever![1]).

The success of the algorithms developed in this work and the results they afford are directly attributable to the careful and accurate modelling of the physical anatomy and disease processes of the breast. The models so developed are the direct links between the algorithms and the realities they seek to describe. A consequence of developing the algorithms from the basis of these models is that we can not only interpret the results they give in a systematic and logical way, relating those results back to statements about the physical anatomy that a given mammogram represents, but more importantly, we can *believe* those results and subsequent interpretations. This confidence in the results of the algorithms is a key requirement of any automated analysis system. Recall from the introductory comments to chapter 3 that:

> "... formulating an automated system based upon sound principles, enables a logical explanation of results produced, and is therefore important in developing the trust of the radiology community for those system capabilities that are not directly comparable to their experience. A system *perceived* (not necessarily correctly) to inadequately or inaccurately analyse a mammogram is of no use to the clinician or the community at large. Conversely an inaccurate system perceived to be correct, is downright dangerous and must be avoided."

The work presented in this thesis is based on the sound principles of model-based analysis, and gives results that can be directly interpreted in terms of the underlying breast anatomy and disease processes. It solves the dilemmas associated with the curvi-linear structures, develops a new technique for detecting mammographic densities, and presents the application of a new statistical analysis technique that overcomes the difficulties of insufficient sample space that have previously plagued this field. Furthermore the high-level feature descriptions of the CLS features and the densities can be applied to many alternative endeavours, further aiding the development of automated mammogram interpretation systems.

[1]Details of the first application of novelty analysis (to the assessment of sleep disorders) can be found in [173].

7.2 Future Work

The material presented in this thesis can be extended in two natural directions: direct improvements of the algorithms presented here; and alternative applications of those algorithms and the results they give. Occasionally these extensions are inter-related, and we therefore consider them together as appropriate.

As noted in section 4.4.8, the selection of a single second-difference kernel scale to facilitate the determination of the image surface curvature does impose a limit on the maximum width of a CLS feature that can be detected. This fact was exposed in figure 4.31(b) on page 109 which illustrates an incomplete identification of a wide surface vein in the image of figure 4.30(b). It is further highlighted in the SAR images of figures 5.9 and 5.10 (pages 134 and 136) at the intersection of the roads towards the bottom right corner of the image. As noted previously, a full scale-space treatment is necessary to accommodate these cases, such as Fleck [67], which describes a multi-scale approach to the implementation of finite-difference filters which would naturally apply in this situation. Whilst not strictly difficult, it is noted that the additional requirement of combining the difference information from different directions (section 4.4.9) is not strictly trivial either, and this must be carefully considered and successfully implemented to achieve a multi-scale implementation of the CLS detector algorithm.

Successful identification of CLS features of any width would naturally improve the CLS-removal procedure of chapter 5. Whilst not adversely effecting the results of the blob and subsequent significant density identification procedures of sections 6.3 and6.4, it is acknowledged that these procedures would also benefit from such improvements in the identification of all CLS features.

Other limitations of the current design and implementation of the CLS detector algorithm should be addressed by further work. The construction or acquisition of a suitable phantom with a priori known locations of the CLS features would facilitate an objective assessment of the accuracy of the CLS detector algorithm, as opposed to the stability analysis presented in this thesis. Images of such a phantom could also be used to determine the best image spatial resolution to undertake the search for CLS features. This could be achieved by reducing an original image of the phantom to a number of different resolutions, reformulating the CLS difference kernels for that scale, and comparing the results with the known locations of the features. This type of analysis could also be used to identify the most suitable image intensity resolution by digitising the original phantom image at a number of different intensity depths and again comparing the CLS features identified in each image with the known results.

As noted in section 4.8 the CLS detector responds to all regions of negative image surface curvature, including such regions associated with edges rather than CLS features. Whilst this difficulty was easily addressed during the CLS-removal procedure described in section 5.1, in general it would be more appropriate to deal with these features during the CLS detector itself. Perhaps the implementation of third difference filters could be used to identify the double curvature inflections at either side of an identified CLS feature to establish that in fact it was

a real CLS feature. In this instance the lack of a curvature inflection on one side of the feature would signal it as simply an edge of a larger feature.

Although the details of the model of a CLS feature developed in section 4.4.1 that precipitated the CLS feature detector are definitely application specific, the final specifications of that detector (that the features be nominally one dimensional, locally linear, and have negative surface curvature) are relatively generic, and can therefore accommodate many other applications. Clearly the extensions of these techniques to other applications, such as the identification of roads in SAR images, is limited only by one's imagination.

Some spiculated masses located in particularly dense glandular regions may only be mammographically identifiable from the spiculations surrounding the mass. Similarly, due to the lack of a radiodense centre, the spiculations of a radial scar process are the only indicators of its presence. There have been very few attempts to identify these processes since identifying the spiculations is a problem that has not in general been tackled by the literature, and thus there has not been a suitable domain within which to identify these features. The notable (only?) exception to this generalisation has been the recent work of Karssemeijer [110] which assembled a pixel-based assessment of local line directions at each location into a measure of the radial spiculations surrounding that location. As mentioned in section 5.2, this technique could be greatly enhanced by directing the search from the domain of the CLS feature description, rather than the entire image, since the CLS feature description includes information on CLS direction.

The novelty analysis techniques of section 6.4 are used to segment the list of blobs previously identified into mammographically significant and insignificant ones. Obviously, this can approach can only identify the significant densities if the blob detector identifies them at the first stage. For completeness it should be noted that whilst the blob-detector as formulated in the present work is capable of identifying all locally salient blobs fitting the model of a blob given in figure 6.1, and therefore nearly all the blobs of a mammographic image, there is a situation (and its variants) that this model of a blob cannot adequately describe. Consider a salient density, a cyst say, in the fatty region of a breast peripheral to a dense region of parenchymal tissue. Imagine that the majority of the cyst is located in the fatty region, and is therefore quite salient, whilst a small portion is occluded by the extremities of the parenchymal tissues. Furthermore imagine that the intensity of the pixels of the cyst are lower than those in the region of the parenchymal tissue (for clarity imagine each region is of a uniform intensity). Such a situation is not described by the model of a blob used in this work, and consequently the blob-detector will never identify this region as a salient blob. Such blobs could be identified in a number of fashions, such as by using dynamic contours (snakes). Clearly any attempt to fully automate the analysis of mammographic images would need to address these limitations.

As noted in section 6.5 the extension and optimisation of the feature vector describing each blob used in the novelty analysis provides a great deal of scope for improvements in that analysis. It has been noted that the clustering in feature space of the blobs is particularly

dominated by relatively few clusters accounting for a very large proportion of the training data, with the remainder of the data dispersed over a disproportionately large number of clusters. There is a need to return to the data and determine which blobs have given rise to the highly probable clusters and which have been dispersed over the remainder of clusters. It is expected that further optimisation of the feature vector describing each blob will result from this analysis. It should be noted that this work is in fact currently being undertaken [90]. Note that the question of how to determine the optimal feature vector describing each blob is one for which there is no answer. It is through a thorough understanding of the data and the operation of the novelty analysis that well-informed and logical decisions can be made. Observe that the feature vector used in the present work includes measures of the bounding contour edge strength and the grey level intensities of the pixels contained therein. As an example, an obvious extension of this work is to include descriptors encoding a measure of the blob shape. The continued development of suitable feature descriptors is an area that would considerably improve the results of the novelty analysis presented in this work.

As presented in chapter 6, novelty analysis is used to identify the relatively few significant potential densities from the many insignificant blobs identified by the blob detector. Whilst utilising thousands of data points to capture the statistics of the normal/insignificant blobs during the training phase of the analysis, the ability of the technique to correctly identify significant potential densities has been evaluated over only a small number of cases (232). Further data should be acquired in order to make the assessment more significant.

Recall from section 3.6 some authors attempt to determine the existence of densities from an assessment of bilateral asymmetry [73, 231, 232], rather than assess densities in some other approach and then use that feature description to drive the search/assessment of bilateral asymmetry. Asymmetry is, after all, not confined simply to densities, but distributed over broader parenchymal processes. A consequence of this broad distribution, and the uncertain and vague terms in which 'asymmetry' is described, is that the task of automating the assessment of asymmetry is a very difficult one, and in general this has not been undertaken. Consider using the blob and density description evaluated in chapter 6 and possibly even the CLS description when assessing temporal change in a single breast, as the basis for a comparison of higher-level features in the assessment of asymmetry. This approach allows the uncalibrated and non-normalised absolute image intensities typical of any pixel-based method to be disassociated from the analysis, and instead, base it on knowledge of the glandular structure interpreted from these higher-level feature descriptions. Clearly image registration is an important method here and has been tackled in a variety of forms in the literature [24, 73, 131, 180, 198]. Unlike the pixel-based applications of these registration approaches it is not imperative to achieve a perfect alignment of the breast structure when assessing asymmetry from the higher-level feature descriptions as proposed here. In agreement with Szeliski [205] we have found a quadratic transformation (12 free parameters) to be a good compromise between flexibility and complexity as it has sufficient degrees of freedom

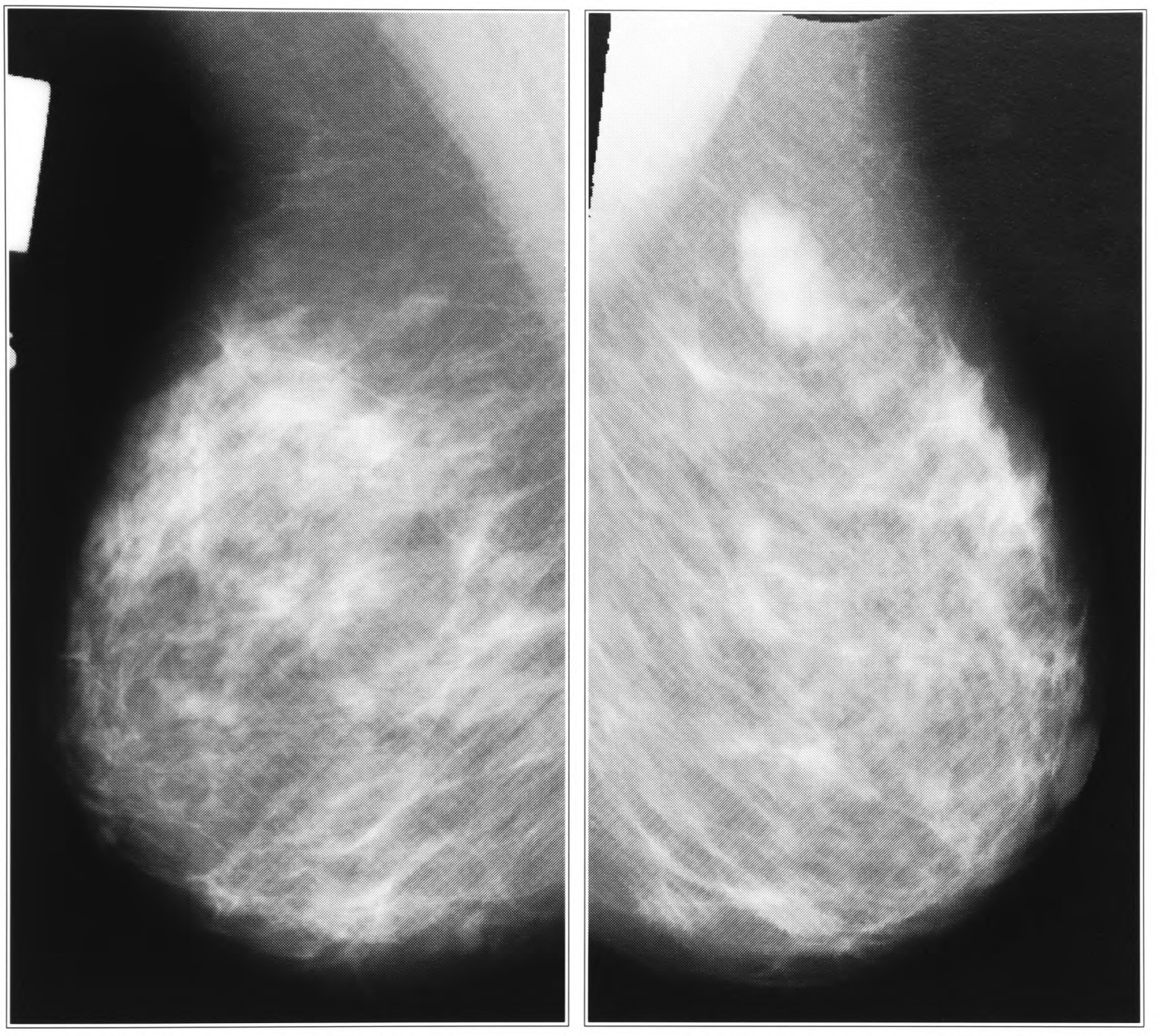


Figure 7.1: The image on the right has been warped with a quadratic transformation to give a rough match between the original images of figure 2.7, in preparation for an assessment of bilateral asymmetry. Note that a perfect matching of the images is not required (and is indeed not achieved—for instance at the very top of the image) since the high-level feature descriptions of the CLS and density structures allows bilateral comparison at this higher-level, rather than at a pixel-based level.

to accommodate the necessary deformations of an image to achieve alignment, yet retains an acceptable computational expense. As an example of this consider the alignment of the images of figure 2.7 by a quadratic transformation, as shown in figure 7.1.

The forward transformation that allowed the deformation resulting in the right-hand image of figure 7.1 is calculated by simple least-squares fit of the known fiducials to the 12 parameters of the quadratic transformation. Acquiring the inverse transformation is more problematic as it is strictly a root-solving process for *every* pixel of an image, a process that is not guaranteed to converge given poor initialisation. We have found that the results of calculating a least-squares *inverse* transformation of the fiducials gives an approximation to the real solution, and when used to initialise the root-solving process, allows it to converge very rapidly in every case. It is

our experience that the process of finding an accurate inverse transformation is realisable and executes in reasonable time. This problem, therefore, should not limit any attempt to analyse asymmetry. With the image registration in place, the question then is to decide how to assess the vague concept of asymmetry. This is the open question to which the CLS and blob/mass feature descriptions could be applied. Solutions could possibly adopt a regional correspondence measure based on 'feature-activity'. For instance, a measurement of blob density weighted by each blob's corresponding novelty value might be considered. The possibilities here are endless, and would naturally reflect the choice of model adopted to describe the parenchymal tissues. Irrespective of the model chosen, the CLS and blob/mass feature descriptions could certainly provide information suitable for this analysis.

As noted in section 5.3 the CLS feature space could be used to develop a dense correspondence mapping between the differential compression mammogram images described by Highnam and colleagues [97, 99].

The CLS-removed images have already found another applications in the textural analysis of the breast parenchymal tissues by Kok et al. [124]. In this work the CLS-removed image is used to assess the textural patterns of the glandular tissues of the parenchyma without the cluttering effects of the CLS features which are normally present in an image.

In other continuing collaborative work, the CLS feature descriptions are being combined with the mass descriptions extracted in chapter 6 to develop a model of spiculated masses which includes both the central density and the spiculations. This technique has initially been applied to the synthesis of artificial lesions for use in the Xmammo mammographic image processing system [97]. The principle purpose of this application is to function as a training tool to assist the development of clinicians training to read and interpret mammograms. Within the realms of automated analysis, it is anticipated that the further development of this model will lead to a spiculated mass detector, drawing on the original image, the CLS description, and the CLS-removed image to accomplish the pattern recognition task.

A

Region thinning

The CLS detector algorithm of section 4.3 is formulated such that the extraction of a higher-level description of the CLS features from the regions of CLS pixels identified in section 4.4.10 is achieved by analysing a topologically equivalent, skeletal representation of those regions. It is the algorithm for extracting the skeletal representation of the binary input regions that is developed in this appendix.

We begin by establishing the desired requirements of the algorithm, during which the nomenclature will also be specified. This is followed by a review of a common algorithm, and the observation of its shortcomings with regard to the requirements of the current application. The following section develops an extension to the algorithm that defeats those shortcomings, allowing the CLS algorithm to proceed as specified in section 4.3.

A.1 Thinning algorithm requirements

In this section we specify the desired characteristics of the thinning algorithm on the basis of the properties of the skeleton that is required to sustain the subsequent processing of the CLS detector algorithm. The concepts are presented in order of decreasing importance and precedence, and therefore a given requirement is subordinate to those that have been previously established.

To make the extraction of a higher-level description of the CLS features as easy as possible, it is required that the skeletal representation of the CLS pixel regions be as simple as possible. To avoid many difficulties during subsequent interpretation, we do not desire the location of the skeleton to be determined to sub-pixel accuracy. On the contrary, we *require* the skeleton to be specified to pixel accuracy, such that if the true location of the skeleton lies between two pixels, then its position should be specified as at the neighbouring pixels. This specification of

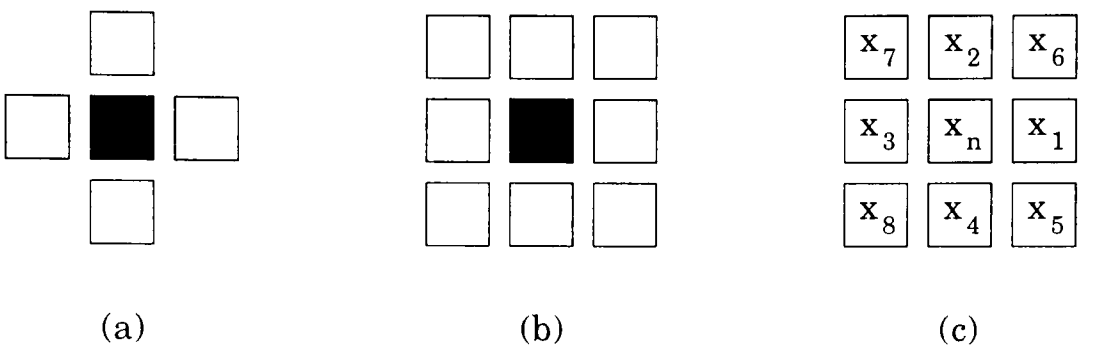


Figure A.1: Pixel conventions. (a) the 4-connected neighbourhood of the central (black) pixel, (b) the 8-connected neighbourhood, and (c) the labelling conventions for pixels surrounding a pixel x_n.

its location is deliberately worded rather loosely at this stage, since it impinges on questions of topology preservation which are to be considered shortly. The important point to note at this stage is that:

Requisite A.1 *The format of the skeleton should be a list of pixels describing those pixels that lie on the skeleton (identified to pixel accuracy).*

We require the high-level CLS description to preserve the topology of the original CLS pixel regions, and therefore since the extraction of the high-level CLS feature description is driven from an analysis of the skeletal representation of those regions, we also require that that representation preserve the topology of the original regions.

Requisite A.2 *The skeleton must preserve the topology of the input regions.*

Excepting the border pixels of an image, each pixel has four horizontal or vertical neighbours, and four diagonal neighbours. A pixel is 4-connected with its neighbourhood if can recognise/acknowledge/respond-to the existence of the four horizontal or vertical neighbours, and is 8-connected with its neighbourhood if it can recognise/acknowledge/respond-to all eight neighbours surrounding it. Figure A.1(a) and (b) illustrate the 4-connected and 8-connected neighbourhoods surrounding a pixel.

Consider a two-dimensional binary image plane with *foreground* pixels $\mathcal{F}$, and *background* pixels $\mathcal{G}$. To preserve region topology within the rectangular tesselation of our images, we allow foreground pixels to be 8-connected and background pixels to be 4-connected (see [65, 104] for further information).

Consider two foreground pixels A and B, within the binary image plane. A *path* $\mathcal{P}$ linking A and B is a sequence of 8-connected foreground pixels including A and B at either ends of the sequence. In situations where a foreground pixel has both a vertical/horizontal foreground neighbour and an adjacent diagonal foreground neighbour, the vertical/horizontal neighbours take precedence over the diagonal neighbours. Consequently the sequence of pixels for such a path P would pass from the current pixel to the horizontal/vertical neighbour and then to the pixel that was the diagonal neighbour (but is a vertical/horizontal neighbour of the current pixel).

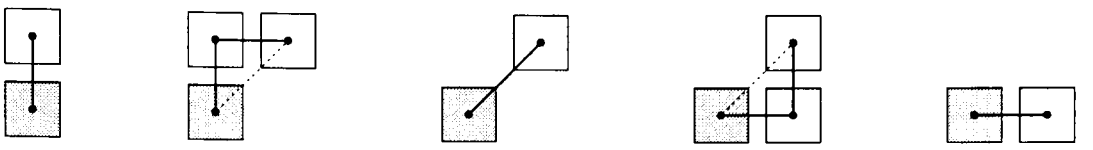
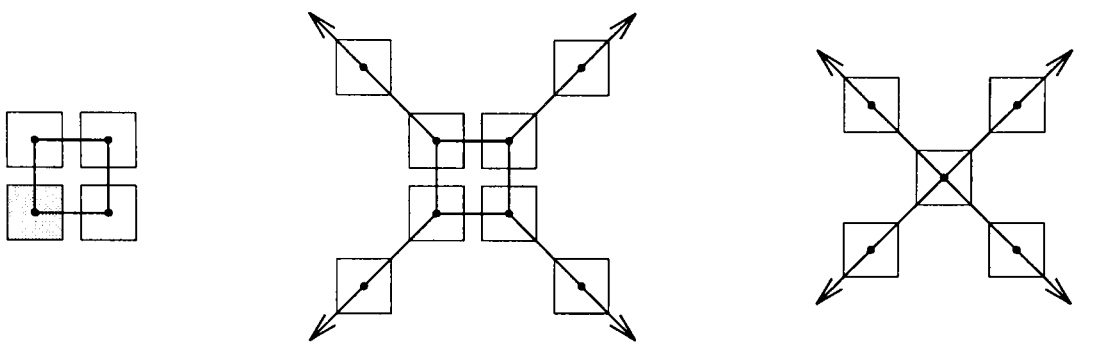


Figure A.2: Connectivity precedence for the top-right quadrant of the neighbourhood surrounding a central (shaded) pixel. The solid lines joining the dots at the centre of each pixel indicate the skeletal connectivity adopted in this work. Not the two instances where both a horizontal/vertical and a diagonal neighbour exist, illustrating the precedence of 4-connectivity over 8-connectivity. The dashed path directly joining the diagonal 8-connected neighbours is rejected in favour of the two 4-connected paths shown. By symmetry the remaining three quadrants behave analogously.



(a) (b) (c)

Figure A.3: Complex skeletal connections. (a) Any 2×2 block of pixels is complex connected with two paths joining diagonally opposite pixels. (b) A situation in which complex-connectivity cannot be avoided in order to preserve topology. The arrows indicate that the regions extend off in some direction roughly indicated by the direction of the arrow. (c) A simply-connected skeleton is acquired if the pixel arrangement is fortuitous to have a central block of odd dimension.

Definition A.1 *Foreground 4-connectivity takes precedence over foreground 8-connectivity.*

This situation is illustrated in figure A.2. With this definition then, consider the situation in which all three pixels of a single quadrant in the neighbouring region of a foreground pixel are also foreground pixels. Figure A.3(a) illustrates such a situation and the corresponding connectivity paths are shown as the solid line overlayed. Note that between, say the lower left pixel (grey colour) and the top right pixel, there are *two* paths joining these endpoints, assuming as we do, that 4-connectivity takes precedence over 8-connectivity.

The formulation of the CLS detector algorithm of section 4.3 aims to extract a single entity (a canton) describing each primitive segment of the CLS pixel regions. Each canton is directly extracted from a segment of the skeletal representation of the CLS pixel regions, and we therefore require that a single segment of the skeletal representation should describe each segment of the CLS pixel regions. Since the segments of the skeletal region are found by breaking the skeleton at each and every junction (see section 4.6) this requirement is equivalent to insisting that a *single* path P traverse each topologically unique segment of the CLS pixel regions. Adherence to this requirement returns a *simply* connected skeleton.

Definition A.2 *A simply-connected skeleton contains only a single path traversing each topologically unique segment of the foreground regions.*

Requisite A.3 *Subject to the preservation of topology, the skeleton of the CLS regions must be simply-connected.*

The definition of a simply-connected skeleton can be expressed more formally as follows. Consider the pixel labelling for a given pixel's neighbourhood as shown in figure A.1(c). Then *every* pixel of a simply-connected skeleton must yield $C(x_n) = 0$ for the following set of equations:

$$\begin{aligned} H(a,b,c) &= \begin{cases} 1 & if(a = \mathcal{F} \text{ and } b = \mathcal{F} \text{ and } c = \mathcal{F}), \\ 0 & \text{otherwise,} \end{cases} \\ h_1 &= H(x_1, x_2, x_6) \\ h_2 &= H(x_2, x_3, x_7) \\ h_3 &= H(x_3, x_4, x_8) \\ h_4 &= H(x_4, x_1, x_5) \\ C(x_n) &= \sum_{1}^{4} h_i \end{aligned}$$

Compliance with this set of equations at each pixel is equivalent to ensuring that situations such as the illustration of figure A.3(a) do not arise. There is however, a situation in which it is impossible to achieve *both* a topologically equivalent and simply-connected skeleton. Consider the arrangement of skeletal pixels shown in figure A.3(b). Observe that the central group of four pixels comprises a complex-connection (implicitly defined in the logical sense to be 'not simply-connected'). Note however, that removal of any one or more of the central pixels would break the topology of the overall figure by splitting the single 8-connected foreground figure into two (or more) separate 8-connected foreground regions. Since it is imperative that the topology of the skeleton reflect that of the original figure, it is acceptable for the skeleton to be complex-connected in these situations.

Clearly, it is desirable to obtain a skeletal representation that describes the medial-axis of each segment of the CLS pixel regions, that is, the locus of points equidistant from the boundary on either side of the current segment. We are of course, constrained by the requirements of a simply-connected skeleton. Consider the situation of a horizontal strip of pixels, of even width, say two pixels. The medial-axis of this figure obviously lies mid-way between the central two rows of pixels. In such a scenario we are free to choose either row of pixels, or indeed a mix of pixels from either raw, as the skeletal representation of that region, provided the constraints of simple-connectivity are observed. This is the case in all situations, and it allows us to require that:

Requisite A.4 *The skeleton shall approximate to pixel accuracy the medial-axis of the original figure.*

In conclusion then, we require a skeletal representation of the input regions consisting of a list of pixels, which preserves the topology of the input regions, is simply-connected where possible, and lies to within a pixel of the medial-axis of the input regions.

A.2 Related work in region thinning

There are a number of different ways to identify the skeleton of a given region. Much of the early work in this field arose from the attempts to obtain a minimal representation, or primal-sketch of a feature by identifying its symmetry-set. Blum [14, 15] set the wheels in motion with the introduction of the Medial-axis transformation (MAT) (or symmetric axis transform (SAT)). There have been many developments and modifications to these approaches [7, 19, 20, 22, 64, 167] and more recently extensions to parallel implementations [182, 214]. The many variants of these concepts (symmetric axis (SAT/MAT), smoothed local symmetries (SLS), local rotation symmetries (LRS), process inferring symmetry analysis (PISA), etc.) all implement a different philosophy of what constitutes the minimal representation of a figure. Since the skeleton of a figure is generally a subset of the information comprising its minimal representation, considerable use has been made of this work in developing dedicated skeletonisation algorithms. Other skeletonisation algorithms have developed more recently from mathematical morphology [143, 142, 162], and dynamic/deformable contours [136].

The selection of an algorithm to achieve the tasks specified above for the present application is a problematic one in the sense that there are more options amongst the plethora of algorithms related to figure symmetry/medial-axis/skeletonisation to choose from than there are blends of tea to select from in the morning. The difficulty arising from this situation, and indeed the reason behind the diversity of such ideas and algorithms, is that the final implementation of any algorithm naturally reflects the axioms and initial specifications established prior to its development. That is, these algorithms tend to be highly application specific, and the success or failure of an algorithm can only be measured by the relative ability of that algorithm to meet the stated specifications.

Lam et al. [130] provide a comprehensive review of the published literature concerning these algorithms, concluding that in general, the requirements that should be met by a thinning algorithm "include preservation of topological and geometric properties, isotropy, reconstructibility and high processing speed". They also conclude "that comparison of the quality of skeletons remains a largely subjective, visual decision", and the choice of algorithm should depend on the application at hand.

The published implementations can generally be broken into two main groups: (a) thinning, and (b) distance transform methods. Following Blum's 'grassfire' methodologies (see below), much of the symmetric-axis based work falls into the thinning class of operation. Distance transform methods involve determining the distance from each pixel of a region to the background (where the 'distance' metric has been defined in many different ways) and then identifying the ridge of pixels with the locally largest distance. As previously mentioned, the

implementation details of these methods reflect the specific requirements of each task at hand. Those attempts to develop 'generic' algorithms often find little or no practical use since the assumptions do not precisely match the particular conditions of the task at hand, and so it is in the present situation. We begin by analysing a common 'generic' thinning procedure, identifying its weaknesses, and then proceed to develop the necessary extensions to that algorithm to comply with the requirements of the present situation.

A.3 Haralick & Shapiro's thinning algorithm

Haralick and Shapiro [86] describe an iterative thinning procedure that falls into the class of thinning operations commonly called the 'grassfire' algorithms. Contemplate for a moment, the somewhat more realistic case of a field of sugar-cane. Considering the size of most commercial sugar-cane plantations, the phenomenal growth rates of the cane during the peak summer months and the impenetrable mesh of foliage that accompanies the cane-stalks, it is little wonder that a developing cane-field is essentially left to its own devices by the farmer. When the time comes to harvest the cane, it quickly becomes apparent, through mishaps or more likely, experience, that the field is home to more than just the sugar-cane — it is generally infested with snakes, which of course, add somewhat to the difficulties posed by the foliage when undertaking the task of harvesting. The time-honoured solution to this problem has been to light a match and stand back. The fires burn rapidly, doing little damage to the cane whilst removing both the foliage and reptile problems in one easy hit. Now consider for a moment the fanciful situation that after the long summer months of idleness watching another Ashes Series victory (this is not the fanciful bit), Farmer Joe has grown impatient and is looking forward to lighting his fire and reducing the home turf to ashes to begin the harvest. He decides that he wants to burn his cane-field as quickly as possible, however the weather prediction is for calm and still conditions and therefore Farmer Joe cannot rely on a strong breeze to drive the fire. Being an industrious chap Farmer Joe reasons that if he got all the local school children to spread out around the boundary of his field, standing shoulder-to-shoulder for instance, and at the first cock-crow at dawn they all lit their matches and set the cane in front of them alight, then the flame-front created around the boundary would burn in towards the centre where it would meet the flame front initiated on the opposite side of the field, at which point the flame-fronts would burn each other out. Farmer Joe further reasons that the burning field will create a rising column of hot air, and thus draw air in from around the field, driving the fire towards the centre of the field and not back towards the pretty school children. This pleases their parents. In a masterful touch of community service, Farmer Joe organises his brother, the local helicopter television reporter, to film the fire so that all the school children can have a copy of the event as a souvenir of the occasion. Well, the big day dawned, the cock crowed, the fire blazed and the harvest was a complete success, all over inside the first day (this is the fanciful bit). A week later, the school children were watching the video of the fire during their physics lesson, having just finished a maths lesson on geometry and symmetry. The children,

some of whom had lost their homes in the big bush fires of last summer, all watched with fascination as the fire fronts approached each other from opposite sides, and then burnt each other out right along a line that was equi-distant to the sides of the field, remarking what a shame it was that they couldn't put the bush fires out like that. One of the students piped up that the answer was obvious. She said that the fire was started along the entire boundary at the same time, and since there was no breeze, the fire burned at the same rate everywhere, and therefore flame fronts initiated on different sides of the field would reach the local symmetry axes at the same time. Indeed it is the case.

Returning to the application of region-thinning, if one records the locations of where the flame-fronts meet, then these points will describe the *local* medial-axis, indicating the location of the local skeleton of the region.

The Haralick and Shapiro algorithm [86, page 278] is built from the sequential application of three primitive operators: the *Mark-Interior/Border-Pixel* operator (MIBP), the *Pair-Relationship* operator (PR), and finally the *Connected-Shrink* operator (CS). The MIBP and PR operators are really pre-processing stages that pass symbolic data to the CS operator where the real work is done.

Further to the notation developed above, let the pixels of the foreground regions $\mathcal{F}$ be subdivided into either *border* pixels $\mathcal{B}$, or *interior* pixels $\mathcal{I}$, such that:

Definition A.3 *A border pixel $\mathcal{B}$ is defined to be a foreground pixel that has a background pixel in its neighbourhood. A pixel can be either 4- or 8-connected to it's neighbourhood depending upon the application/implementation, and*

Definition A.4 *An interior pixel $\mathcal{I}$ is defined to be a foreground pixel that is not a border pixel.*

The general idea behind the algorithm is to initially identify all border pixels of the feature (i.e., where to start the 'fire'). Then at each border pixel, if it is next to an interior pixel (and therefore some of the feature would remain if all the currently labelled 'border' pixels were deleted, i.e., it is not about to hit another flame-front), and if the connectivity of neighbouring pixels would not be destroyed by deleting this pixel (i.e., record the location at which the flame fronts meet and burn each other out), then the pixel is deleted. This procedure is applied iteratively (the flame-front advances a step for each application of the loop) until no further pixels are deleted at the last stage. All remaining pixels are deemed to constitute the skeleton of the feature. The algorithm is shown in figure A.4 and the reader may find it convenient to cross-reference this figure during the following discussion of the algorithm details.

Consider a binarised image I for which the feature skeleton is sought. Without loss of generality, consider those pixels that are components of the image features to be labelled as foreground pixels $\mathcal{F}$, and all other pixels as background, $\mathcal{G}$.

The MIBP operator marks all *interior* pixels with the label $\mathcal{I}$, and all *border* pixels with the label $\mathcal{B}$. Using these definitions for the neighbourhood of a pixel, figures A.5(a) and (b) show the results of applying the MIBP operator to a small group of foreground pixels.

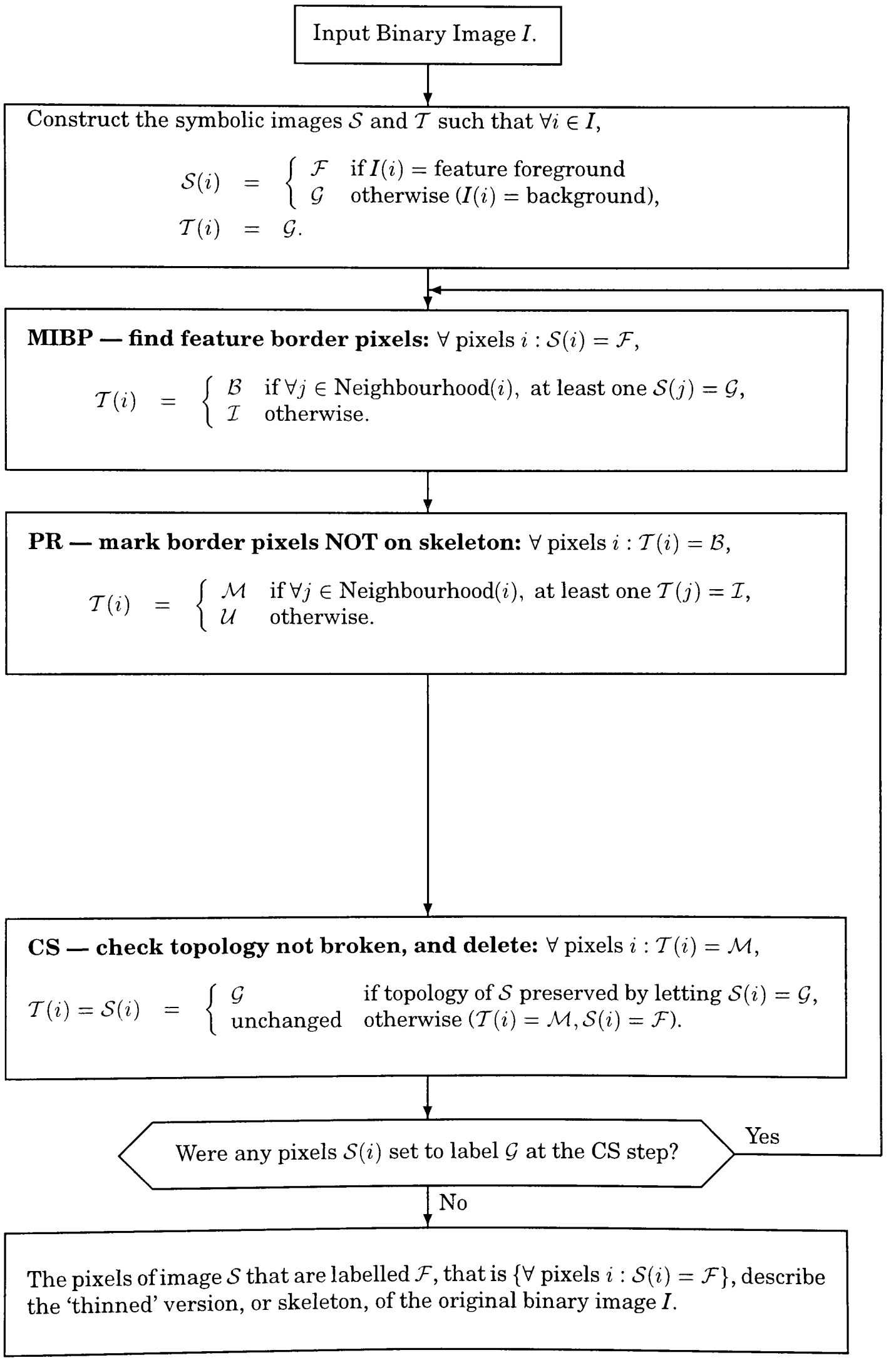


Figure A.4: Haralick and Shapiro's Thinning Algorithm

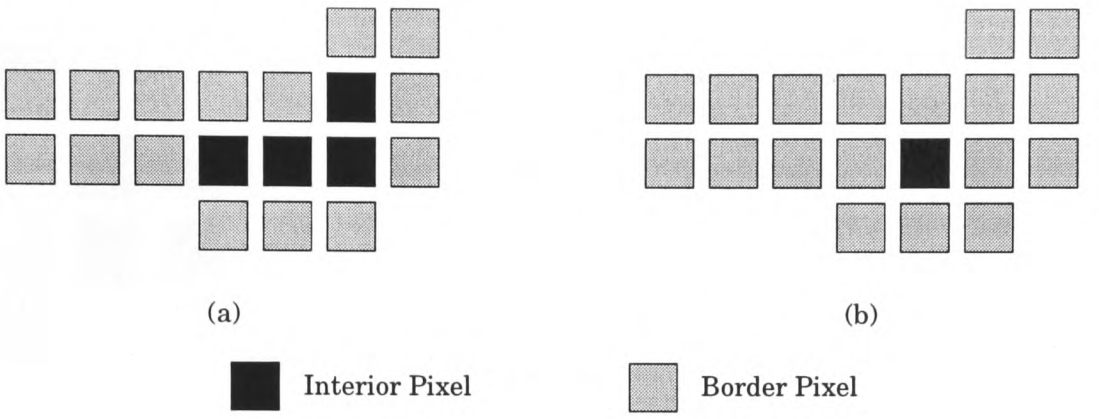


Figure A.5: Results of applying the Mark-Interior/Border-Pixel (MIBP) operator (a) using 4-connected neighbourhood criterion, and (b) 8-connected neighbourhood. Note that for clarity the *background* pixels have not been explicitly displayed, and blank space surrounding the pixels displayed in this figure should be taken to describe the background.

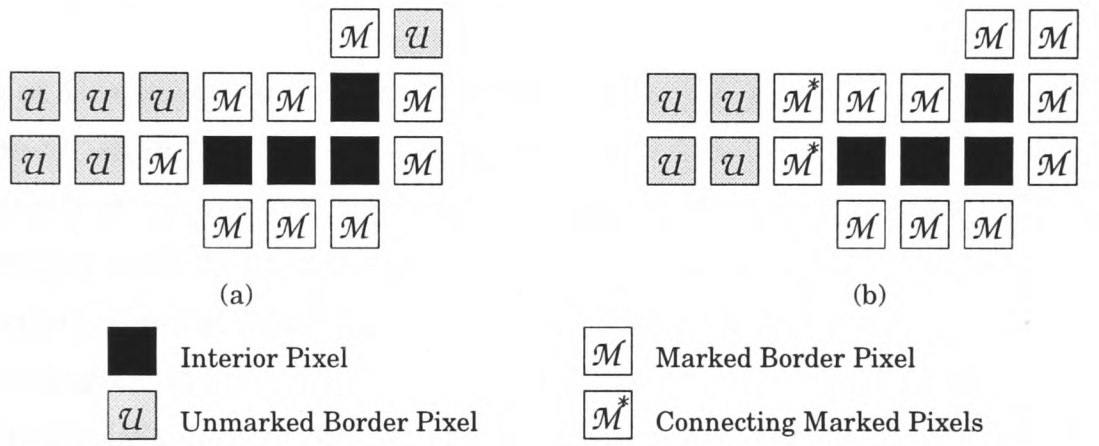


Figure A.6: Effect of the Pair-Relationship (PR) operator applied to the pixel group of figure A.5(c). (a) Results from using a 4-connected PR operator, and (b) an 8-connected PR operator. Note that the PR operator makes no distinction between the $\mathcal{M}$ and $\mathcal{M}^*$ pixels shown here (they are all reported as $\mathcal{M}$ pixels by the PR operator). Rather the Connected-Shrink (CS) operator (which deletes the $\mathcal{M}$ pixels) must not delete *both* $\mathcal{M}^*$ pixels since that would break the topology of the original figure by splitting the remaining pixels into two groups, the unmarked border pixels to the left and the interior pixels to the right.

The Pair-Relationship (PR) operator is a general operator that labels a pixel on the basis of whether it stands in the specified relationship with its neighbourhood pixels. For the current application, this operator is used to label border pixels $\mathcal{B}$ that have at least one interior pixel $\mathcal{I}$ as a neighbour. Those pixels that comply with this are labelled as *marked border* pixels $\mathcal{M}$, and those that do not are labelled as *un-marked border* pixels $\mathcal{U}$.

Definition A.5 *A marked border pixel $\mathcal{M}$ is a border pixel with at least one interior pixel $\mathcal{I}$ as a neighbour.*

Definition A.6 *An un-marked border pixel $\mathcal{U}$ is a border pixel with no neighbouring interior pixels $\mathcal{I}$.*

Figure A.6 shows the effect of applying this operator to the pixel group of figure A.5 (a).

The PR operator may be formulated in either the 4-connected or the 8-connected mode, and use of these modes produces different results in the final thinning operation. In fact it

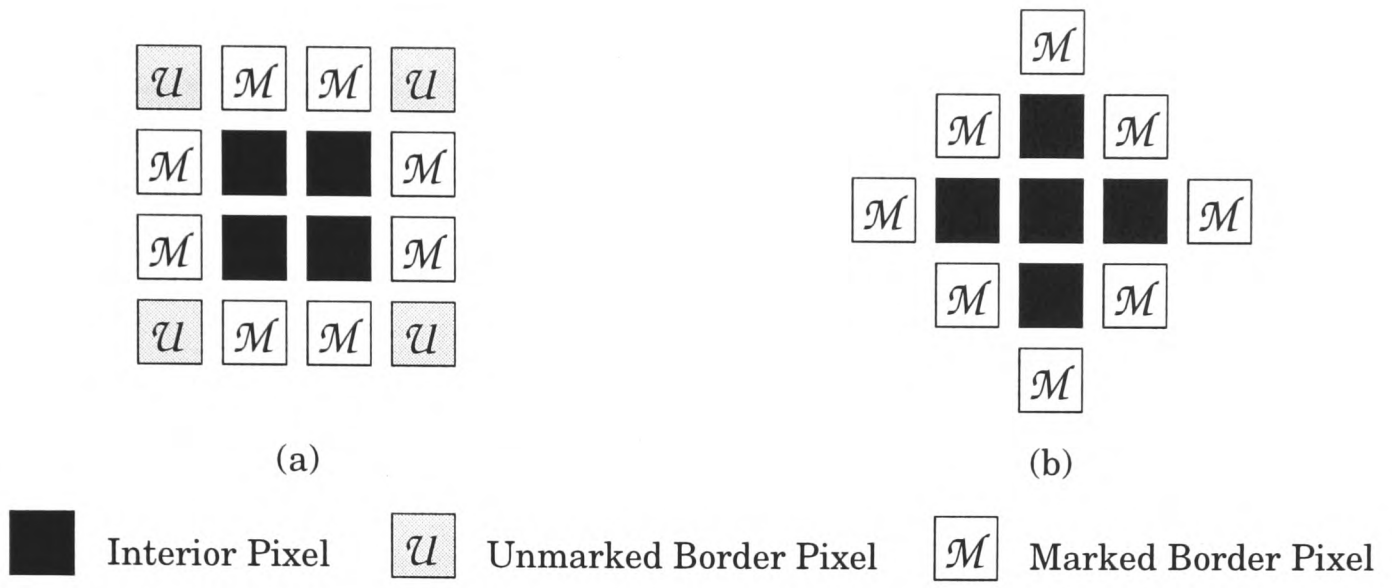


Figure A.7: Rotational variance of the 4-connected Pair-Relationship (PR) operator.

is convenient at this stage to observe the effect that this choice will have, and to keep this knowledge handy when tackling the CS operator which follows next. Observe the top right-hand corner pixel of figures A.6 (a) and (b) (marked as $\mathcal{U}$ and $\mathcal{M}$ respectively). The 4-PR operator leaves the pixel as an un-marked border pixel $\mathcal{U}$, since its 4-neighbours are either $\mathcal{M}$ or $\mathcal{G}$ (background). On the other hand, the 8-neighbourhood of the pixel includes an interior pixel $\mathcal{I}$, and therefore the 8-PR operator will label the pixel as a marked pixel $\mathcal{M}$. In this example, when the CS operator deletes the $\mathcal{M}$ pixels (i.e., relabels them as background $\mathcal{G}$), the skeleton of figure A.6(a) will extend to the corner pixel, whilst that of figure A.6(b) will not.

Although the ability to trace the skeleton into the corner of a feature, and so capture the local symmetry axes correctly, might be deemed a desirable feature, it should be noted that this procedure is not rotationally invariant. Consider the pixel arrangements of figure A.7 and the subsequent effects of applying the 4-PR operator. In figure A.7(a) the 4-PR operator will detect the feature corners, whilst in figure A.7(b) it will not. To this question of rotational variance we shall return at a later stage.

The final stage of the algorithm, the Connected-Shrink (CS) operator, ensures that the topology of the original feature is preserved in the final skeleton that is returned. Its function is to identify those marked-pixels that can be deleted *without* disconnecting the remaining pixels (breaking the pixels into two groups). Figure A.6(b) shows a situation where a number of the marked pixels (labelled $\mathcal{M}^*$) provide the only link between two subgroups of pixels within the original figure (the un-marked pixels $\mathcal{U}$ on the left and the interior pixels $\mathcal{I}$ on the right). If *all* the $\mathcal{M}^*$ pixels were deleted then the topology of the original figure would not be preserved in the final representation of the skeleton. In this particular example, since we will accept an 8-connected skeleton, it matters not which one of the two $\mathcal{M}^*$ pixels is deleted (retained), so long as one of them is retained.

The CS operator decides if a pixel labelled as $\mathcal{M}$ can be relabelled as $\mathcal{G}$ (deleted from the feature) by assessing the neighbourhood of the pixel. This assessment is conducted within the space of the binarised image I (which has elements $\mathcal{F}$ and $\mathcal{G}$) rather than in the image recording

the currently identified label of the pixel ($\mathcal{M}, \mathcal{U}, \mathcal{I}$ or $\mathcal{G}$).

The theoretical basis of the CS operator has been explored by Rosenfeld and Pfaltz [178], Rosenfeld [177] and Stefanelli and Rosenfeld [199], and in its 8-deletable variety (which is the appropriate case in the thinning application since an 8-connected skeleton is desired) is defined according to Yokoi et al. [233] as follows. Consider the primitive functions:

$$h(b,c,d,e) = \begin{cases} 1 & \text{if } c \neq b \text{ and } (d = b \text{ or } e = b) \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

$$f(a_1,a_2,a_3,a_4,x) = \begin{cases} \mathcal{G}^1 & \text{if exactly one of } a_1,a_2,a_3,a_4 = 1 \\ x & \text{otherwise} \end{cases} \tag{A.2}$$

Observing the pixel neighbourhood labelling convention of figure A.1(c) the new value x_{n+1} for a pixel with a current value of x_n is given by

$$\begin{aligned} a_1 &= h(x_n,x_1,x_6,x_2) \\ a_2 &= h(x_n,x_2,x_7,x_3) \\ a_3 &= h(x_n,x_3,x_8,x_4) \\ a_4 &= h(x_n,x_4,x_5,x_1) \end{aligned} \tag{A.3}$$

$$x_{n+1} = f(a_1,a_2,a_3,a_4,x_n). \tag{A.4}$$

Keeping in mind that the purpose of this operator is to determine those marked pixels $\mathcal{M}$ that cannot be relabelled as the background $\mathcal{G}$ because doing so would break/alter the region topology (for example, the pixels marked $\mathcal{M}^*$ in figure A.6(b)), it is straight-forward to grasp the workings of this operator. Observe that as the CS operator is only applied to marked pixels $\mathcal{M}$, which are necessarily foreground pixels $\mathcal{F}$ also, the CS operator (as defined in equations (A.1) through (A.4)) reduces to[2]

$$h'(c,d,e) = \begin{cases} 1 & \text{if } c \neq \mathcal{F} \text{ and } (d = \mathcal{F} \text{ or } e = \mathcal{F}) \\ 0 & \text{otherwise,} \end{cases} \tag{A.5}$$

$$\begin{aligned} a_1 &= h'(x_1,x_6,x_2), \\ a_2 &= h'(x_2,x_7,x_3), \\ a_3 &= h'(x_3,x_8,x_4), \\ a_4 &= h'(x_4,x_5,x_1), \end{aligned} \tag{A.6}$$

$$x_{n+1} = \begin{cases} \mathcal{G} & \text{if } \sum_1^4 a_i = 1, \\ \text{unchanged} & \text{otherwise.} \end{cases} \tag{A.7}$$

Figure A.8 works through the execution of the CS operator as defined by these equations. Although this figure looks a little daunting at first, the persevering reader will be rewarded

[1]The label $\mathcal{G}$ represents the background.

[2]Note that equations (A.5) and (A.6) precisely implement the calculation of the Hilditch [103] crossing number X_H.

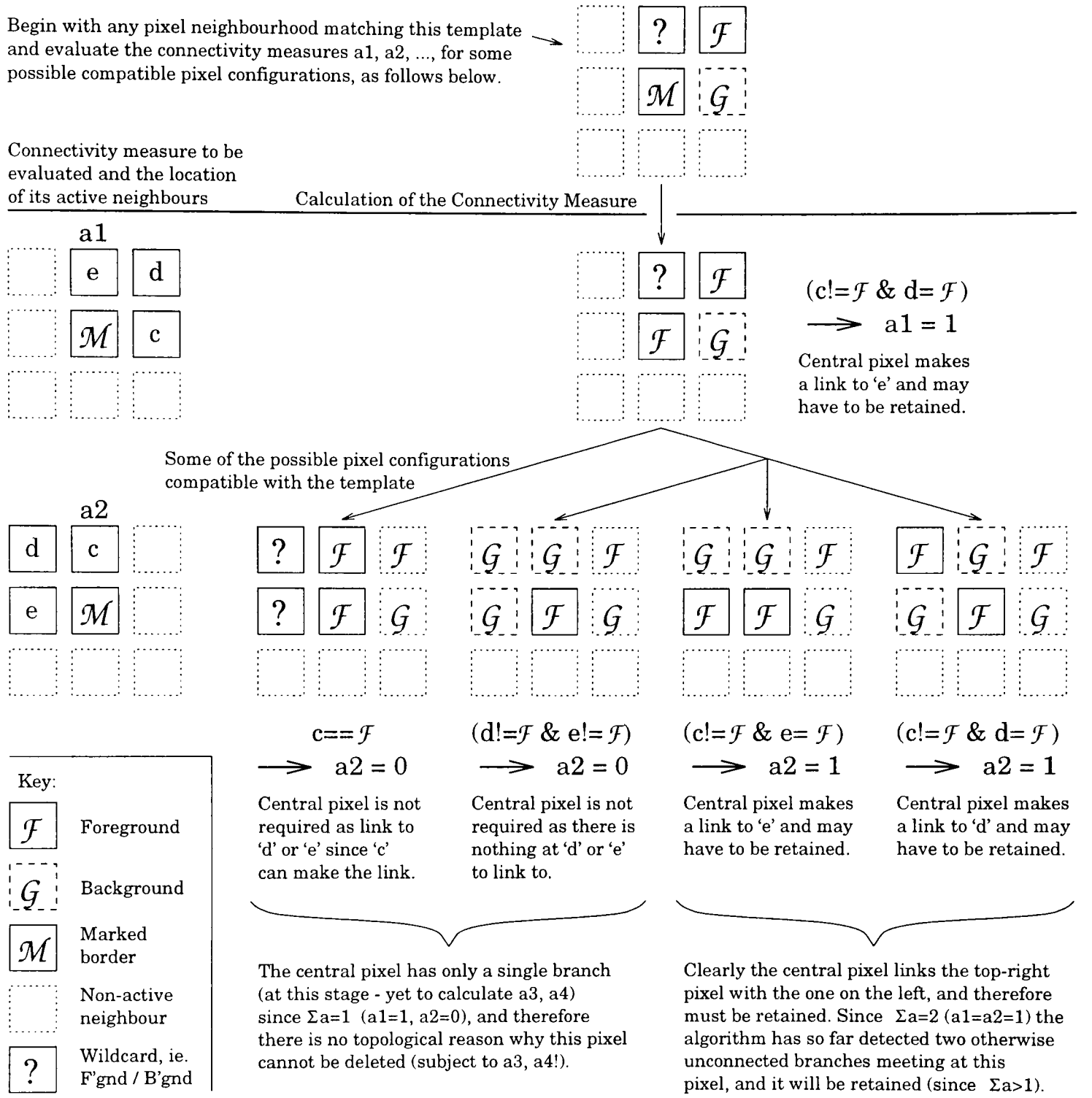


Figure A.8: Workings of the CS operator. Starting with the template at the top of the figure, the table shows the calculation of the connectivity measures a_1 and a_2 (equations (A.6)) for some possible alternative yet compatible pixel configurations. The calculation of a_3 and a_4 follow naturally, allowing final assessment of the neighbourhood connectivity associated with the central pixel (equation (A.7)).

with an intuitive feel for the details of this operator, an asset that will prove valuable when tackling the extension to the Haralick and Shapiro thinning algorithm developed in the following sections.

To illustrate the operation of the CS operator, consider our example region of figure A.9(b) (from figure A.6(b)), and in particular the order that the $\mathcal{M}^*$ pixels will be tackled by the CS operator. Figure A.10, whilst again looking rather daunting, details the decisions made by the CS operator when applied to some of these $\mathcal{M}^*$ pixels. The reader is again encouraged to embrace this figure in the hope of grasping a fuller understanding of the CS operator.

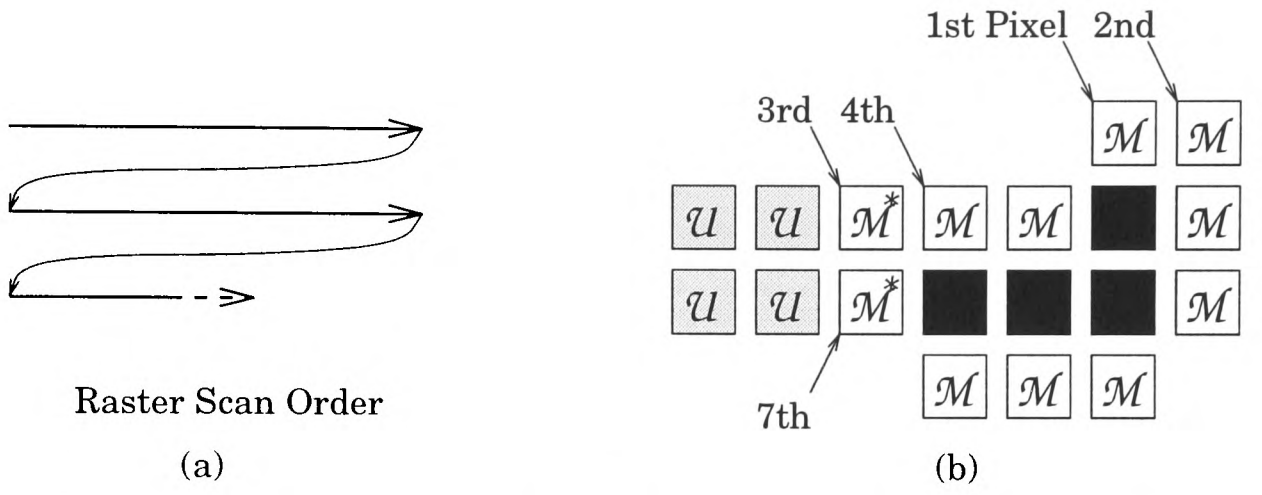


Figure A.9: Order of marked pixel ($\mathcal{M}$) assessment by the CS operator.

For completeness, figure A.11 gives the complete progression of our example figure through the application of the Haralick & Shapiro algorithm.

A.3.1 Discussion

Figure A.12 shows the results of applying this algorithm to a range of pixel regions.

By successively removing the pixels comprising the bounding contour of a region, provided those pixels have *interior* pixels within their neighbourhood, the Haralick & Shapiro thinning algorithm ensures that the pixels returned by the algorithm deemed to comprise the skeleton of the original region are at most one (1) pixel from the true (sub-pixel) location of the skeleton. There is however a nasty consequence of this 'feature' with regard to the present application.

Consider a rectangular region of an *even* number of pixels in width. At each iteration of the thinning procedure a 1-pixel wide strip will be removed from *both* sides/ends of the region, leaving the next intermediate stage of the region with an even number of pixels. Continuing until a foreground region of width two pixels remains, leaves the PR operator unable to find any *interior* pixels, thus all pixels are labelled *unmarked* $\mathcal{U}$, and the CS operator retains *all* pixels. In essence, if a region to be thinned is of even-pixel width, then the skeleton lies between two pixels, and both pixels are retained in the set describing the skeleton. An example of this can be seen in the left most pixels of figure A.11(h). The result is a complex-connected skeleton and as established in the section A.1, this is unsuitable in the present application. Note however that all is not lost, since unlike some other thinning algorithms (e.g., Pavlidis [166]), this method retains the topology of the original figure.

The following section develops an extension to this algorithm that re-assesses the conditions leading to the complex-connectivity highlighted here, developing an algorithm for defeating this difficulty.

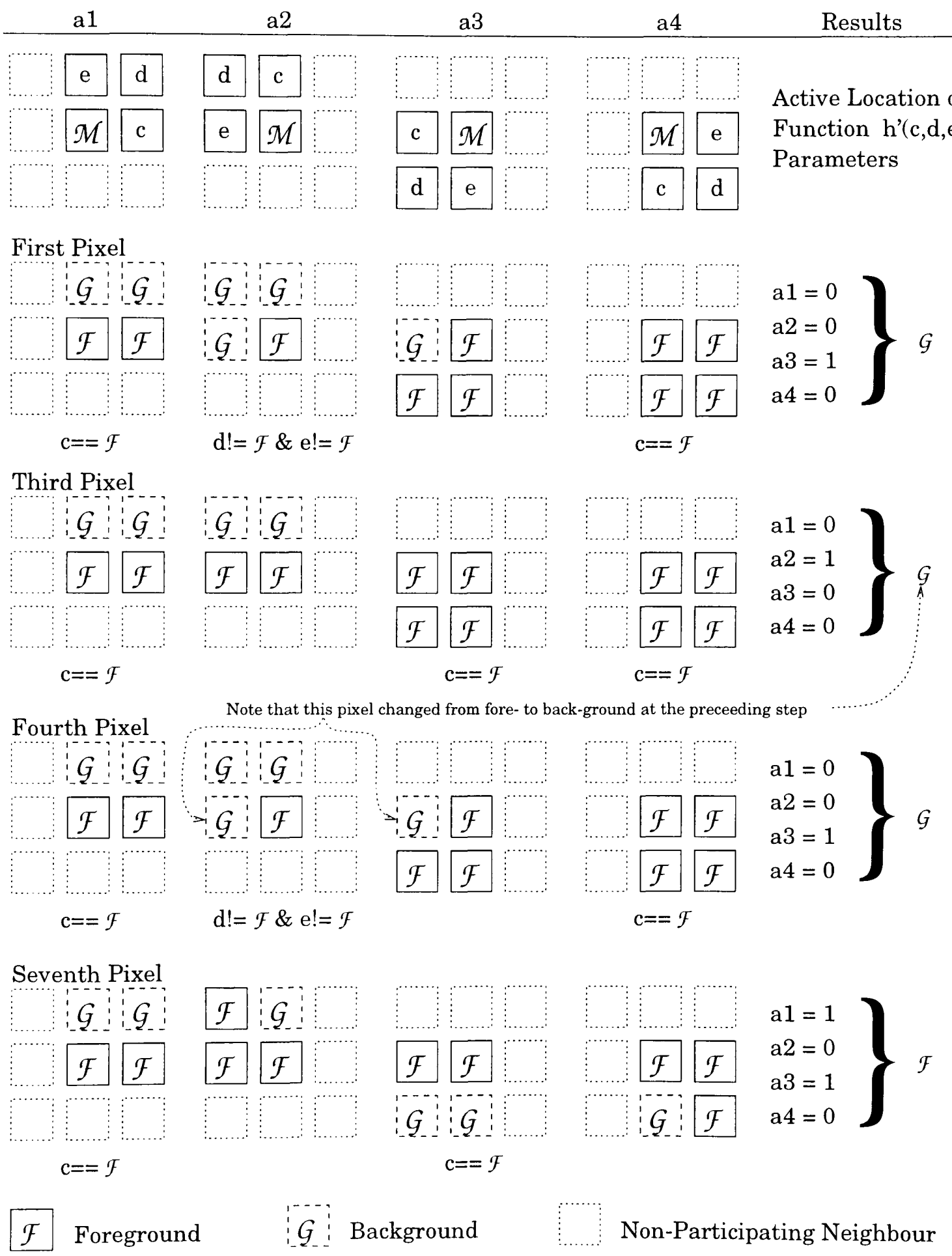


Figure A.10: The CS operator in action. The columns of this figure detail the calculation of equations (A.6) for some of the $\mathcal{M}$ pixels of figure A.9 (the rows of this figure). The first row identifies the active locations of the parameters of equation (A.6) for each of the four equations. The evaluation of equation (A.7) is given in the last column of this figure, and the new value is assigned to the pixel at each step *before* proceeding to assess the next pixel, as can be seen in the step between the 3^{rd} and 4^{th} pixels. Beneath each neighbourhood that made the assignment $a = 0$ is shown the reason for this assignment (equation (A.5)). Note that the last row identifies a pixel that is the sole bridge (since the 3^{rd} pixel was deleted at an earlier stage) connecting two neighbouring regions (twice the assignment $a = 1$ is made) and therefore this pixel *cannot* be deleted as this would split its neighbours into two separate groups.

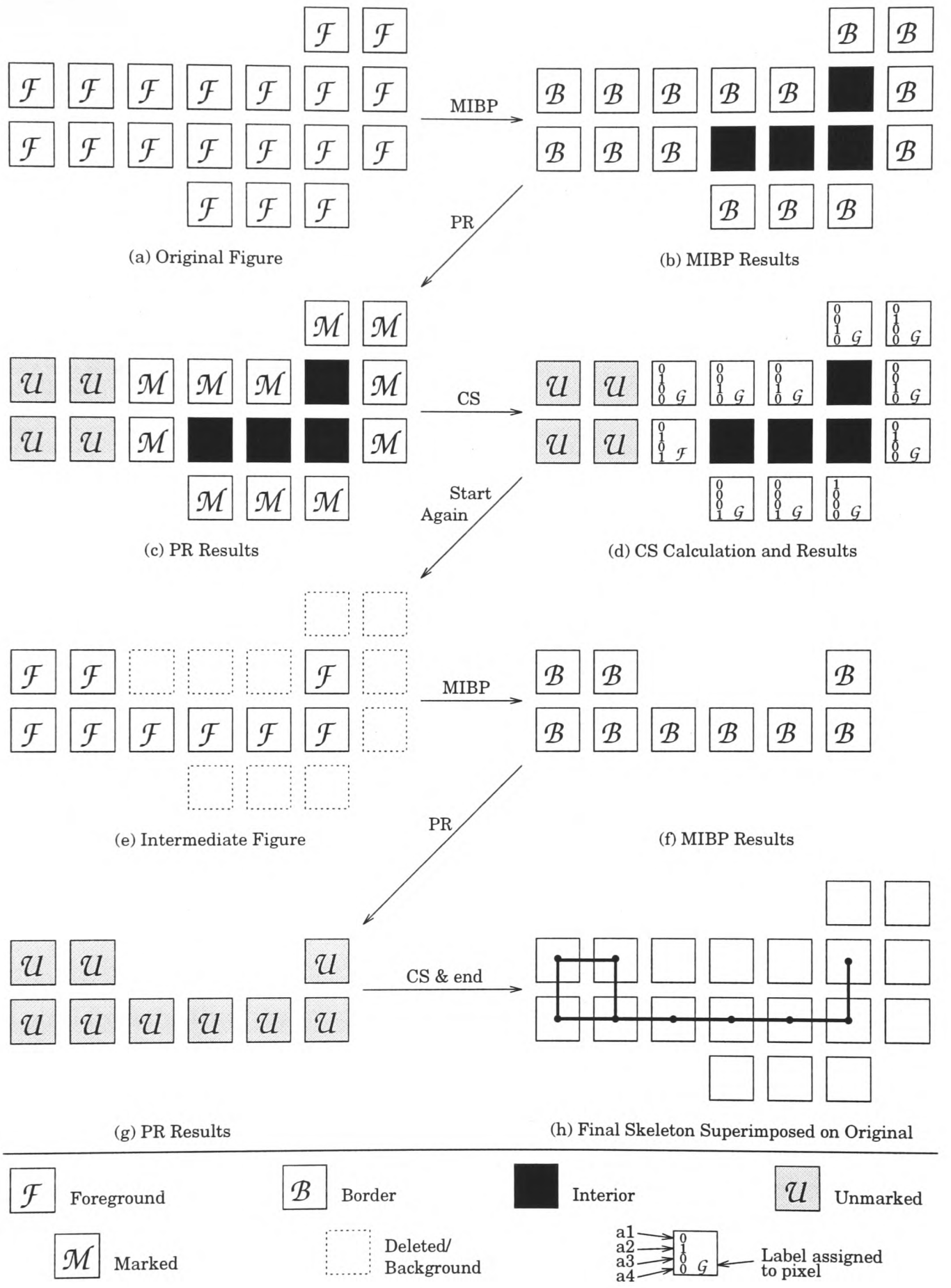


Figure A.11: Progression of the region of figure A.5 through the Haralick & Shapiro thinning algorithm. Note that the final steps, labelled here as 'CS & end' have been omitted for brevity. Since (g) does not contain any *marked* pixels $\mathcal{M}$ the CS operator does not do anything, and therefore no pixels are deleted/relabelled. Thus the algorithm terminates, returning the pixels shown in (g), and displayed in (h), as the skeleton of the original figure (a).

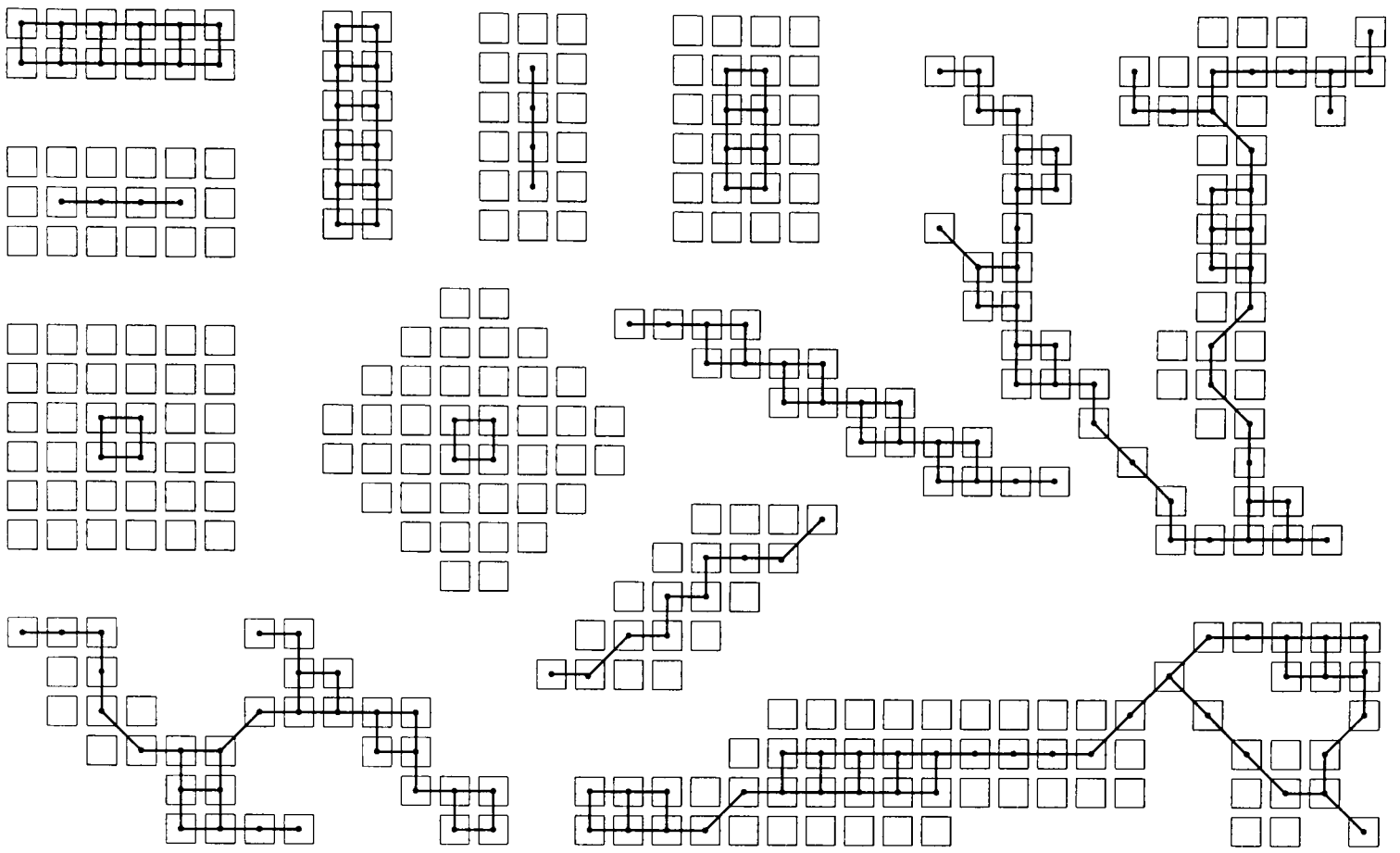


Figure A.12: The skeleton found by the Haralick & Shapiro algorithm for a range of foreground regions. The bottom-left, bottom-right and top-right regions have been taken from real data during the course of finding the CLS features. The remaining features have been generated manually.

A.4 An extension to Haralick & Shapiro's thinning algorithm

A.4.1 Introduction

The thinning algorithm of Haralick & Shapiro (previous section) returns a skeleton that lies at most a pixel from the true (analytic) skeleton. At regions where the true skeleton lies between two pixels they are both retained with the set describing the skeleton. This last feature means that the skeleton of Haralick & Shapiro is often complex-connected.

When thinned to the final Haralick & Shapiro representation of the skeleton, application of the MIBP operator reveals there are no further *interior* pixels $\mathcal{I}$ (otherwise the thinning operation would not be complete), and therefore all pixels will be given the *unmarked* label $\mathcal{U}$ by the PR operator. If a simply-connected skeleton is desired (as it is), it is clear that some of the *unmarked* pixels $\mathcal{U}$ must be deleted (relabelled as the *background* $\mathcal{G}$).

This section describes a method for reassessing the *unmarked* pixels $\mathcal{U}$ returned by the PR operator, and where possible (the definition of 'possible' is the crux of this section) relabelling those pixels as *marked* pixels $\mathcal{M}$. This reassessment is done before passing the data to the CS operator which then functions as before. The net result is to return a simply connected skeleton that is a subset of the pixels returned by the Haralick & Shapiro method. Therefore, this extended algorithm retains the features of preserving the topology of the original feature

and returning a skeleton that is at most a (half) pixel from the true location of the skeleton.

For clarity, we call this operator the Reassess Un-marked (RU) operator.

Figure A.13 illustrates the flow chart of the complete algorithm, clearly showing the similarities with the flow chart for the algorithm of Haralick & Shapiro given in figure A.4. The reader may find this flow chart useful during the discussion of the algorithm details below.

A.4.2 The details

The task of the RU operator parallels that of the CS operator in many ways, and it is worth initially establishing the similarities, thus providing the rationale for the details of the RU operator.

The task of the CS operator is to establish from a list of candidate pixels given to the operator (the *marked* pixels $\mathcal{M}$, which are *foreground* pixels $\mathcal{F}$ in the feature-space — see below) which ones can safely be removed from the list (deleted/relabelled as *background* pixels $\mathcal{G}$) without altering the topology of some feature space (the image $\mathcal{S}$ describing the region being thinned itself). Similarly, the task of the RU operator is to establish from a list of candidate pixels given to the operator (the *unmarked* pixels $\mathcal{U}$) which ones can safely be removed from the list (relabelled as *marked* pixels $\mathcal{M}$) without altering the topology of some feature space (the image $\mathcal{T}$ describing the *unmarked* and *marked* pixels $\mathcal{U}$ and $\mathcal{M}$ respectively).

Note the subtle difference between these operations with regard to the feature space in which the topology is assessed. In the case of the CS operator, the feature space ($\mathcal{S}$) does not contain the candidate list of pixels ($\mathcal{M} \in \mathcal{T}$), whilst the feature space of the RU operator ($\mathcal{T}$) *does* contain the candidate pixels ($\mathcal{U}$). This difference will necessitate both a slight alteration to the ordinary CS operator and an additional test for topology preservation over those of the ordinary CS operator, as will become apparent in the following discussion.

The sequence of operator application (see figure A.13) allows some leniency in the specification of the RU operator. Since the CS operator, which checks for topology preservation prior to deleting any pixels, *follows* the RU operator, it is not imperative that RU operator return a topology preserving pattern of *unmarked* pixels $\mathcal{U}$, but rather it is sufficient to return enough pixels labelled $\mathcal{U}$ to act as *anchors* for the topology calculation of the CS operator. Thus it is acceptable to return more *marked* pixels $\mathcal{M}$ than can safely be deleted in order to preserve topology. One could depict this procedure by describing the CS operator as possessing a level of redundancy that allows it to correct for the mistakes of the preceding steps/operators. A more charitable view would be to describe the RU operator as taking advantage of the redundancy of the CS operator to propose the maximum possible range of pixels that might be deleted, and thus ensure the minimum-pixel set describing the feature skeleton is obtained. For this reason, the RU operator assesses the neighbourhood topology of each $\mathcal{U}$ pixel in the space of the *original* $\mathcal{T}$ image, recording those pixels to be remarked as $\mathcal{M}$ in a temporary image, to be transcribed to $\mathcal{T}$ following the completion of the topology assessment phase of the operator[3].

[3]This procedure can of course be implemented without the need for a temporary image by using a more sophisticated

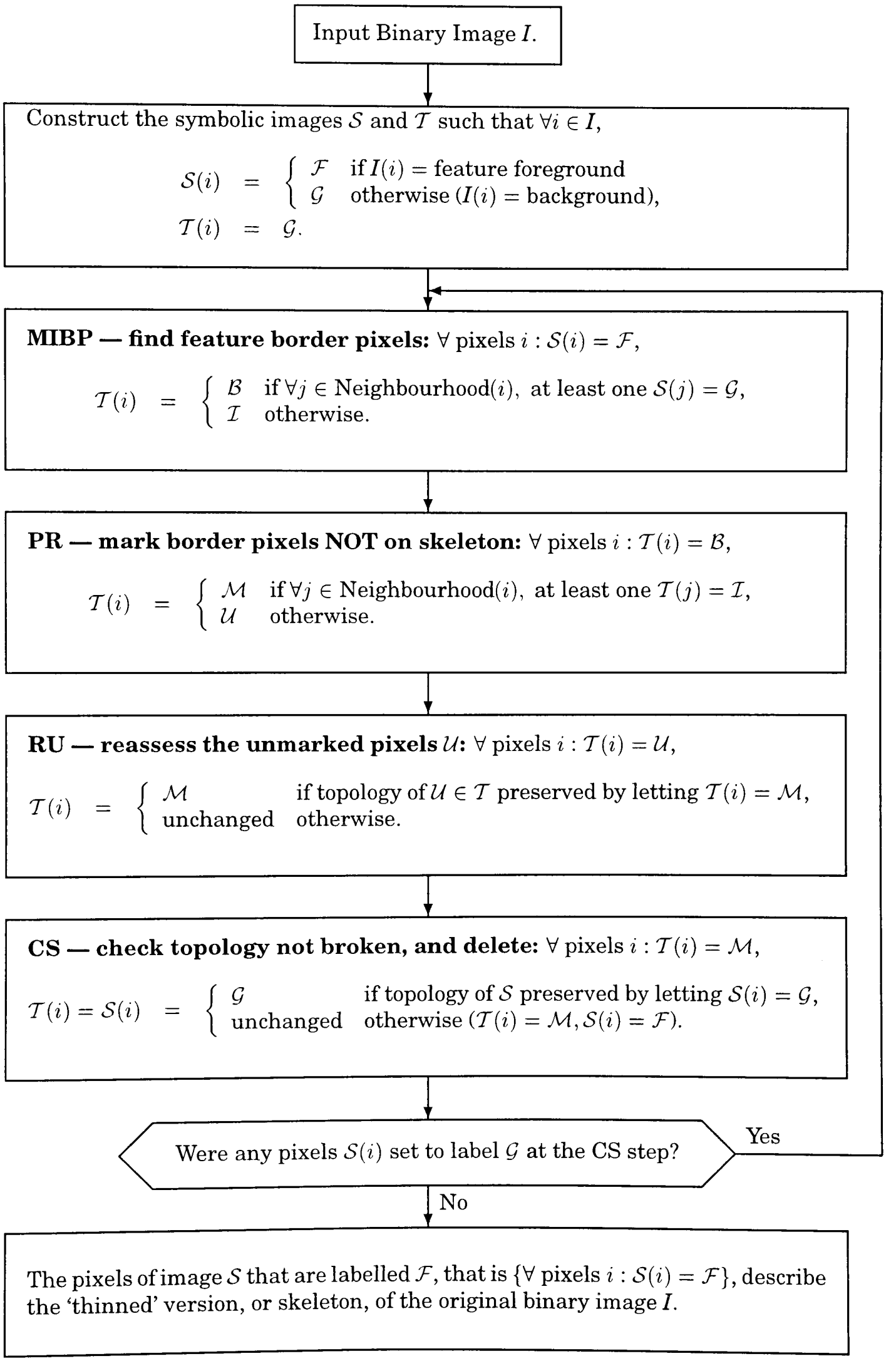


Figure A.13: Extended Haralick and Shapiro Thinning Algorithm

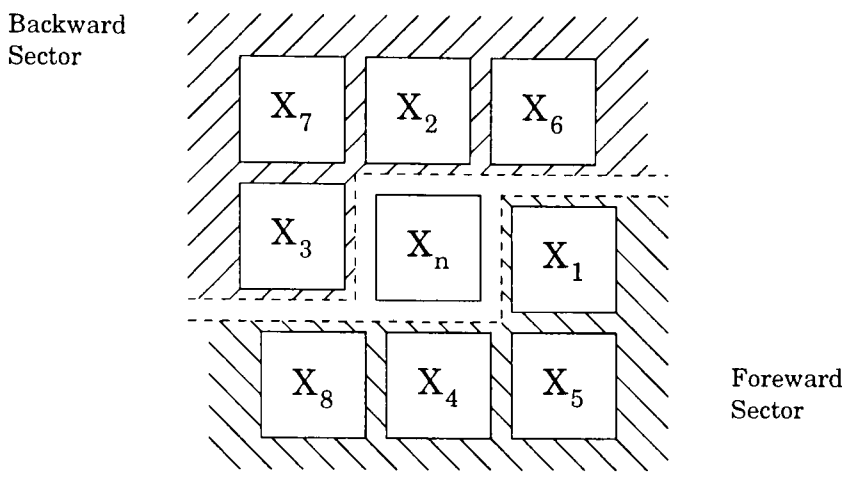


Figure A.14: The forward and backward sectors of a pixel's neighbourhood. Consider the raster-scan order of figure A.9(a). The backward sector comprises those pixels that have already been visited by the current operation (whatever that may be) when the x_n pixel is assessed, whilst the forward sector comprise those pixels yet to be visited.

By saving the results of each assessment in a temporary location, the assessment of topology at each pixel is independent of the assessment made at neighbouring pixels (unlike the formulation of the CS operator which updates a pixel's label upon calculating a new value, and the new label is used in all subsequent calculations involving the pixel), and therefore a pixel's connectivity is assessed purely on the basis of the original figure. Clearly there exists situations where a group of pixels might *individually* qualify for deletion/relabelling due to the existence of the other pixels in the group, however, not *all* of them may be deleted/relabelled (for example the pixels labelled $\mathcal{U}$ in figure A.6(b)). By adopting this technique, the algorithm is sure to propose all the pixels that might possibly be deleted/relabelled during application of the CS operator.

Clearly, one can take this too far, say, relabelling all *unmarked* pixels as *marked*. In such a situation, the CS operator has no evidence upon which to retain any pixels (since there are no *unmarked* pixels that the *marked* pixels could possibly link together), and they are all deleted—hardly a rewarding result! This matter is resolved by observing that the spatial extent over which the CS operator can gather evidence to support the retention of a pixel *marked* for deletion/relabelling is the 8-neighbourhood of that pixel. Thus if a pixel is to be relabelled from *unmarked* to *marked*, there must be sufficient evidence for the CS operator to return that pixel to *unmarked* should it be required to preserve the topology of the figure in the $\mathcal{S}$ image. Pixels which retain their *unmarked* label $\mathcal{U}$ due to this, were termed *anchor* pixels in the preceding paragraphs.

Consider the segmentation of a pixel's 8-neighbourhood into two regions, the forward and backward sectors, as shown in figure A.14. By inspection (and examination of the examples which follow), to ensure that a pixel has sufficient neighbouring information for the CS operator to reassign the *unmarked* label $\mathcal{U}$ if required, it suffices to check that a pixel has an *unmarked* pixel in both its forward and backward sectors before remarking it as a *marked* pixel $\mathcal{M}$. If a

system of labels than that described here. The system described here has been chosen for clarity and ease of explanation.

pixel does not enjoy the company of *unmarked* pixels in both its forward and backward sectors (due to either the lack of *unmarked* pixels to begin with, or because they have been relabelled as *marked* pixels during this operation), then that pixel must retain its *unmarked* label $\mathcal{U}$, and is therefore an *anchor* pixel.

In consideration of these points, the algorithm for the RU operator is set out as follows. In keeping with the notation of the previous sections, consider the primitive functions:

$$h''(c,d,e) = \begin{cases} 1 & \text{if } c \neq \mathcal{U} \text{ and } (d = \mathcal{U} \text{ or } e = \mathcal{U}) \\ 0 & \text{otherwise,} \end{cases} \tag{A.8}$$

$$u(a) = \begin{cases} 1 & \text{if } \mathcal{Z}(a) = \mathcal{U}, \text{ (see below for definition of } \mathcal{Z}) \\ 0 & \text{otherwise,} \end{cases}$$

$$fb(d_1,d_2,d_3,d_4) = \sum_{i=1}^{4} u(d_i). \tag{A.9}$$

Then the RU operator is formulated as follows:

- Construct the temporary image $\mathcal{Z}$ such that for all pixels $i \in \mathcal{T}$

$$\mathcal{Z}(i) = \begin{cases} \mathcal{U} & \text{if } \mathcal{T}(i) = \mathcal{U}, \\ \mathcal{G} & \text{otherwise.} \end{cases}$$

- Then, for all x_n such that $\mathcal{T}(x_n) = \mathcal{U}$,

$$\begin{aligned} c_1 &= h''(x_1,x_6,x_2), \\ c_2 &= h''(x_2,x_7,x_3), \\ c_3 &= h''(x_3,x_8,x_4), \\ c_4 &= h''(x_4,x_5,x_1), \end{aligned} \tag{A.10}$$

$$\begin{aligned} f &= fb(x_1,x_8,x_4,x_5), \\ b &= fb(x_3,x_6,x_2,x_7), \end{aligned} \tag{A.11}$$

$$\mathcal{Z}(x_n) = \begin{cases} \mathcal{M} & \text{if } \sum_1^4 c_i = 1 \text{ and } f > 0 \text{ and } b > 0, \\ \text{unchanged} & \text{otherwise.} \end{cases} \tag{A.12}$$

- Finally, for all i such that $\mathcal{Z}(i) = \mathcal{M}$,

$$\mathcal{T}(i) = \mathcal{M}.$$

The similarity (equivalence excepting labels) of equations (A.8), (A.10) and (A.12) with those of the CS operator (equations (A.5), (A.6) and (A.7)) should be apparent, and accordingly the reader is referred to the description of the CS operator of section A.3 (and in particular figures A.10 and A.11) for an explanation of the workings of this component of the RU operator.

Equations (A.9), (A.11) and (A.12) implement the test for the *anchor* pixels described previously (compare the parameters of equations (A.11) with the details of figure A.14 to see how this is achieved).

Figure A.15 shows the complete passage of our example region through the extended Haralick & Shapiro algorithm, which should be compared with figure A.11 detailing the passage of the region through the ordinary Haralick & Shapiro algorithm.

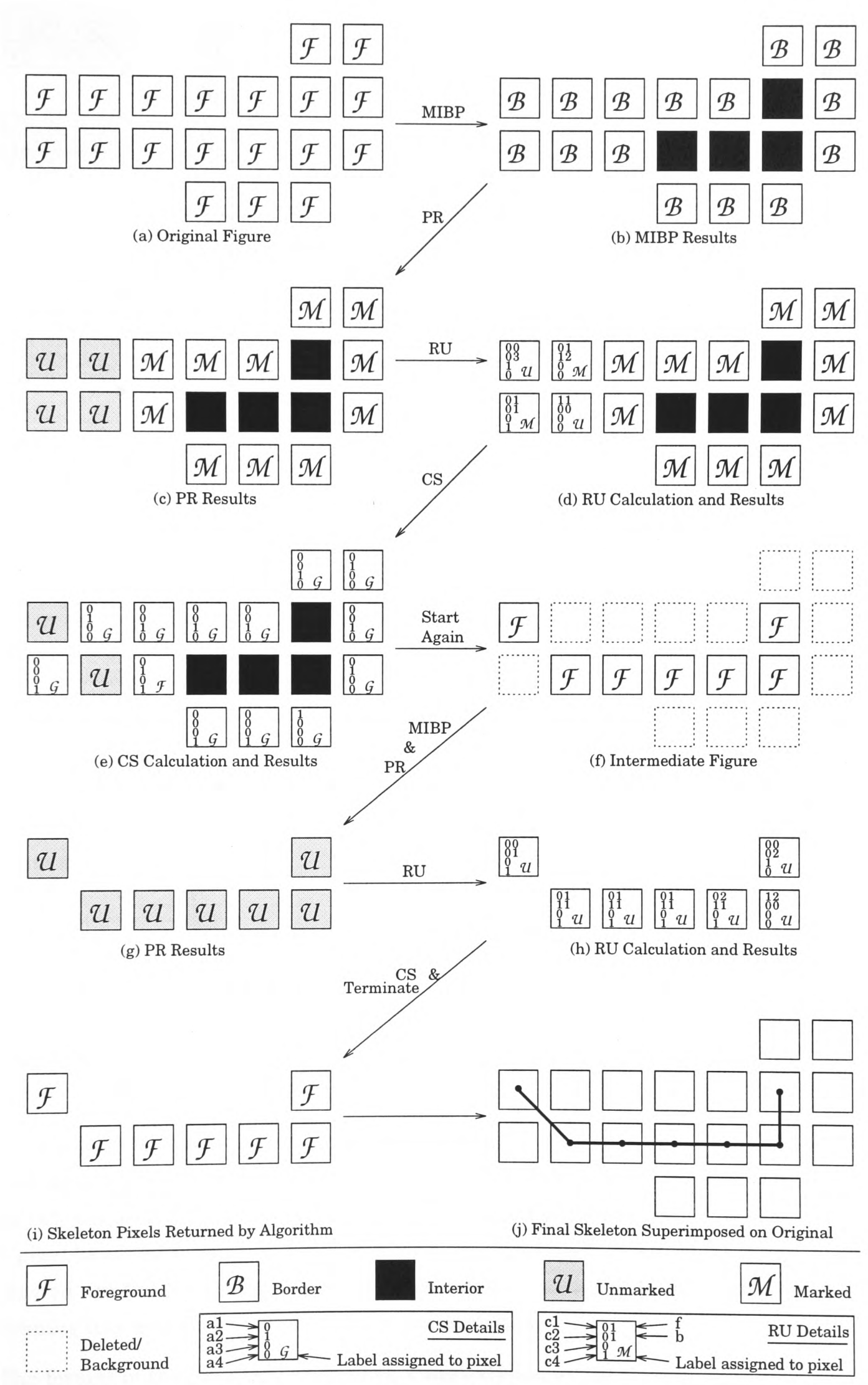


Figure A.15: Progression of the region of figure A.5 through the extended Haralick & Shapiro thinning algorithm. For a comparison of the original and extended Haralick & Shapiro thinning algorithms, compare this figure with figure A.11.

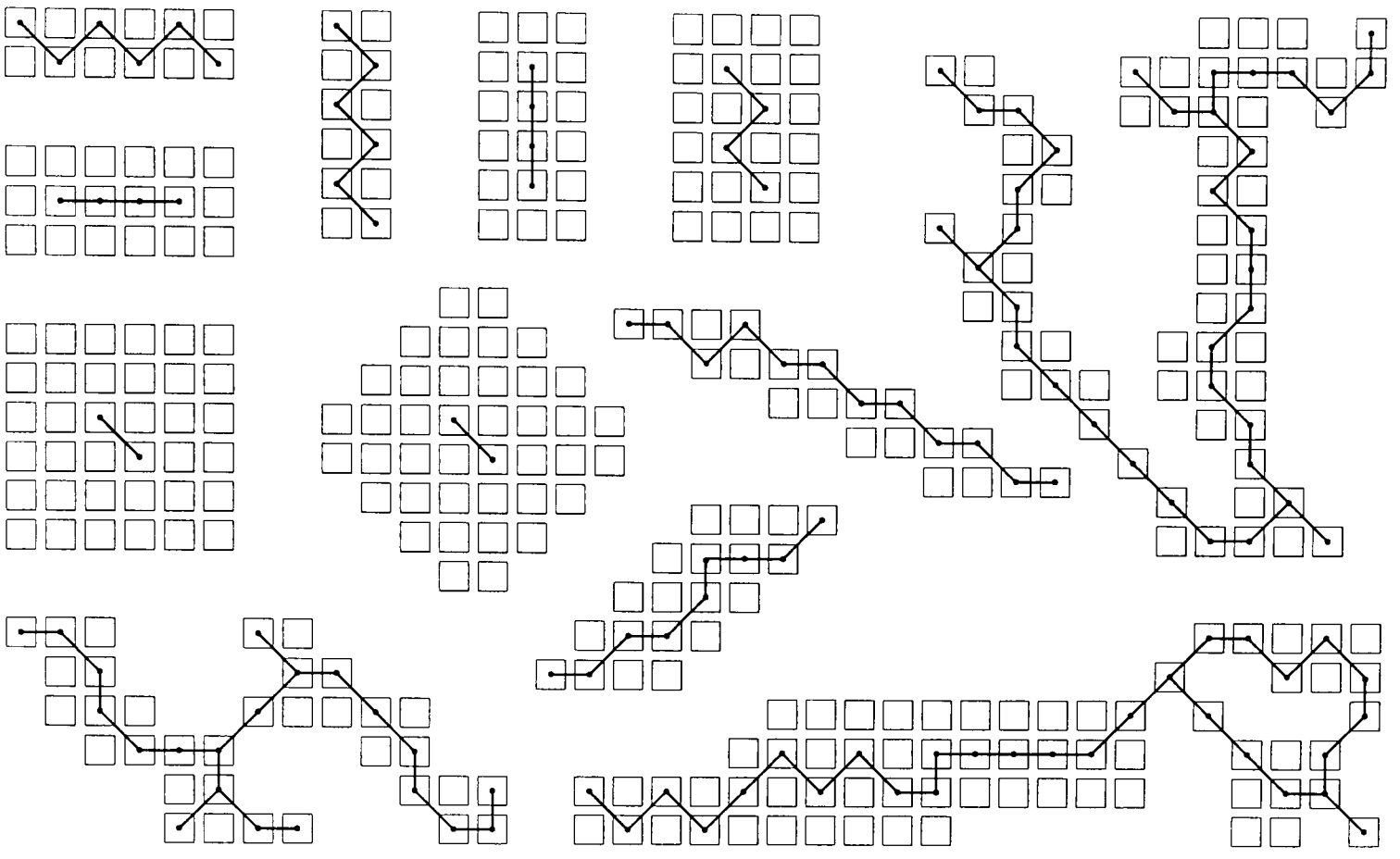


Figure A.16: The skeleton found by the extended Haralick & Shapiro algorithm for a range of foreground regions. The bottom-left, bottom-right and top-right regions have been taken from real data during the course of finding the CLS features. The remaining features have been generated manually.

Figure A.16 shows the results of applying this algorithm to a range of pixel regions (compare with figure A.12). Note that a more complete example of the operation of this algorithm is shown in figure 4.26 on page 100.

A.5 Limitations of the Extended Haralick & Shapiro algorithm

Given the formulation of the RU operator described above, there are three (and three only) situations in which the skeleton identified by the extended algorithm will *incorrectly* return a complexly-connected skeleton. For completeness, these situations, which all involve a single 2×2 pixel block such as that of figure A.3(a), are illustrated in figure A.17 (top row). Also illustrated are the remaining possible arrangements involving a 2×2 pixel block and the corresponding (correct) skeletons returned by the algorithm.

The format of the incorrectly extracted complexly-connected skeletons is due to the raster scan order of the analysis (see figure A.9(a)), explaining why the symmetrical arrangements of these three cases do not return complexly-connected skeletons.

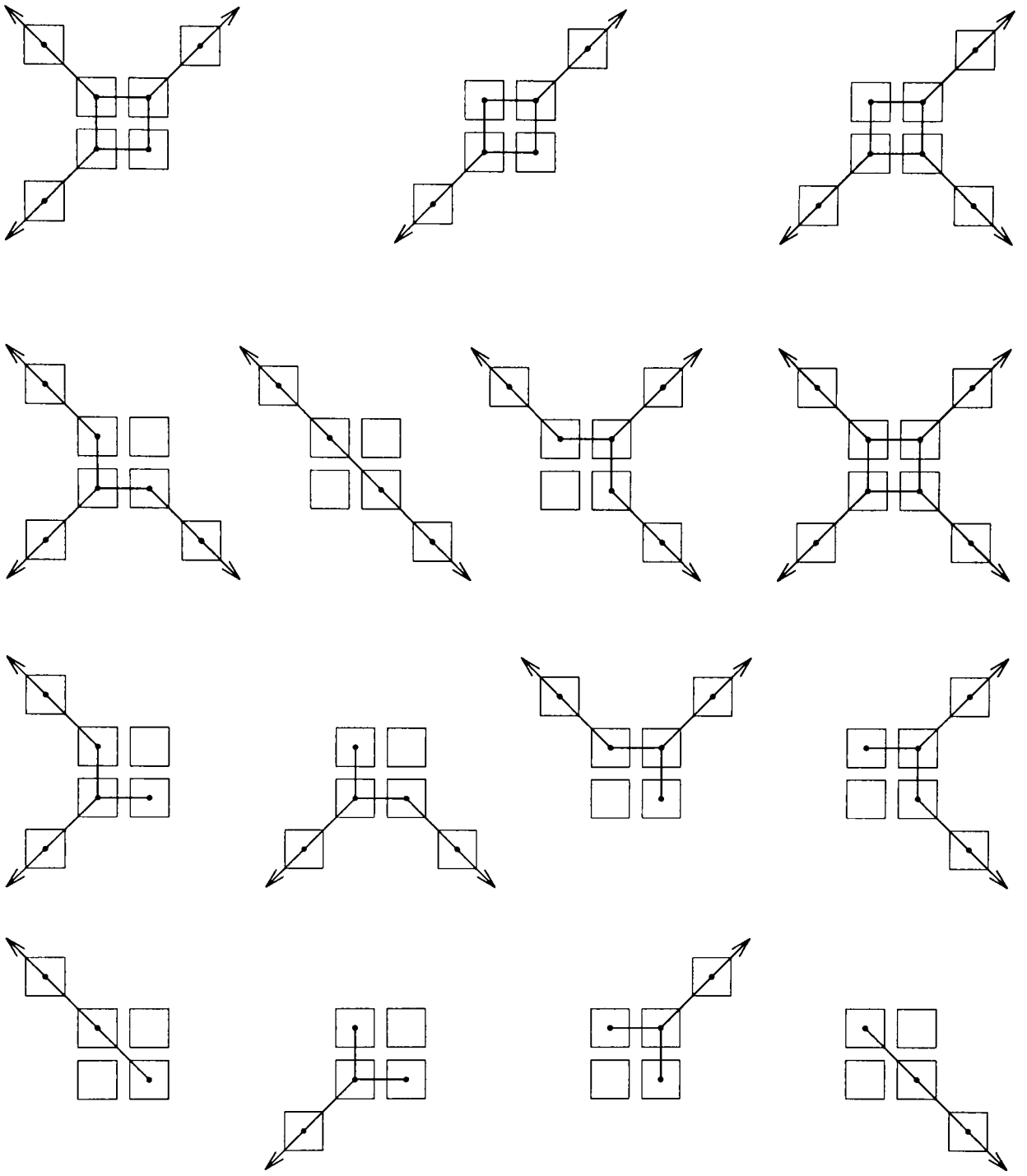


Figure A.17: The skeletons identified by the extended Haralick & Shapiro algorithm for the underlying pixel arrangements. The top row shows the three scenarios for which the algorithm incorrectly extracts complexly-connected skeletons. The remaining regions show the correct skeletons identified for the other possible scenarios involving a 2×2 pixel block. The arrows indicate the allowable wildcard extensions of the region.

A.6 Region Thinning Conclusions

This appendix has established that the CLS feature detector requires a pixel-based, topologically equivalent, simply-connected skeletal representation of the CLS pixels previously identified. Analysis of the Haralick & Shapiro grass-fire based thinning algorithm identified that it returned a pixel-based, topologically equivalent, complexly-connected skeleton. An extension to this algorithm was developed in the present work to add the attribute of returning a simply-connected skeleton to its qualities.

It was also acknowledged that the extension did not quite return a simply-connected skele-

ton, failing to fully remove those pixels causing the additional complexity in only three situations. The existence of these three anomalies is stated here for completeness, yet no attempt is made to correct for them, either by reformulation of the extension to the Haralick & Shapiro algorithm, or by attaching a post-processing step to identify and correct these situations. Clearly, this last alternative, whilst the simplest, detracts from the simplicity and execution speed of the algorithm, yet the former requires a complete reappraisal. It is noted that within the present work these situations both appear rarely, and have little or no consequence on the final outcome of the CLS feature description and its applications. It is for this reason that no attempt is made to correct for the anomalies. Obviously, if it was desired to be able to say that "the algorithm is *guaranteed* to identify a simply-connected, topologically equivalent skeleton lying at most a pixel from the medial-axis of each region" then the conditions giving rise to the three anomalous situations would need to be corrected.

Bibliography

[1] L Ackerman and E Gose. Breast lesion classification by computer and xeroradiograph. *Cancer*, 30:1025–1035, 1972.

[2] AH Adams, JR Brookeman, and MB Merickel. Breast lesion discrimination using statistical analysis and shape measures on magnetic resonance imagery. *Computerized Medical Imaging and Graphics*, 15(5):339–349, 1991.

[3] DD Adler, PL Carson, JM Rubin, and D Quinn-Reid. Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. *Ultrasound in Medicine and Biology*, 16:553–559, 1990.

[4] MP Albert, E Sachsse, NPW Coe, et al. Correlation between mammography and the pathology of nonpalpable breast lesions. *Journal of Surgical Oncology*, 44(1):44–46, 1990.

[5] S Anderson-Engels, R Berg, S Svanberg, and O Jarlman. Time-resolved transillumination for medical diagnostics. *Optics Letters*, 15:1179–1181, 1990.

[6] VF Andolina, SL Lillé, and KM Willison. *Mammographic Imaging*. JB Lippincott Company, Philadelphia, 1992.

[7] H Asada and M Brady. The curvature primal sketch. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–8(1):2–14, January 1986.

[8] N Ayache, JD Boissonat, E Brunet, et al. Building highly structured volume represenations in 3D medical images. In *Proceedings of teh 3rd Symposium on Computer Assisted Radiology*, pages 765–772, Berlin, 1989.

[9] LW Bassett and C Kimme-Smith. Breast sonography. *American Journal of Roentgenology*, 156(3):449–455, 1991.

[10] M Baum. *Breast cancer: The facts*. Oxford Medical Publications. Oxford University Press, Oxford, UK, 1981.

[11] R Beccherle, A del Guerra, M Gambaccini, et al. Developments in digital mammography with monochromatic and narrow energy band x-rays. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 153–162, York, England, 10–12 July 1994. Elsevier Science.

[12] R Berg, O Jarlman, and S Svanberg. Medical transillumination imaging using short-pulse diode lasers. *Applied Optics*, 34(4):574–579, February 1993.

[13] J Black and B Young. A radiological and pathological study of the incidence of calcifications in diseases of the breast and neoplasms of other tissues. *British Journal of Radiology*, 58:596–598, 1965.

[14] H Blum. A transformantion for extracting new descriptors of shape. In W Wathen-Dunn,

editor, *Models for the Perception od Speech and Visual Form*, pages 363–380, Cambridge, MA, 1967. MIT Press.

[15] H Blum. Biological shape and visual science. *International Journal of Theoretical Biology*, 38:205–287, 1973.

[16] CC Boring, TS Squires, T Tong, and S Montgomery. Cancer statistics, 1994. *CA-A Cancer Journal for Clinicians*, 44(1):7–26, 1994.

[17] NF Boyd, E O'Sullivan, E Fishell, et al. Mammographic patterns and breast cancer risk: methodoligic standards and contradictory results. *Journal of the National Cancer Institute*, 72:1253–1259, 1984.

[18] S Bradamante, E Barchiesi, S Pilotti, and G Borasi. High resolution H-1 NMR spectroscopy in the diagnosis of breast cancer. *Magnetic Resonance in Medicine*, 8:440–449, 1988.

[19] JM Brady. Criteria for representations of shape. In A Rosenfeld and J Beck, editors, *Human and Machine Vision*, pages 39–84, New York, 1983. Academic Press.

[20] JM Brady and H Asada. Smoothed local symmetries and their implementations. *International Journal of Robotics Research*, 3(3):36–61, 1984.

[21] DS Brettle, G M^{c}Leod, RJ Oddy, et al. Automatic microcalcification localisation using matched Fourier filtering. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 21–30, York, England, 10–12 July 1994. Elsevier Science.

[22] FL Brookstein. The line-skeleton. *Computer Graphics and Image Processing*, 11:123–137, 1979.

[23] D Brzakovic, XM Luo, and P Brzakovic. An approach to automated detection of tumors in mammograms. *IEEE Transactions on Medical Imaging*, 9(3):233–241, September 1990.

[24] D Brzakovic, N Vujovic, M Neskovic, et al. Mammogram analysis by comparison with previous screenings. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 281–287, York, England, 10–12 July 1994. Elsevier Science.

[25] P Burns, M Halliwell, and PNT Wells. Ultrasonic doppler studies of the breast. *Ultrasound in Medicine and Biology*, 8:127–143, 1982.

[26] CB Caldwell, SJ Stapleton, DW Holdsworth, et al. Characterization of mammographic parenchymal pattern by fractal dimension. *SPIE Medical Imaging III: Image Processing*, 1092, 1989.

[27] CB Caldwell, SJ Stapleton, DW Holdsworth, et al. Characterisation of mammographic parenchymal pattern by fractal dimension. *Physics in Medicine and Biology*, 35(2):235–247, 1990.

[28] J Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–8(6):679–698, 1986.

[29] I Carlbom, D Terzopoulos, and KM Harris. Reconstructing and visualizing models of neuronal dendrites. In NM Patrikalakis, editor, *Scientific Visualization of Physical Phenomena*, pages 623–638. Springer-Verlag, New York, 1991.

[30] J Caseldine, R Blamey, E Roebuck, and C Elston. *Breast Disease for Radiographers*. Wright, London, 1988.

[31] H Chan, K Doi, CJ Vyborny, et al. Digital mammography: development of a computer-aided system for the detection of microcalcifications. *SPIE*, 767:367–370, 1987.

[32] HP Chan, K Doi, S Galhotra, et al. Image feature analysis and computer-aided diagnosis

in digitial radiography 1. automated detection of microcalcifications in mammography. *Medical Physics*, 14(4):538–548, July 1987.

[33] HP Chan, K Doi, KL Lam, et al. Digital characterisation of clinical mammographic microcalcifications: applications in computer-aided detection. *SPIE*, 914:591–593, 1988.

[34] HP Chan, K Doi, CJ Vyborny, et al. Computer-aided detection of microcalcifications in mammograms — methodology and preliminary clinical-study. *Investigative Radiology*, 23:664–671, 1988.

[35] HP Chan, CJ Vyborny, H M^{ac}Mahon, et al. Digital mammography — ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications. *Investigative Radiology*, 22(7):581–589, 1987.

[36] CHJ Chang, DE Nesbit, DR Fisher, et al. Computed tomographic mammography using a conventional body scanner. *American Journal of Roentgenology*, 138:553–558, March 1982.

[37] CHJ Chang, JL Sibala, SL Fritz, et al. Computed tomographic evaluation of the breast. *American Journal of Roentgenology*, 131:459–464, September 1978.

[38] Y Chitre, AP Dhawan, and M Moskowitz. Classification of mammographic microcalcifications using image structure and binary shape features. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 31–40, York, England, 10–12 July 1994. Elsevier Science.

[39] E Claridge. Characterisation of mammographic lesions. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 241–250, York, England, 10–12 July 1994. Elsevier Science.

[40] E Claridge and JH Richter. Measuring edge blur in mammographic lesions. In HU Lemke, K Inamura, CC Jaffe, and R Felix, editors, *Computer Assisted Radiology '93*, pages 612–617. Springer-Verlag, 1993.

[41] HJA Collette, JJ Rombach, NE Day, and F de Waard. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *The Lancet*, 1:1224–1226, 2 June 1984.

[42] A Cooper. *Anatomy and Diseases of the Breast*. Lea and Blanchard, Philadelphia, 1845.

[43] AR Cowen, DS Brettle, NJ Coleman, and GJS Parkin. A preliminary investigation of the imaging performance of photostimulable phosphor computed radiography using a new design of mammographic quality control test object. *British Journal of Radiology*, 65(774):528–535, June 1992.

[44] JS DaPonte and MD Fox. Enhancement of chest radiographs with gradient operators. *IEEE Transactions on Medical Imaging*, MI–7(2):109–117, June 1988.

[45] BB Das, KM Yoo, and RR Alfano. Ultrafast time-gated imaging in thick tissues — a step toward optical mammography. *Optics Letters*, 18(13):1092–1094, 1993.

[46] N Dash. Magnetic reasonance imaging in the diagnosis of breast disease. *American Journal of Roentgenology*, 146:119–125, January 1986.

[47] C David and SW Zucker. Potentials, valleys, and dynamic global coverings. *International Journal of Computer Vision*, 5(3):219–238, 1990.

[48] DH Davies and DR Dance. Automatic computer detection of clustered calcifications in digital mammograms. *Physics in Medicine and Biology*, 35(8):1111–1118, August 1990.

[49] DH Davies, DR Dance, and CH Jones. Automatic detection of microcalcifications in digital

mammograms using local area thresholding techniques. In *SPIE Medical Imaging III: Image Procesing*, pages 153–159, 1989.

[50] UK Department of Health Advisory Committee. *Breast Cancer Screening 1991: Evidence and Experience since the Forrest Report*. NHS Breast Screening Programme, January 1991.

[51] AP Dhawan. A review on biomedical image processing and future trends. *Computer Methods and Programs in Biomedicine*, 31(3–4):141–183, 1990.

[52] AP Dhawan, G Buelloni, and R Gordon. Enhancement of mammographic features by optimal adaptive neighbourhood image processing. *IEEE Transactions on Medical Imaging*, MI–5(1):8–15, March 1986.

[53] AP Dhawan and E Le Royer. Mammographic feature enhancement by computerised image processing. *Computer Methods and Programs in Biomedicine*, 27:23–35, 1988.

[54] JM Dixon, J Walsh, D Paterson, and U Chetty. Colour Doppler ultrasonography studies of benign and malignant breast lesions. *British Journal of Surgery*, 79:259–260, 1992.

[55] W Dock, F Grabenwoeger, V Metz, et al. Tumor vascularization: assessment with duplex sonography. *Radiology*, 181:241–244, 1991.

[56] JS Duncan and T Birkhölzer. Reinforcement of linear structure using parametrized relaxation labelling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–14(5):502–515, May 1992.

[57] RL Egan and MB McSweeney. Mammographic parenchymal patterns and risk of breast cancer. *Radiology*, 133:65–70, 1979.

[58] RL Egan and RC Mosteller. Breast cancer mammography patterns. *Cancer*, 40:2087–2090, 1977.

[59] RW Ehrich and PF Lai. Elements of a structural model of texture. *IEEE Proceedings on Pattern Recognition and Image Processing*, pages 319–326, 1978.

[60] BW Fam. The detection of calcification clusters in film-screen mammograms: a detailed algorithmic approach. *SPIE*, 914:620–634, 1988.

[61] BW Fam, SL Olson, PF Winter, and FJ Scholz. Algorithm for the detection of fine clustered calcifications an film mammograms. *Radiology*, 169:333–337, 1988.

[62] SA Feig. Low-dose mammography: assessment of theoretical risk. In SA Feig and R M^{c}Lelland, editors, *Breast Carcinoma: Current Diagnosis and Treatment*, page 69. ACR and Masson, New York, 1983.

[63] EJ Feuer, LM Wun, CC Boring, et al. The lifetime risk of developing breast-cancer. *Journal of the National Cancer Institute*, 85(11):892–897, 1993.

[64] MM Fleck. Local rotational symmetries. Technical Report AI-TR-852, MIT Artificial Intelligence Laboratory, 1985.

[65] MM Fleck. Boundaries and Topological Algorithms. PhD Thesis. Technical Report 1065, MIT Artificial Intelligence Laboratory, 1988.

[66] MM Fleck. Spectre: an improved phantom edge finder. In *Proceedings of the 5th Alvey Vision Conference*, pages 127–132, University of Reading, UK, 25–28 September 1989.

[67] MM Fleck. Multiple widths yield reliable finite differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–14(4):412–429, April 1992.

[68] T Forester. Bugs and glitches in computers. *Engineering World*, 4(2):45–51, April 1994.

[69] P Forrest. Breast cancer screening. report to the Health Ministers in England, Wales, Scotland and Northern Ireland. Technical report, HMSO, 1986.

[70] ET Fossel, G Brodsky, JL DeLayre, and RE Wilson. Nuclear magnetic resonance for

the differentiation of benign and malignant breast tissues and auxillary lymph nodes. *Annals of Surgery*, 198(4):541–545, October 1983.

[71] S Fox, UM Pujare, WG Wee, et al. A computer analysis of mammographic microcalcifications: global approach. In *IEEE 5th Pattern Recognition Conference*, pages 624–631, 1980.

[72] KE Fredfeldt, E Christensen, K Conradsen, et al. Automatic screening of plain film mammography. *Seminars in Ultrasound, CT and MRI*, 13(2):135–139, April 1992.

[73] M Giger, FF Yin, K Doi, et al. Investigation of methods for the computerized detection and analysis of mammographic masses. *SPIE*, 1233:183–184, 1990.

[74] ML Giger, P Lu, Z Huo, et al. CAD in digital mammography: computerized detection and classification of masses. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 281–287, York, England, 10–12 July 1994. Elsevier Science.

[75] A Glatt, H Longbotham, Thomas Arnow, et al. An application of weighted majority minimum range filters in the detection and sizing of tumors in mammograms. *SPIE Medical Imaging IV: Image Processing*, 1652:447–488, 1992.

[76] R Gordon and RM Rangayyan. Feature enhancement of film mammograms using fixed and adaptive neighbourhoods. *Applied Optics*, 23(4):560–564, 1984. Correction in *Applied Optics*; 23(13),1984.

[77] RE Graham. Snow removal — a noise-stripping process for picture signals. *IEEE Transactions on Information Theory*, IT-8:129–144, 1962.

[78] IH Gravelle, JC Bulstrode, DY Wang, and JL Hayward. The relationship between radiographic features and determinants of risk of breast cancer. *British Journal of Radiology*, 53:107–113, 1980.

[79] RL Gregory. *Eye and Brain: the psychology of seeing*. World University Library. McGraw-Hill Book Company, New York, 1973.

[80] R Guissin and JM Brady. Iso-intensity contours for edge detection. Technical Report OUEL 1935/92, Robotics Research Group, Oxford University, May 1992.

[81] R Gupta and C Srinivas. Image warping for accurate digital subtraction angiography. Technical report, AAAI Spring Symposium on Applications of Computer Vision in Medical Image Processing, Stanford University, USA, 21–23 March 1994.

[82] DA Hale, JF Cook, Z Baniqued, et al. Selective digital enhancement of conventional film mammography. *Journal of Surgical Oncology*, 55(1):42–46, 1994.

[83] ECB Hall-Craggs. *Anatomy as a Basis for Clinical Medicine*. Urban & Schwarzenberg, Munich, second edition, 1990.

[84] W Hand, JL Semmlow, LV Ackerman, and FS Alcorn. Computer screening of xeromammograms: a technique for defining suspicious areas of the breast. *Computer and Biomedical Research*, 12:445–460, 1979.

[85] RM Haralick. Cubic facet model edge detector and ridge-valley detector: implementation details. In ES Gelsema and LN Kanal, editors, *Pattern Recognition in Practice II*, pages 81–90, Oxford, 1986. Elsevier Science Publishers B.C. (North Holland).

[86] RM Haralick and LG Shapiro. *Computer & Robot Vision*, volume 1. Addison-Wesley Publishing, Reading, Mass., 1992.

[87] RM Haralick and L Watson. A facet model for image data. *Computer Graphics and Image Processing*, 15:113–129, 1981.

[88] AG Haus. Recent advances in screen/film mammography. *Radiologic Clinics of North*

America, 25(5):914, 1987.

[89] P Hayton. Detection of globally salient structures in X-ray mammograms. Third year project report, Honours School of Engineering and Computing Science, Oxford University, 21 March 1994.

[90] P Hayton. Fourth year project report, Honours School of Engineering and Computing Science, Oxford University, June 1995 1995. In preparation.

[91] G Hermann, RJ Keller, et al. Non-palpable ductal carcinoma insitu verses infiltrating carcinoma of the breast — can they be differentiated by mammography? *Journal of the Canadian Association of Radiologists*, 42(3):219–222, 1991.

[92] SH Heywang, R Basserman, G Fenzi, et al. MRI of the breast-histopathologic correlation. *European Journal of Radiology*, 7:175–182, 1987.

[93] SH Heywang, T Hilbertz, R Beck, et al. Gd-DTPA enhanced MR imaging of the breast in patients with postoperative scarring and silicon implants. *Journal of Computer Assisted Tomography*, 14:348–356, 1990.

[94] SH Heywang-Köbrunner. Nonmammographic breast imaging techniques. *Current Opinion in Radiology*, 4(V):146–154, 1992.

[95] PF Hickman, NR Moore, and BJ Shepstone. The indeterminate breast mass: assessment using contrast enhanced magnetic resonance imaging. *British Journal of Radiology*, 67(793):14–20, January 1994.

[96] R Highnam. Automated analysis of mammograms. First year report, Oxford University Computing Laboratory, August 1990.

[97] R Highnam. Model-based enhancement of mammographic images. D.Phil thesis, Programming Research Group, Oxford University, Trinity 1992.

[98] RP Highnam. Personal communication, 1994. Robotics Research Group, Oxford University.

[99] RP Highnam, JM Brady, and B Shepstone. Mammograms at different compression plate widths for the detection of breast cancer. *Radiology and Oncology*, 1991. Abstract of work in progress.

[100] RP Highnam, JM Brady, and BJ Shepstone. Model-based enhancement of mammographic images. To be published, Robotics Research Group, Oxford University, May 1992.

[101] RP Highnam, JM Brady, and BJ Shepstone. Computing the scatter component of mammographic images. *IEEE Transactions on Medical Imaging*, 13(2):301–313, 1994.

[102] RP Highnam, BJ Shepstone, and JM Brady. Differential Compression Mammography. To be published, Robotics Research Group, Oxford University, 1992.

[103] CJ Hilditch. Linear skeletons from square cupboards. In B Meltzer and D Michie, editors, *Machine Intelligence*, volume 4, pages 403–420. Elsevier, New York, 1969.

[104] BKP Horn. *Robot Vision*. MIT Press, 1986.

[105] DR Hush and BG Horne. Progress in supervised neural networks. *IEEE Signal Processing Magazine*, pages 8–39, January 1993.

[106] PC Johns and MJ Yaffe. X-ray characterisation of normal and neoplastic breast tissue. *Physics in Medicine and Biology*, 32:675–695, 1987.

[107] J Kaprio, A Alkanko, L Kivisaari, et al. Mammographic patterns in twins discordant for breast cancer. *British Journal of Radiology*, 60:459–462, 1980.

[108] N Karssemeijer. A stochastic method for automated detection of microcalcifications in digital mammograms. In *XII Conference on Information Processing in Medical Imaging*, Wye College, Kent, 7–12 July 1991.

[109] N Karssemeijer. Stochastic model for automated detection of calcifications in digital mammograms. *Image and Vision Computing*, 10(6):369–375, July/August 1992.

[110] N Karssemeijer. Recognition of stellate lesions in digital mammograms. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 211–219, York, England, 10–12 July 1994. Elsevier Science.

[111] N Karssemeijer, JTM Frieling, and JHCL Hendriks. Spatial-resolution in digital mammography. *Investigative Radiology*, 28(5):413–419, 1993.

[112] N Karssemeijer and L van Erning. Iso-precision scaling of digitized mammograms to facilitate image analysis. In *Proceedings of the Medical Imaging V Conference*, San Jose, 1991.

[113] M Kass and A Witkin. Analysing oriented patterns. *Computer Vision, Graphics and Image Processing*, 37(3):362–385, 1987.

[114] M Kass and A Witkin. Analyzing oriented patterns. In MA Fischler and O Firschein, editors, *Readings in computer vision: Issues, problems, principles, and paradigms*, pages 268–276. Morgan Kaufmann, Los Altos, California, 1987.

[115] M Kass, A Witkin, and D Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[116] M Kass, A Witkin, and D Terzopoulos. Snakes: active contour models. In *ICCV87, First International Conference on Computer Vision, London, England*, pages 259–269, 8–11 June 1987.

[117] WP Kegelmeyer Jr. Computer detection of stellate lesions in mammograms. In *Biomedical Image Processing and Three-Dimensional Microscopy*, volume 1660, pages 446–454. SPIE, 1992.

[118] WP Kegelmeyer Jr. Evaluation of stellate lesion detection in a standard mammogram data set. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(6):1477–1492, 1993.

[119] WP Kegelmeyer Jr. and MC Allmen. Dense feature maps for detection of calcifications. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 3–12, York, England, 10–12 July 1994. Elsevier Science.

[120] WP Kegelmeyer Jr., JM Pruneda, PD Bourland, et al. Computer-aided mammographic screening for spiculated lesions. *Radiology*, 191(2):331–337, May 1994.

[121] JL Kelsey. A review of the epidemiology of human breast cancer. *Epidemiology Review*, 1:74–109, 1979.

[122] C Kimme, BJ O'Loughlin, and J Sklansky. Automatic detection of suspicious abnormalities in breast radiographs. In TL Kunii et al., editors, *Data structures, computer graphics and pattern recognition*, pages 427–447. Academic Press, New York, 1977.

[123] C Kimme-Smith, LW Bassett, RH Gold, and L Gormley. Digital mammography, a comparison of two digitization methods. *Investigative Radiology*, 24:869–875, 1989.

[124] SL Kok, JM Brady, and L Tarassenko. The detection of abnormalities in mammograms. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 261–269, York, England, 10–12 July 1994. Elsevier Science.

[125] DB Kopans. "Early" breast cancer detection using techniques other than mammography. *American Journal of Roentgenology*, 143:465–468, September 1984.

[126] DB Kopans and JE Meyer. Computed tomography guided localization of clinically occult breast carcinoma — the "N" skin guide. *Radiology*, 145:211–212, 1982.

[127] A Kundu. Local segmentation of biomedical images. *Computerized Medical Imaging and Graphics*, 14(3):173–183, 1990.

[128] SM Lai, X Li, and WF Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Transactions on Medical Imaging*, MI–8(4):377–386, December 1989.

[129] A Laine, S Schuler, J Fan, et al. A framework for contrast enhancement by dyadic wavelet analysis. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 91–100, York, England, 10–12 July 1994. Elsevier Science.

[130] L Lam, SW Lee, and CY Suen. Thinning methodologies — a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–14(9):869–885, September 1992.

[131] TK Lau and WF Bischof. Automated detection of breast-tumors using the asymmetry approach. *Computer and Biomedical Research*, 24(3):273–295, 1991.

[132] HP Leis Jr. Risk factors for breast cancer: An update. *Breast Disease*, 6(4):24, 1980.

[133] J Levy-Vehel. Early detection of breast cancer by studying microcalcifications on mammography: State of the art and new ideas. , Inria.

[134] HG Lewis-Jones, GH Whitehouse, and SJ Leinster. The role of magnetic resonance imaging in the assessment of local recurrent breast carcinoma. *Clinical Radiology*, 43(3):197–204, 1991.

[135] F Leymaire. Tracking and describing deformable objects using active contour models. Technical Report TR-CIM-90-9, McGill Research Center for Intelligent Machines, McGill University, Montreal, 1990.

[136] F Leymarie. Simulating the grassfire transform using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–14(1):56–75, January 1992.

[137] T Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283, 318 1993.

[138] F Liu, KM Yoo, and RR Alfano. Ultrafast laser-pulse transmission and imaging through biological tissues. *Applied Optics*, 32(4):554–558, 1993.

[139] H Madjar, W Sauerbrei, S Münch, and H Schillinger. Continuous-wave and pulsed Doppler studies of the breast: clinical results and effect of tranducer frequency. *Ultrasound in Medicine and Biology*, 17:31–39, 1991.

[140] IE Magnin, F Cluzeau, CL Odet, and A Bremond. Mammographic texture analysis: An evaluation of risk for developing breast cancer. *Optical Engineering*, 25(6):780–784, 1986.

[141] R Malladi, JA Sethian, and BC Vemuri. Front propagation: a framework for topology independent shape modelling and recovery. Technical report, AAAI Spring Symposium on Applications of Computer Vision in Medical Image Processing, Stanford University, USA, 21–23 March 1994.

[142] P Maragos and RD Ziff. Threshold parallelism in morphological feature extraction, skeletonization and pattern spectrum. In *Visual Communications and Image Processing*, volume 100, pages 106–115. SPIE, 1988.

[143] PA Maragos and RW Schafer. Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP–34(5):1228–1244, October 1986.

[144] DE March, RJ Weschsler, AB Kurtz, et al. CT-pathologic correlation of auxillary lymph nodes in breast carcinoma. *Journal of Computer Assisted Tomography*, 15:440–443, 1991.

[145] DC Marr. *Vision*. WH Freeman & Co, New York, 1982.

[146] JE Martin, M Moskowitz, and JR Milbrath. Breast cancer missed by mammography. *American Journal of Roentgenology*, 132:737–739, 1979.

[147] L mendell, M Rosenbloom, and A Naimark. Are breast patterns a risk index for breast cancer? A resppraisal. *American Journal of Roentgenology*, 128:547, 1977.

[148] P Miller and S Astley. Classification of breast tissue by texture analysis. In *BMVC91, British Machine Vision Conference*, pages 258–265, University of Glasgow, UK, 24–26 September 1991.

[149] P Miller and S Astley. Automated detection of breast asymmetries. In J Illingworth, editor, *Fourth British Machine Vision Conference, University of Surrey*, volume 2, pages 519–528. BMVA, BMVA Press, 21–23 September 1993.

[150] P Miller and S Astley. Automated detection of breast asymmetries. In *BMVC-93 British Machine Vision Conference*, pages 519–528, University of Surrey, Guilford, UK, 21–23 September 1993.

[151] B Monsees. Thermography, transillumination light scanning, and magnetic resonance imaging. In GW Mitchell, Jr. and LW Bassett, editors, *The Female Breast and its Disorders*, chapter 13, pages 161–171. Williams and Wilkins, Baltimore, 1990.

[152] WM Morrow, RB Paranjape, RM Ranjayyan, and JEL Desautels. Region based contrast enhancement of mammograms. *IEEE Transactions on Medical Imaging*, MI–11(3):392–406, September 1992.

[153] M Moskowitz. Age-specific growth rates and screening strategies. *Radiology*, 161:37, 1986.

[154] CE Mountford, GL May, PG Williams, et al. Classification of human tumors by high resolution magnetic resonance spectroscopy. *The Lancet*, 1(8482):651–653, 1986.

[155] PA Newcomb and PM Lantz. Recent trends in breast-cancer incidence, mortality, and mammography. *Breast Cancer Research and Treatment*, 28(2):97–106, 1993.

[156] RM Nishikawa, ML Giger, K Doi, et al. Effect of case selection on the performance of computer-aided detection schemes. *Medical Physics*, 21(2):265–269, February 1994.

[157] RM Nishikawa, ML Giger, K Doi, CJ Vyborny, and RA Schmidt. Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals. *Medical Physics*, 20(6):1661–1666, 1993.

[158] B Nordenström. Electric potentials in pulmonary lesions. *Acta Radiologica*, 11:1–16, 1971.

[159] SL Olson, BW Fam, PF Winter, et al. Breast calcifications: analysis of imaging properties. *Radiology*, 169:329–332, 1988.

[160] MM Osman and EM Afify. Thermal modeling of the normal woman's breast. *Journal of Biomedical Engineering*, 106:123–130, May 1984.

[161] MM Osman and EM Afify. Thermal modeling of the malignant woman's breast. *Journal of Biomedical Engineering*, 110:269–276, November 1988.

[162] TW Pai and JHL Hansen. Boundary-constrained morphological skeleton minimization and skeleton reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intel-*

ligence, PAMI–16(2):201–208, February 1994.

[163] M Pamilo, M Soiva, I Anttinen, et al. Ultrasonography of breast lesions detected in mammography screening. *Acta Radiologica*, 32(3):220–225, 1991.

[164] J Parker, DR Dance, and J Suckling. The hysteresis technique for detection of microcalcifications in digital mammograms. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 41–50, York, England, 10–12 July 1994. Elsevier Science.

[165] EA Patrick, M Moskowitz, VT Mansukhani, and EI Gruenstein. Expert learning system network for diagnosis of breast calcifications. *Investigative Radiology*, 26(6):534–539, June 1991.

[166] T Pavlidis. *Algorithms for Graphics and Image Processing*. Springer-Verlag, 1982.

[167] S Peleg and A Rosenfeld. A min-max medial axis transformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–3(2):208–210, March 1981.

[168] ME Peters, DR Voegeli, and KA Scanlan. *Handbook of Breast Imaging*. Churchill Livingstone, New York, 1989.

[169] FW Prior, DS Channin, SH King, et al. High resoltuion digital mammography using DIMA^{TM} and CR: preliminary results. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 143–151, York, England, 10–12 July 1994. Elsevier Science.

[170] JH Richter and E Claridge. Extraction and measurement of lesion edge blur im mammograms by computer image analysis. In HU Lemke, ML Rhodes, CC Jaffe, and R Felix, editors, *Proceedings of the 5th International Symposium in Computer Assisted Radiology*, Berlin, 3–6 July 1991. Springer-Verlag.

[171] JH Richter and E Claridge. Extraction of quantitative blur measures for circumscribed lesions in mammograms. *Medical Informatics*, 16(2):229–240, April-June 1991.

[172] DF Rideout and PY Poon. Patterns of breast parenchyma on mammography. *Journal of the Canadian Association of Radiologists*, 28:257–258, 1977.

[173] S Roberts and L Tarassenko. A probabilistic resource allocating network for novelty detection. *Neural Computation*, 6:270–284, 1994.

[174] EJ Roebuck. *Clinical Radiology of the Breast*. Heinemann Medical Books, Oxford, 1990.

[175] J Rogowska, K Preston, and D Sashin. Evaluation of digital unsharp masking and local contrast stretching as applied to chest radiographs. *IEEE Transactions on Biomedical Engineering*, 35:817–827, October 1988.

[176] PR Rosen, RE Snyder, and G Robbins. Specimen radiology for nonpalpable lesions found by mammography: Procedures and results. *Cancer*, 34:2028–2033, 1974.

[177] A Rosenfeld. Connectivity by digital pictures. *Journal of the Association for Computing Machinery*, 17:146–160, 1970.

[178] A Rosenfeld and J Pfaltz. Sequential operation in digital picture processing. *Journal of the Association for Computing Machinery*, 12:471–494, 1966.

[179] RJ Ross, JS Thompson, K Kim, and RA Bailey. Nuclear magnetic resonance imaging and evaluation of human breast tissue: preliminary clinical trials. *Radiology*, 143:195–205, 1982.

[180] M Sallam and K Bowyer. Registering time sequences of mammograms using a two-dimensional image unwarping technique. In A Gale, SM Astley, DR Dance, and

AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 121–130, York, England, 10–12 July 1994. Elsevier Science.

[181] RA Schmidt, RM Nishikawa, K Schreibman, et al. Computer detection of lesions missed by mammography. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 289–294, York, England, 10–12 July 1994. Elsevier Science.

[182] GL Scott, SC Turner, and A Zisserman. Using a mixed wave/diffusion process to elicit the symmetry set. In *AVC88, Fourth Alvey Vision Conference*, pages 221–228, University of Manchester, 31 August–2 September 1988.

[183] LM Scoutt, KJ Taylor, C Lee, et al. Correlation of continuous-wave doppler detected neovascularity in breast cancers with pronosis. *Radiology*, 181(P):201, 1991.

[184] JL Semmlow, A Shadagopan, and LV Ackerman others. A fully automated system for screening xeromammograms. *Computer and Biomedical Research*, 13:350–362, 1980.

[185] MI Sezan, AM Tekalp, and R Schaetzing. Automatic anatomically selective image enhancement in digital chest radiography. *IEEE Transactions on Medical Imaging*, MI–8(2):154–162, June 1989.

[186] A Sha'ashua and S Ullman. Structural saliency: the detection of globally salient structures using a locally connected network. In *ICCV88, Second International Conference on Computer Vision*, pages 321–327, 1988.

[187] S Shapiro, W Venet, P Strax, et al. Ten to fourteen year effect of breast cancer screening on mortality. *Journal of the National Cancer Institute*, 69:349–355, 1982.

[188] Dr B Shepstone. Personal communication, May 1994. Breast Care Unit, Churchill Hospital, Oxford, UK.

[189] JH Shroff, LR Lloyd, and DM Schroder. Open breast biopsy: A critial analysis. *The American Surgeon*, 57(8):481–485, 1991.

[190] EA Sickles. Breast cancer detection with transillumination and mammography. *American Journal of Roentgenology*, 142:841–844, 1984.

[191] EA Sickles. Mammographic features of "early" breast cancer. *American Journal of Roentgenology*, 143:461–464, September 1984.

[192] EA Sickles. Breast imaging: a view from the present to the future. *Diagnostic Imaging and Clinical Medicine*, 54:118–125, 1985.

[193] EA Sickles, RA Filly, and PW Callen. Breast cancer detection with sonography and mammography: Comparison using state-of-the-art equipment. *American Journal of Roentgenology*, 140:843–845, 1983.

[194] EA Sickles, RA Filly, and PW Callen. Benign breast lesions: Ultrasound detection and diagnosis. *Radiology*, 151:467–470, 1984.

[195] KM Simpson and KW Bowyer. Comparison of spatial noise removal techniques for digital mammography. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 325–334, York, England, 10–12 July 1994. Elsevier Science.

[196] KT Smith, SL Wagner, RB Guenther, et al. The diagnosis of breast cancer in mammograms by the evaluation of density patterns. *Radiology*, 125:383–386, 1977.

[197] W Spiesberger. Mammogram inspection by computer. *IEEE Transactions on Biomedical Engineering*, BME–26:213–219, 1979.

[198] A Stamatakis, AY Cairns, IW Ricketts, et al. A novel approach to aligning mammograms.

In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 355–364, York, England, 10–12 July 1994. Elsevier Science.

[199] T Stefanelli and A Rosenfeld. Some parallel thinning algorithms for digital pictures. *Journal of the Association for Computing Machinery*, 18:255–264, 1971.

[200] EE Sterns and B Zee. Thermography as a predictor of prognosis in cancer of the breast. *Cancer*, 67:1678–1680, 1991.

[201] P Strax. Control of breast cancer: the why and how of mass screening. *Administrative Radiology*, pages 30–36, September 1989.

[202] RN Strickland and HI Hahn. Detection of microcalcifications using wavelets. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 79–88, York, England, 10–12 July 1994. Elsevier Science.

[203] J Suckling, DR Dance, DJ Lewis, and SG Blacker. Parenchymal delineation by human and computer observers. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 315–324, York, England, 10–12 July 1994. Elsevier Science.

[204] J Suckling et al. The MIAS mammographic database. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 375–378, York, England, 10–12 July 1994. Elsevier Science.

[205] R Szeliski. Personal communication, March 1994. Digitial Equipment Corporation.

[206] R Szeliski and S Lavallée. Matching 3-D anatomical surfaces with non-rigid volumetric deformations. Technical report, AAAI Spring Symposium on Applications of Computer Vision in Medical Image Processing, Stanford University, USA, 21–23 March 1994.

[207] L Tabar, G Fagerberg, SW Duffey, et al. Update of the Swedish two-country program of mammographic screening for breast cancer. *Radiologic Clinics of North America*, 30:187–210, 1992.

[208] L Tabár, A Gad, LH Holmberg, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group if the Swedish National Board of Health and Welfare. *The Lancet*, 1:829–832, 1985.

[209] PG Tahoces, J Correa, M Souto, et al. Enhancement of chest and breast radiographs by automatic spatial filtering. *IEEE Transactions on Medical Imaging*, MI–10(3):330–335, September 1991.

[210] L Tarassenko. Medical applications of neural networks. In *Neural Computing Applications Forum*. St. Hugh's College, Oxford, 30 June 1994.

[211] L Tarassenko. Neural networks: diagnostic systems application II. Lecture notes, Neural Networks Summer School, Cambridge University, 1994.

[212] N Tohnosu, K Okuyama, Y Koide, et al. A comparison between ultrasonography and mammography, computed-tomography and digital subtraction angiography for the detection of breast cancers. *Surgery Today — The Japanese Journal of Surgery*, 23(8):704–710, 1993.

[213] DY Tsai, H Fujita, K Horita, et al. Breast-tumor classification by neural networks fed with sequential-dependence factors to the input layer. *IEICE Transactions on Information and Systems*, E76-D(8):956–962, August 1993.

[214] SC Turner. Parallel algorithms for two dimensional shape representation. D.Phil Transfer Report, Robotics Research Group, Oxford University, December 1988.

[215] L Valatx, IE Magnin, and A Bremond. Automatic microcalcifications and opacities detection in digitized mammographies using a multiscale approach. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 51–57, York, England, 10–12 July 1994. Elsevier Science.

[216] ALM Verbeek, R Holland, F Sturmans, et al. Reduction of breast cancer mortality through mass screening with modern mammography. *The Lancet*, 1:1222–1224, 2 June 1984.

[217] L Wang, PP Ho, X Liang, et al. Kerr-fourier imaging of hidden objects in thick turbid media. *Optics Letters*, 18(3):241–243, 1993.

[218] LM Wang, PP Ho, and RR Alfano. Double-stage picosecond Kerr gate for ballistic time-gated optical imaging in turbid media. *Applied Optics*, 32(4):535–540, 1993.

[219] LM Wang, PP Ho, C Liu, et al. Balistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate. *Science*, 253:769, 16 August 1991.

[220] W Wee, M Moskowitz, NC Chang, et al. Evaluation of mammographic calcifications using a computer program. *Radiology*, 116:717–720, 1975.

[221] SR Wellings. Development of human breast cancer. *Advances in Cancer Research*, 31:287–314, 1980.

[222] F Winsberg, M Elkin, J Macy, et al. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89:211–215, August 1967.

[223] JN Wolfe. Mammography: ducts as a sole indicator of breast carcinoma. *Radiology*, 89(2):206–210, August 1967.

[224] JN Wolfe. A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease. *Radiology*, 89(2):201–205, August 1967.

[225] JN Wolfe. Breast patterns as an index of risk for developing breast cancer. *American Journal of Roentgenology*, 126:1130–1139, 1976.

[226] JN Wolfe. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37:2486–2492, 1976.

[227] JN Wolfe, S Albert, S Belle, et al. Breast parenchymal patterns: analysis of 322 incident breast carcinomas. *American Journal of Roentgenology*, 138:113–118, 1982.

[228] KS Woods and KW Bowyer. Computer detection of stellate lesions in digital mammograms. In A Gale, SM Astley, DR Dance, and AY Cairns, editors, *2nd International Workshop on Digital Mammography*, Excerpta Medica International Congress Series 1069, pages 221–229, York, England, 10–12 July 1994. Elsevier Science.

[229] DQ Wu. Morphological filters in image analysis. First Year Report, Robotics Research Group, Oxford University, 1992.

[230] Y Yasuoka and RM Haralick. Peak noise removal by a facet model. *Pattern Recognition*, 16(1):23–29, 1983.

[231] FF Yin, ML Giger, K Doi, et al. Computerized detection of masses in digital mammograms: analysis of bi-lateral subtraction images. *Medical Physics*, 18(5):955–963, September/October 1991.

[232] FF Yin, ML Giger, K Doi, et al. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effect on bilateral-subtraction

technique. *Medical Physics*, 21(3):445–452, March 1994.

[233] S Yokoi, J Toriwaki, and T Fukumura. An analysis of topological properties of digitized binary pictures using local features. *Computer Graphics and Image Processing*, 4:63–73, 1975.

[234] SJ el Yousef, RH Duchesneau, RJ Alfidi, et al. Magnetic resonance imaging of the breast. *Radiology*, 150:761–766, 1984.

[235] X Zhou and R Gordon. Geometrical unwarping for digital subtraction mammography. In *Proceedings of Vision Interface '88*, pages 25–30. Canadian Image Processing and Pattern Recognition Society, 1988.