An Automated Image Analysis System for the Detection of Microcalcifications

by

S. A. Hojjatoleslami

Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Surrey University

Centre for Vision, Speech and Signal Processing School of Electronic Engineering, IT and Mathematics University of Surrey Guildford, U.K.

September, 1997
To my supportive family,
my patient wife and
my daughter Nilufar
Abstract

The interpretation of medical images is one of the most difficult tasks in computer vision, largely because of the high degree of variability associated with normal and abnormal appearances. This thesis introduces a systematic method for the detection of microcalcifications as one of the most important signs of early breast cancer. It involves a four step procedure. The first step is blob detection to detect regions of microcalcification size range. The second step involves a specially designed directional region growing method to find the best fitting boundaries for each blob region. A newly developed combination of classifiers is then applied to label each region as a microcalcification or background. The final processing step involves a search for the existence of clusters of microcalcifications using a hierarchical nearest mean clustering method.

The contributions of the work to the field of image processing are; a new blob detection system; a novel region growing method and a theoretical framework for combining classifiers which use a combination of shared and distinct representations. Here specifically, we present a blob detection method with the capability of detecting any suspected blob of specific size range. Then a new region growing method is developed based on a unique directional growing process providing predictable behaviour for the method. The application of two discontinuity measures is considered for the extraction of two fitting boundaries representing information about the region and its local background. The information conveyed by the boundaries and their associated regions is used to compute reliable representations for labelling each blob region. The robustness of the region growing method to the choice of a starting point and to Gaussian noise is examined on real images. We demonstrate that commonly used classifiers provide reliable results in labelling the suspected regions.

In spite of achieving an acceptable performance using different individual classifiers, a decision fusion rule involving a weighted combination of classifiers is developed and its performance on the problem is investigated. The combination rule is applicable when mixed mode representations (some shared and some individual features) are used. A comparative study of the individual classifiers and also of conventional classifier com-
bination techniques with the weighted combiner is performed on independent test sets.

The results achieved with the presented algorithm are very promising and approaching a level where a clinical pilot evaluation for screening purposes would be warranted.
Acknowledgement

I would like to express my deepest appreciation and gratitude to my supervisor Professor Josef Kittler for his invaluable guidance and encouragement. I am glad to have worked with him and I have benefited tremendously from his knowledge and experience in theoretical and technical matters. I especially thank him for his patience during my research.

I would like to thank the Ministry of Culture and Higher Education (MCHE) of Iran for their financial support during this research.

I would like to thank all the members of the Centre for Vision Speech and Signal Processing, for providing a nice and friendly environment by giving advice, help and encouragement during my studies. In particular: the assistance of my colleague Dr. Lucia Sardo for providing the package and required information for the RBF neural network used in this study; Dr. Radek Marik, Dr. Majid Mirmehdi, Miss. Sarah Morgan, and Dr. G. Matas for their useful discussions along the way.

I am very grateful to Dr. Julie Cooke from the Jarvis Breast Screening Centre, Guildford and Dr. Kenneth C. Young from the National Coordinating Centre for Physical Mammography, Guildford for spending their valuable time to re-annotate some images in the MIAS database.

Finally, I would like to record my deep gratitude to my family for their support during my studies here in the UK. Especially my father for his encouragement to study for a PhD and my wife, Nasrin, and my dear daughter, Nilufar, for their patience and support during this period.
# Contents

1 Introduction 1
   1.1 Microcalcifications ................................................................. 3
   1.2 Computer Based Analysis .......................................................... 4
   1.3 Required tools for computer interpretation .............................. 5
       1.3.1 Images and Data Set ....................................................... 6
       1.3.2 Receiver Operating Characteristics .............................. 7
   1.4 Scope of this research .............................................................. 8

2 Literature Survey 14
   2.1 Conventional techniques .......................................................... 14
   2.2 Multi-resolution approaches .................................................... 17
   2.3 Other Techniques ................................................................. 19

3 Blob Detection Method 25
   3.1 Image Segmentation: A Brief Overview .................................... 25
       3.1.1 Segmentation Techniques Based on Discontinuity .......... 27
       3.1.2 Segmentation Techniques Based on Similarity ............. 28
           3.1.2.1 Measurement Space Clustering ......................... 28
           3.1.2.2 Region Growing Methods .......................... 31
           3.1.2.3 Spatial Clustering ........................................... 32
           3.1.2.4 Split and Merge ......................................... 32
       3.1.3 Conclusion ............................................................... 33
   3.2 A new approach to Blob Detection ......................................... 34
       3.2.1 Morphological filters ..................................................... 34
           3.2.1.1 Top Hat Transform ........................................... 34
       3.2.2 Thresholding ................................................................. 35
   3.3 Conclusions ................................................................. 36
# A New Region Growing Method

4.1 Growing Process ......................................................... 42
4.2 Discontinuity Measures .................................................. 45
   4.2.1 Average Contrast Measurement .............................. 46
   4.2.2 Peripheral Contrast Measurement ....................... 46
4.3 Test on Gaussian Shape Image ...................................... 48
4.4 Starting Point ............................................................. 51
4.5 Experimental Results .................................................. 51
   4.5.1 Textured Image ......................................................... 51
   4.5.2 Experiments on Medical Images ......................... 53
4.6 Adaptation of the Method for MC Boundary Extraction ...... 56
4.7 Comparison with the Centroid Linkage Region Growing ...... 59
4.8 Conclusions ................................................................. 60

# Pattern Recognition

5.1 Feature Selection ......................................................... 75
   5.1.1 Problem Formulation ................................................. 75
   5.1.2 Sequential Forward/Backward Search Methods ............ 76
   5.1.3 Plus I Take Away r Algorithm ................................. 77
   5.1.4 Sequential Forward Floating Selection and Sequential Backward Floating Selection Methods ............. 78
5.2 Classification .............................................................. 81
   5.2.1 K-Nearest Neighbour Classifier ............................... 81
      5.2.1.1 Parameter Estimation for the K-NN .................. 82
   5.2.2 Gaussian classifier .................................................. 85
   5.2.3 Multi-layer Perceptron ............................................. 85
   5.2.4 Radial basis functions network ............................... 88
5.3 Experimental Method .................................................. 89
   5.3.1 Feature selection ..................................................... 89
   5.3.2 Classification .......................................................... 91
5.4 Performance Measures ................................................ 91
   5.4.1 Clustering .............................................................. 92
5.5 Experimental Results .................................................. 92
5.6 Conclusions ............................................................... 95
List of Figures

1.1 Samples showing mammographic image patches of normal dense (a) and normal fatty breasts (b). ............................................................... 2

3.1 Histogram of normal and abnormal mammograms. ................................. 29
3.2 Co-occurrence matrix of normal and abnormal mammograms .................. 30
3.3 The results of successive steps of our segmentation method on a mammogram. .................................................................................. 37

4.1 Topographical surface of a microcalcification and its grey level mapping during the growing process................................................. 43
4.2 Schematic graph illustrating notations used for the description of our region growing method. .............................................................. 45
4.3 Some of the commonly used gradient operators. ...................................... 48
4.4 The segmentation results of a Gaussian shape image with $\sigma = 25$. .............. 50
4.5 The segmentation results of a part of road centre line. .............................. 63
4.6 Different mappings during the growing process for two starting points. ...... 64
4.7 Segmentation results of an MR image of head. ......................................... 65
4.8 The mappings for brain stem during the growing process. ......................... 66
4.9 Segmentation results of MR image of head. ............................................. 67
4.10 Segmentation error rate of the regions in the MRI image for different levels of Gaussian noise (image range is 0 — 255). ................................. 68
4.11 The segmentation results of holes in the brain. ........................................ 68
4.12 Segmentation results for an individual microcalcification. ....................... 69
4.13 Segmentation results for two neighbouring microcalcifications. ............... 70
4.14 Part of a mammogram containing microcalcifications and its annotation. ...... 71
4.15 Segmentation result of image by the centroid linkage region growing method. 72
4.16 Segmentation result of image by the proposed region growing method. ...... 73

5.1 $K$-NN classifier performance versus $K$ and $m$. ................................... 83
5.2 A Feed-forward Network with one hidden layer. ................................. 86
5.3 Structure of a single neuron in the MLP neural network. ......................... 87
5.4 Block diagram of the RBF neural net ....................................................... 88
5.5 The use of Bayes decision rule for decision making using the RBF classifier. 89
5.6 The results of SFFS and SBFS feature selection methods. ......................... 90
5.7 ROC curve comparison of performance of different classifiers in the clusters of microcalcifications detection. ................................. 93
5.8 Outliers detected by the RBF classifier. .................................................... 94
5.9 Results produced using the MLP classifier on a mammogram. ................. 100

6.1 Weighted averaging fusion of multiple expert opinions. ......................... 110
6.2 Correlation matrix of features used by different classifiers. ..................... 111
6.3 Results produced on a full size abnormal mammogram .......................... 120
6.4 Results produced on abnormal part of the mammogram. ......................... 123

7.1 Schematic diagram of the presented method. ........................................... 125
List of Tables

6.1 Errors produced on the two independent test sets. ............................................ 112
6.2 The performance of the classifiers and their equally weighted average combi-
nation when only the shared features are used for classification purpose. 113
6.3 Errors produced using equations (6.8) and (6.9) on the two independent
data sets. .............................................................................................................. 113
6.4 Errors when conventional classifier combination techniques are applied. . . 114
6.5 The best results for cluster of microcalcification detection achieved using the
weighted average combiner. ............................................................................. 115
6.6 The best results for image identification achieved using the weighted average
combiner. .......................................................................................................... 115
A.1 List of features used for microcalcification detection. ................................. 132
Chapter 1

Introduction

This introduction presents a brief description of breast cancer and explains the motivation and the importance of automatic microcalcification detection in breast screening. It describes the major points which must be considered when a computerised system is designed for breast screening. The final step of this introduction presents an outline of the thesis.

Breast cancer is one of the commonest cancers affecting women in the UK, Western Europe and North America. Approximately one American woman in 9 will develop breast cancer at some time during her life. Breast cancer is second only to lung cancer as a cause of cancer-related death among women. Its incidence still appears to be increasing.

The etiology of breast cancer is unclear and no single dominant cause has been discovered yet. As current methods of treating breast cancer are effective in its early stage of development, early detection can increase the survival rate [5, 21]. Mammography is currently considered to be the best method for early detection of breast cancer. It is used for breast cancer screening in the UK, USA, Sweden, Holland and other countries [3]. The UK screening program generates more than 3 million mammograms per year. Screening is performed every three years for an individual by examination of both breasts using a single view, the medio-lateral or middle to outside view. Some radiologists employ two radiological views, the second one being taken at a different angle.

The most recent data indicate that death rates from breast cancer have begun to decline. The decreases are attributed to earlier detection and improved treatment of breast cancer.
One of the most important parameters affecting the results of the screening program is the performance of the radiologists, in interpreting appropriately the visual information presented on the mammograms. The probability of observer error is high, due possibly to the incidence of cancer being very low (less than 1.5 per 100 women [4, 3]) and the wide variation of normal breast tissue structures together with repetitive examination of mammograms, sometimes with a magnifier, leading to fatigue and subsequent loss of performance [1]. These can be the main causes of poor accuracy in diagnosis, more than 8% of cancers are missed and more than 70% of open surgical biopsies in actuality are benign [22].

To illustrate the large variation of the background in mammograms, sections of two different normal mammograms are shown in Figure 1.1. This figure shows the large variation in texture and intensity of a normal background. As various abnormalities found in mammograms are superimposed on the normal variations, the problem of abnormality detection is extremely difficult [2].

Radiologists look for certain signs and characteristics indicative of cancer when eval-
1.1 MICROCALCIFICATIONS

Clinically unsuspected and non-palpable cancers may be diagnosed on the basis of microcalcifications. A microcalcification is a tiny calcium deposit, often found in clusters, in the breast tissue which appears as a small bright spot in the mammogram. They may be found within a malignant mass or at a distance from the mass. They may be circular or elongated, tubular and branched in the line of the lactiferous ducts with various sizes ranging from less than 0.1\(mm\) up to 5\(mm\) in diameter. On close inspection, these minute calcifications are seen to be irregular in outline [24]. The individual microcalcifications appear in various shapes, eg. circular, bin shapes, the so called \(x, y, z\) shapes and others as classified by Lanyi [19]. Although some differentiating signs between benign and malignant microcalcifications are recognised, e.g. the benign calcifications are usually larger in size than those found in malignancy and their margins are normally more regular than the malignant ones, nevertheless, distinguishing between the two different kinds is found to be difficult [1, 20].

The appearance of three or more microcalcifications within a square centimeter region, known as a cluster, is clinically very significant. According to Lanyi [19], certain patterns of microcalcifications appearing as a cluster are considered to be significant signs of cancer, however, these are not considered as reliable signs of malignancy because their incidence may vary in shape and appearance based on the different projections used [1].

The large variety of microcalcification signs (size, coarseness, smoothness, shape, clustering and distribution) are significant diagnostic indicators for radiologists. It follows, therefore, that these factors should be taken into consideration when designing a computerised detection system, particularly where modelling for microcalcifications is required.
1.2 Computer Based Analysis

Two different scenarios have been suggested for computer-aided diagnosis (CAD) techniques: pre-screening and prompting [26].

The pre-screening techniques aim to pre-read the mammograms and refer to radiologists only those mammograms detected as being abnormal. These systems are designed to reduce the number of normal images in the set so that only a small percentage are referred to the radiologists for further investigation. In the United Kingdom, the breast screening age range is 50 – 65 yrs where the incidence of detected cancers is very low, at approximately 1.5 per 100 women, and the aforementioned techniques can greatly reduce not only the physicians time but also their error rate which in part is caused by investigating the huge number of images [7]. Such a system may be used for breast cancer pre-screening thereby referring only the suspected cases of cancer to the radiologists. Should such a system be advocated, it could be made easily accessible to the public since even non-specialist radiologists can apply this computerised system which in turn, can provide a greater speed of breast cancer detection by a reduced number of specialist radiologists. This would provide a higher rate of accuracy and a huge reduction in the screening costs. Naturally, the reliability of such a system needs to be very consistent and must be capable of detecting all kinds of abnormalities rather than microcalcifications alone.

In the prompting techniques the computer acts as a second reader, resulting in a reduction of the number of expert radiologists where double reading is the current practice [26]. These techniques should be capable of prompting all kinds of abnormalities and should be sufficiently robust to provide the second opinion for an expert radiologist in analysing mammograms. They should perform at least as well as an experienced radiologist in order to improve the observer’s performance and in such a case, the necessity for double reading will be reduced. In this scenario, the number of abnormal images to be considered by the technique is very high and the aim is to detect as many abnormalities as possible with as low a false positive rate as an expert radiologist. If the false positive rate is high, the entire system will be ignored as a useful diagnostic tool by radiologists.

The main difficulty in CAD is testing the system to demonstrate its capability of per-
forming to an acceptable standard. Such a test should be performed on a dataset matched to the real situation which is important in order to get a true feeling for the results that could be expected in practice. The real situation is not the same for the two different scenarios. For pre-screening, the test data should be representative of that generated by the screening program. In this scenario the important characteristic of the data is that the vast majority of mammograms are normal [1], therefore the number of normal images should be much higher than the number of abnormal images. To test computer-aided prompting techniques, a database containing a higher number of abnormal images in comparison to the normal images is required. Such a dataset can be prepared for individual abnormalities and used to test the method designed to detect that specific abnormality. However, as a final test, a combination of the different methods can be used to detect all the various abnormalities from the mixed datasets.

It is quite obvious that both techniques focus on the same problem with different aspects offering different figures of merit which must be considered. These figures for computer-aided microcalcification detection systems are known as “image identification” and “cluster of microcalcification detection”. Image identification is the important parameter for the computer-aided pre-screening system while the microcalcification detection or more precisely, the cluster of microcalcification detection figure is the important parameter to be considered for the computer-aided prompting system. These issues suggest two performance measures to be considered as introduced in Section 5.4.

1.3 Required tools for computer interpretation

When the performance of the system is being considered, the main question is how to test the method. The main points to be addressed relate to the database (including imaging method, digitisation and ground-truth) used for the experiments, and how the results are analysed. The accepted method of presenting the results is based on the receiver operating characteristic (ROC) curve analysis which is described in Section 1.3.2.

The database should satisfy all the following requirements:

- The required properties for the imaging process, as addressed by many researchers [6,
1.3. REQUIRED TOOLS FOR COMPUTER INTERPRETATION

8).

- To include a complete annotation for every abnormality (individual microcalcifications and clusters for our case).

- Must be digitised with high resolution sufficient to represent the small size abnormalities, like microcalcifications.

The second item is important, as a perfect annotation for different abnormalities is required to enable researchers to perform a statistical analysis on the pre-annotated images providing quantitative results, rather than to present their results based on visual evaluation giving a qualitative judgement. This is very important as the opinion of radiologists may be affected by viewing the outcome of the algorithm which could precipitate a biased judgement.

A spatial resolution of at least 50\(\mu m/sample\) is required to represent small microcalcifications which may appear in less than 0.1\(mm\) in diameter. Such a resolution will produce about 4 pixels for microcalcification size blobs (size 0.1\(mm\)).

Only three databases containing clusters of microcalcifications (MIAS, Neijman and USF)\(^1\) are available to the public so far. For the purposes of this research we used the MIAS, as described in the next section, which satisfies most of the required properties to test the performance of our algorithm.

1.3.1 Images and Data Set

As mentioned in the previous section, a digital mammography database produced by the Mammographic Image Analysis Society (MIAS) is used in this study [23]. The database collected by the United Kingdom National Breast Screening Program contains 207 normal images and a large variety of abnormalities including, 25 microcalcification mammograms. Each microcalcification image contains at least one biopsy-proven cluster of microcalcifications. The mammograms are digitised to a spatial resolution of 50 microns per

\(^{1}\)Information about the databases are available on http://marathon.csee.usf.edu/Mammography/Database.html site.
1.3. REQUIRED TOOLS FOR COMPUTER INTERPRETATION

pixel with grey level resolution of 8 bits per pixel. The location of each cluster of microcalcifications is specified by the \(x, y\) coordinates of the centre of abnormality and the radius of a circle covering the cluster. Since the size of each image in the database is very large, from 6.9 Mbyte to 20.8 Mbyte depending on the size of breast and its projection on the film, we selected a part of each normal mammogram, covering more than 80% of the breast tissue for our extensive experiments.

Our local radiologists considered the digitised images using the 'xv' viewer on a high quality monitor. They considered every abnormal image by changing the contrast of images to verify the supplied annotation. They labelled three of the benign microcalcification images as normal and two of the normal images as suspected of being benign. Since the MIAS database is annotated before the mammograms are digitised, this inconsistency may be due to a mis-adjustment during the digitisation. In order to avoid any controversy, we excluded the 5 images from the database and used the rest (22 microcalcifications and 205 normal images) with the MIAS annotations in this experiment.

1.3.2 Receiver Operating Characteristics

The accuracy of a detection system can be characterised by its receiver operating characteristic (ROC) curve. The ROC is the single analytical technique known to provide both a useful performance accuracy index, and a basis for assessing the usefulness of the classification procedure in terms of cost and benefit [25].

The ROC curve is a plot of true positive (TP) versus false positive (FP) detection rates. It demonstrates the variation of the two probabilities (TP versus FP) when a parameter in the diagnosis (or recognition) system is changed [25]. Therefore it also helps to choose the parameter for the required performance (TP/FP). For our application, the ROC curve is used to consider the performance of a classifier by plotting TP versus FP for various a-priori probabilities.

For our application, two ROC curves have been produced for the two figures of merit: "image identification" and "cluster of microcalcifications detection". The ROC for "cluster of microcalcifications detection" is produced by plotting the percentage of the TP clu-
1.4. SCOPE OF THIS RESEARCH

The accuracy of a classifier detecting clusters of microcalcifications. Similarly, an ROC curve is produced to assess the performance of a classifier for image identification which is the plot of the TP image rate (the percentage of abnormal image identification based on the correctly detected clusters) versus the FP image rate (percentage of normal images detected as abnormal).

1.4 Scope of this research

A number of attempts aimed at the design of a system for automatic interpretation of mammograms has already been reported, most of which focus on microcalcification detection. In so far as the reported results are concerned (see Chapter 2) none of these techniques have been thoroughly tested, clinically.

Our aim is to design a system for the detection of microcalcifications to be used as a part of a CAD system for digital mammography. Before designing such a system, we considered conventional image processing techniques and designed various processing sections to match with the properties of microcalcifications. The system here involves four different processing steps; segmentation, feature selection, classification and clustering in the spatial domain.

This thesis contains several novel contributions to the field of image processing and pattern recognition ([9] – [18]). These contributions include: i) a novel blob detection technique with the capability of detecting all the small bright blobs in a texture background. ii) a novel region growing method with the capability of outlining any bright/dark region in a textured background. iii) a new approach to the combination of classifiers which guarantees a higher performance in comparison to the best classifier and conventional classifier combination techniques. The outline of the thesis is as follows:

Chapter 2 presents a review of the literature addressing the problem of microcalcification detection. Existing algorithms for detecting microcalcifications are cited and the results are given. However the reported detection rates vary and these should not be compared because the results have been obtained on different data sets.
1.4. SCOPE OF THIS RESEARCH

Chapters 3 and 4 are concerned with the problem of segmentation of microcalcifications. Chapter 3 presents a novel blob detection method for the detection of small blobs in the image. The technique employs nonlinear filters, top-hat transform and median filter, to segment the grey level image into binary regions representing suspected blobs. Top-hat transform is applied to elaborate all the blobs in the microcalcification size range and adaptive thresholding based on the median of a local neighbourhood is applied to segment the bright regions appearing in the transformed image.

Chapter 4 describes a novel region growing technique outlining two boundaries for each suspected region seeded by the blob detector. Our region growing method, like other region growing techniques, starts from a point which meets a detection criterion and grows through high grey level pixels into the background. The method considers two different discontinuity measures representing important characteristics of the region as a function of its evolving boundary during the growing process. Once a coarse stopping criterion is satisfied, a reverse test is applied to find a unique boundary associated with the highest discontinuity measure for the region. Experiments have been performed both on synthetic and real images to evaluate this new approach. The main strengths of the method are its ability to segment out from a textured background, a locally bright/dark region with fuzzy boundaries, as well as its simplicity and immunity to global intensity shifts.

Then a set of 39 measurements is computed from the two regions and their associate boundaries of each segmented region. The list of the derived measurements is given in Table A.1 in the Appendix.

Chapter 5 considers the application of pattern recognition to reduce the number of falsely detected regions. This consideration includes the use of feature selection techniques to select the best subset of features out of the available measurements and the application of different pattern recognition techniques to label the detected regions as microcalcification or normal background. The floating search feature selection methods are applied to choose the best set of features by maximising the performance of two simple
classifiers. This chapter continues by outlining the application of four different classifiers (multi-layered perceptron (MLP), radial basis function (RBF) neural network, \(k\)-nearest neighbour with locally optimum metric and Gaussian classifiers) to be used for labelling the patterns as either microcalcification or normal background. The chapter also considers the performance of the classifiers on independent images in the MIA S database and compares their capabilities for the problem at hand. The capability of the classifiers for the detection of outliers, regions which are neither microcalcifications nor normal background, are also presented.

In chapter 6, we refer to recent studies suggesting that the object labelling performance can be improved by means of combining the opinion of individual experts, which is analogous to using the opinion of several specialists instead of only one, to finalise the decision.

Further to the aforementioned studies, we present a theoretical framework for the combination of classifiers which uses a mix of distinct and shared representations. We start by developing fusion strategies for this mixed mode data. We show that strategies are defined in terms of both the opinions of the classifiers which use the shared representation alone, and the decision outputs of experts each employing the union of the classifier specific and the shared representations. Weighted fusion strategies are then derived by taking the confidence of the individual experts into account. We show that substantial gains in performance can be achieved by fusing the opinions of multiple experts by using the methodology developed.

Each chapter provides a conclusion on individual processing steps. Chapter 7 presents final concluding remarks on the system, including a summary of the technique and outlines further areas of research to be carried out in the future.

References

REFERENCES


Chapter 2

Literature Survey

Not surprisingly, the problem of computerised microcalcification detection has received considerable interest in the literature [1, 18, 27], where the use of various techniques ranging from image analysis to decision making, have been addressed. Judging from the literature on computerised mammography techniques, acceptable results have been reported, but these have not yet been adequately tested.

Several computer-aided diagnosis (CAD) schemes using digital image processing techniques have been presented for detecting microcalcifications. Typically most of the work is focused on the design of CAD systems to improve the performance of radiologists by providing a second opinion, i.e. computer-aided prompting. These methods exploit one or more of the specific characteristics of microcalcifications [2, 1, 3, 6, 7, 8, 13].

Most of the techniques developed for medical image processing and especially mammographic image analysis are quite specific and appear to have no equivalent in non-biomedical applications [25].

2.1 Conventional techniques

A group of methods based on conventional image processing techniques are reported by researchers [2, 3, 6, 7, 13, 20].

Davies and Dance [6] applied a local thresholding method based on a histogram mode examination to detect suspected calcifications and then a feature analysis was performed to reduce the number of false positives. A clustering technique was incorporated to group
2.1. CONVENTIONAL TECHNIQUES

the detected regions into a cluster. Their method was tested successfully on 50 test images, half of which contained clusters of microcalcifications. They reported detecting 47 out of 49 clusters with a total of 9 FP clusters. However, the technique performed quite well on well-defined calcifications but they could not achieve acceptable results when tested with subtle cases [13]. They presented a new pre-processing technique by employing a hysteresis thresholding technique, in contrast to the earlier work using the histogram mode separation technique, followed by a region growing technique [13].

Woods et. al. [28] used local thresholding and region growing to segment candidate pixels followed by the application of pattern recognition techniques to classify the regions as calcification or non-calcification. The maximum and minimum grey levels in a local neighbourhood were used to segment the image. Then a region growing technique was performed to group pixels into objects. They considered 3% of the total number of segmented objects with the highest contrast, as the suspected regions. Then six classifiers, including two Bayesian (Linear and Quadratic), a K-Nearest Neighbour and three neural network classifiers, were considered for classifying the suspected regions based on seven representative features. In their experiments, the simple Linear and Quadratic classifiers performed better for both individual calcification detection and cluster detection. They reported 17 true detections out of 18 clusters with 3 false positive clusters on 15 test images. In their more recent experiment [26], they reported results on the MIAS database using an updated technique when 50% of the normal and abnormal images were used for training and the rest for testing the classifiers and vice versa. The method achieved a cluster detection rate of 80% at the cost of approximately one false positive cluster per image.

Shen et. al. [20] presented an algorithm based on an especially designed region growing technique, called multi-tolerance region growing. They used all pixels with grey level values exceeding a threshold based on the mean of the whole image as seed pixels for the region growing algorithm [21]. Two thresholds based on the maximum and minimum pixel values in the region being grown were used to specify a tolerance level to stop the growing process and calculate a set of features (including shape, compactness, centre of
2.1. CONVENTIONAL TECHNIQUES

The normalised distance between features at the successive tolerance levels was computed and the region with the minimum distance variation was selected as the final set. Among the segmented regions, those within a pre-specified size range and contrast were categorised as microcalcifications. The performance of the method as tested on four mammograms (two benign and two malignant) was 81% TP with zero FP on benign calcifications and 85% TP with 29 FP on malignant ones.

Nishikawa et al. [11] developed an automated system (containing special purpose hardware and software) for the detection of clustered microcalcifications and masses. The detection of clustered microcalcifications is accomplished in three processing steps. First the suppression of background to increase the signal to noise ratio is accomplished by applying difference image techniques (matched and box- rim filters). Then microcalcifications are segmented and relevant features are extracted for classification purposes. Using an ANN, the performance of their algorithm was 85% TP cluster detection versus 1.5 false cluster detection per mammogram. They reported a higher performance (approximately 87% TP versus 0.5 FP cluster per image) using an updated technique where 78 images were utilised, half of which contained at least one cluster of microcalcifications [12].

Dengler et al. [7] used an algorithm relying on a Gaussian filter to segment blobs of microcalcification size range and morphological filters to extract the shape of the detected blobs. They did not apply any pattern recognition technique to increase the performance. The algorithm was deemed successful based on a subjective test on 25 microcalcification images. The technique achieved a sensitivity of 97% with a specificity of about 70% in individual microcalcification detection.

Bankman et al. [3] used features of iso-intensity contour maps of the image as a basis for microcalcification detection. The contour maps are obtained by global thresholding using different thresholds. In a contour map, each micro-structure (peak) is represented by a nested set of contours. The sequence of features produced for each peak during the mapping is used to find clusters of microcalcifications. The features used were called departure (maximum second derivative of area sequence in the lower half of the peak), prominence (the number of contours above a threshold relative to the local con-
2.2. MULTI-RESOLUTION APPROACHES

Contrast, steepness (first derivative area), distinctness (the number of contours between the tip of the peak and the level where its contour merges with that of the nearest peak), and compactness (compactness of the contour obtained one level above the level where the first merging occurred). The performance of their algorithm tested on only 9 mammograms was 95% TP versus 0.22 FP clusters per mammogram.

Zheng [32] applied features computed on both spectral and spatial domains on sliding windows for classification purposes using a specially designed neural network, where the images were initially enhanced using a wavelet transform. The performance is assessed on 30 images, including 20 biopsy proven microcalcification clusters and 10 normal images, using a modified cross-validation method. He reported a more promising result of 90.1% TP rate with an average of 0.71 FP cluster per image.

2.2 Multi-resolution approaches

The application of multi-resolution techniques to digital mammography is at its early stage but has already shown significant advantages. Methods based on multi-resolution techniques, such as wavelet transform, originally developed in the field of signal processing and have recently been proposed for image processing [17, 19, 24, 25, 29]. In multi-resolution techniques, having a higher number of levels in comparison to conventional techniques produces a wider opportunity to focus on more specific features of the image.

The general approach in the wavelet techniques is to compute the forward wavelet transform of the image, to transform the wavelet coefficients using adaptive or nonlinear weighting functions and finally, to compute the inverse wavelet transform. The result of such techniques is an image which is enhanced, based on the wavelet function used. Typically the detection is either performed on the transformed image, in the second stage of the algorithm, or on the enhanced image.

Qian et. al. [14] applied a multistage structured nonlinear filter, which uses several centrally weighted median filters, to enhance microcalcifications, followed by a wavelet technique to extract features relevant to microcalcifications. The weighted median filter is accomplished by applying a set of straight line and curved shape windows. The enhanced
image is processed by a tree-structured wavelet transform to produce relevant features for microcalcifications. In tree-structured wavelet transform, separable wavelets are obtained from products of one dimensional half-band (low-pass and high-pass) filters. The wavelet bases used in their experiments were biorthogonal B-spline functions proposed by Cohen et. al. [5]. A subjective evaluation on the results of their algorithm on a set of 15 abnormal images showed 0.1 – 0.2 FP microcalcification clusters per image for 100% TP detection. In a similar experiment, they employed a three channel quadrature mirror filter [16] as compared with the results produced by the wavelet technique. They achieved a performance of 0.6 FP clusters per image on the same dataset when all the clusters were labelled correctly.

In a later work [17], they advocate enhancing the visibility of microcalcifications by the application of a hybrid architecture that includes an adaptive multistage nonlinear filter and a tree structure wavelet transform. One of five filters (a linear filter, three nonlinear trimmed-mean filters with different window sizes and a median filter) is chosen according to a local measure of variance of the original image [15]. Then a tree structure wavelet transform is applied to enhance microcalcifications further. If the measurement value is high, a smoothing filter with smaller window size is chosen and, therefore, the effects of the filtering stage will reduce and the output will be very close to the original input image. If it is low the effect of the filtering stage will increase by selecting a mean filter with wider window sizes and, therefore, increasing the smoothing effect. As a result of such a process, it is expected that a group of small structures like microcalcifications will be enhanced and low variation background is smoothed. Then an adopted tree structured wavelet transform is used to produce two separate images relevant to microcalcifications and background. Finally a back-propagation neural network is employed to classify the enhanced image. The neural network classifies all of the sliding windows of size $6 \times 6$ to detect microcalcifications. Using the algorithm [17], the true positive detection rate was 90%, with a relatively low false positive rate of 0.85 clusters per image using a set of 30 full mammograms, including 20 abnormal and 10 normal mammograms.

Yoshida et. al. [30] used wavelet transform to enhance microcalcifications and thresh-
2.3. OTHER TECHNIQUES

Olding to segment microcalcifications from the reconstructed image. They used a set of weighting factors for each level of the wavelet decomposition to enhance clusters of microcalcifications. The weighting factors are defined using a jackknife method to minimise a cost function defined by the difference between the segmented image and its annotation. Their method achieved a sensitivity of 90% and a specificity of 80% when it was tested on a set of normal and abnormal image patches with the size of $12.8 \times 12.8 \text{mm}^2$. Their updated method achieved a sensitivity of 92% and a specificity of 75% when it was tested on 41 abnormal and 41 normal image patches [29] using a jackknife method.

Strickland [22, 23] employed a wavelet transform which acts as a bank of multi-scale matched filters for detecting microcalcifications. In their work, they used a Markov texture model for background noise and a Gaussian model, adapted for microcalcifications. Like Qian [17], the wavelet bases used by Strickland were biorthogonal B-spline functions proposed by Cohen et. al. [5]. The performance of the method on a set of 40 abnormal mammograms was 90% true positive with one false cluster detection per image. Strickland [24] updated the method by employing a linear classifier to specify the weights for the wavelet coefficients. The method could detect less than 90% of clusters when one false positive cluster per image was detected on 38 images from the Neijmen database. He, like Richardson [19], concluded that components in the lowest octaves are more useful in enhancing and detecting the microcalcifications.

Yu et. al. [31] used neural networks to classify features extracted in different layers of wavelet transform. The reported results on 40 image patches was approximately 90% true positive versus 0.25 false positive image patches.

2.3 Other Techniques

Other efforts based on morphological image processing [2, 4, 10], Bayesian statistics [2] and a Markov random field model [8] are also reported, with promising results. Morphological filters are used by Astley and Taylor [2] to extract regions which possess the two main properties of microcalcifications, namely of being bright regions of small size and having sharp edges. In particular they use the response of a morphological top-hat oper-
ator and a morphological edge detector to categorise each region. Their method achieves a 97% true positive versus approximately a 2% false positive rate in 900 microcalcifications. Their more recent considerations [1] suggest that other cues are required to provide a more robust system for the detection of microcalcifications. The main reason derives from the fact that morphological top-hat and gradient filters locate any bright region smaller than the size of the structuring element, so the noise caused by the film granularity, spike noises and inhomogeneous tissues, may also provide a high response to the filters and, therefore, can be segmented out.

Betal [4] presented an algorithm based on grey level morphological operations to detect microcalcifications. He used the top hat transform and the morphological gradient to find peaks and edges for a watershed algorithm. The primary results are quite promising, however, the performance of his algorithm is not examined on a database. More recent efforts based on the watershed algorithm and a classification technique is reported by Braga Neto et. al. [10].

Karssemeijer [8] reported an iterative segmentation method using a Markov random field model. His method detects less than 90% of the clusters to approximately 0.6 FP clusters per image.

In summary most of the techniques used in the computerised analysis of mammographic microcalcifications, first pre-process the digitised grey-level image to provide features representing microcalcifications. Feature analysis is then performed to detect microcalcifications.

Because mammographic images are normally very poor in contrast and severely lacking in the definition of microcalcification regions, the segmentation process for such regions is absolutely critical. If the segmentation process is not robust enough, the microcalcifications may be joined together or missed. This will result in producing noisy region descriptors which do not provide correct information about the grey-level distributions in the local structure (inside the region and its local background), since they play a significant role in actual mammographic analysis by radiologists [9]. Thus, classification of the segmented regions based on such noisy features will be a hard task.
We present here an algorithm very similar to the conventional methods by focusing on every part of the technique from segmentation through region growing to feature selection and classification. As mentioned in the introduction, our segmentation method includes a novel seed point detection algorithm followed by a new approach to region growing by pixel aggregation. Then a sequential floating search feature selection method is applied to select four different sets of features producing four combinations for different classifiers. Then four pattern recognition techniques have been applied to classify the suspected regions as calcifications or non-calcifications based on each set of the available features.

This thesis also presents a theoretical framework for weighted combinations of classifiers which use different representations. A very promising performance is achieved on the database when a weighted combination of four classifiers is used for decision making. Finally a clustering method is applied to locate clusters of detected microcalcifications in the spatial domain as the output of the method.

References


Chapter 3

Blob Detection Method

A microcalcification (MC) can be viewed as a small bright region on a textured background. Our approach to the detection of MCs in a mammogram is to identify the regions suspected to be MCs and then to classify them as normal or abnormal. First, the detection of seed blobs of pixels suspected to correspond to MCs is performed using a novel blob detection method. Subsequently, the extent of each blob is determined by using a novel region growing method. We consider detection of the seed blobs in this chapter and introduce the region growing method in the next chapter.

In the first section of this chapter we attempt to consider some general purpose image segmentation techniques and evaluate their capabilities from the MC detection point of view. Then a two step technique for the detection of seed blobs of possible calcifications is presented in Section 3.2 where the application of morphological filters to the problem is also considered. The last section evaluates the performance of the seed blob selection method adopted and offers concluding remarks.

3.1 Image Segmentation: A Brief Overview

The segmentation of an image into regions is an important first step for a variety of image analyses and visualisation tasks. A wide variety of image segmentation techniques is presented in the literature, some considered general purpose and some designed for a specific class of images [1, 5, 8, 15, 19, 24, 23]. The segmentation techniques for monochromatic images can be categorised into two different approaches [4, 12, 17]. One is region based,
which relies on the homogeneity, or similarity, of spatially localised features, whereas the other is based on abrupt changes in grey level, using discontinuity measures. If segmentation of a region is the aim, the two methods exploit two different definitions of the region which should ideally yield identical results. Homogeneity is the characteristic of a region and methods based on that are called region based methods while non-homogeneity or discontinuity is the characteristic of the boundary of a region and the methods which exploit this concept are called boundary finding methods.

If a region is homogeneous with relatively high contrast, the detection of the region boundary becomes a simple task using any of the two conventional methods. However, the problem arises when the high frequency information characteristic of the image is indistinguishable from that of a boundary. In such situations, the boundaries of image regions will not be well defined (i.e. uncertain boundaries exist) and boundary finding methods will fail, especially in the presence of noise. Although region based techniques are less affected by noise, they commonly suffer from the problem of over-growing into neighbouring regions or background especially when these are textured. Furthermore, since conventional boundary finding methods rely on changes in grey level, rather than on their actual values, they are less sensitive to changes in image contrast than the region based segmentation methods. Also boundary finding methods in general do a better job of boundary localisation [4, 9].

Many studies investigating the properties of the two approaches to image segmentation have been reported [1, 5, 7, 14, 17, 19, 24, 23]. As the two methods use complementary information, they involve conflicting objectives and therefore their direct comparison is not straightforward. Most of the reported techniques rely on a region growing method and use some discontinuity measures as a stopping criterion to avoid the problem of merging two neighbouring regions or over-growing into the background. The quality of these techniques is highly dependent on the edge operator used [14, 17] as a measure of discontinuity. Other approaches use the slope of a local planar approximation of the image surface. The idea is to test the hypothesis that the slope of the plane has changed as a characteristic of the boundary between two neighbouring regions. Fitting a plane to
image intensities over a set of pixels requires information about the region which is not always accessible in real situations. Consequently, these methods often exhibit a poor performance in defining the boundary.

This section aims to review the main representatives of the two classes of segmentation techniques and consider their capabilities from the point of view of microcalcification detection. Comprehensive surveys of different approaches to segmentation techniques, their merits and limitations can be found in Zucker [26], Fu and Mui [11], and Haralick and Shapiro [17].

3.1.1 Segmentation Techniques Based on Discontinuity

The approach is to partition an image based on discontinuities in a local neighbourhood. The most common way to look for discontinuities is to run a differentiating mask over the image and associate the grey level value of the output with the centre pixel. The process can be expressed as:

\[ g(x, y) = T[f(x, y)] \]  

(3.1)

where \( f(x, y) \) is the input image, \( g(x, y) \) is the processed image and \( T \) is an operator on \( f \), defined over some neighbourhood of \((x, y)\). Different functional forms of the operator \( T \) will produce different results. This approach is effective in finding many kinds of discontinuities, like points and lines, in the image, but the most popular area of interest within this category is the detection of edges in the image. An edge is the boundary between two regions with relatively distinct grey level transition properties.

Several local operators have been introduced to detect edges, eg. gradient and Laplace operators [12]. Since edge detection is a point based process, it is very sensitive to noise. So in real situations, discriminating the true edges at the object boundaries from discontinuities generated by other sources, eg. noise and unknown textures, is not an easy task. To minimise these problems, a higher level operation called edge linking is applied. Various strategies have been proposed to address this problem. Some are based on local measures
of continuity and smoothness, with no prior information about the object shape [18], others apply a rigid template at the object boundary [3], or one can use a mixture of the two methods known as deformable models [25]. However even with these post-processing methods the detection of edges delineating the boundaries of small regions like microcalcifications on a highly varying textured background is difficult. Edge detectors tend to produce huge amounts of noise within the textured area which cause any edge linking methods to fail. In Section 3.2 we address this particular problem and develop a new method based on intensity variations over a local neighbourhood to identify all the suspected blobs of microcalcification size range.

3.1.2 Segmentation Techniques Based on Similarity

The popular methods of finding similarity among pixels can be classified into the following categories: measurement space clustering, region growing, region split and merge, and spatial clustering schemes. Since they apply various strategies to group pixels into regions, based on similarity of the pixel properties, they can all be viewed as clustering schemes which incorporate spatial domain context to produce mutually exclusive groups of spatially coherent pixels.

The most commonly used similarity measurements for this category of image segmentation techniques are pixel grey-level value, and mean and variance of a local neighbourhood. However some discontinuity measures in conjunction with the similarity measures may also be used to improve the performance of the methods. These techniques are explained in the following sections.

3.1.2.1 Measurement Space Clustering

Measurement space clustering is applied to establish partitions in the measurement space. Each pixel in the spatial domain is then labelled based on the partition it belongs to. The connected components of the pixels in the spatial domain, having the same label, then form regions.

The basic idea behind these approaches is that homogeneous objects on the image manifest themselves as clusters in the measurement space. Typically the method works
well, if the measurements reflecting the properties of pixels within distinct objects in the image produce distinct clusters.

Most of the thresholding techniques, which attempt to separate the modes in the grey level histogram, belong to this category of image segmentation [17, 21]. These thresholding techniques use histogram statistics to find the best threshold to segment the image. Various approaches are suggested to optimise the thresholds [21].

![Figure 3.1: (a) Shows part of a mammogram containing one microcalcification and (b) shows part of a normal mammogram (window size = 25 \times 25\text{pixels}). (c) Illustrates the histogram of images (a) and (b).](image)

In order to show the capabilities of thresholding methods in the context of microcalcification detection, the histograms of two different parts of a mammogram are presented.
3.1. IMAGE SEGMENTATION: A BRIEF OVERVIEW

Figure 3.2: Co-occurrence matrix of normal and abnormal sections of a mammogram are shown in (a) and (b) respectively (pixel distance=1, directions= 0, 90° and window size= 25 × 25).

Figure 3.1 gives two examples illustrating a mammographic image patch of (a) a breast microcalcification and (b) a mammographic image patch of a normal area (non-calcification). A histogram of the image patch is shown in 3.1(c). As microcalcifications manifest themselves as blobs of bright pixels, one might have hoped for these properties to be reflected in a local image histogram. However, as we can see from Figure 3.1(c) no relevant difference between the histogram of the normal and abnormal section of the mammogram can be found. Therefore, the presence of any calcifications in the abnormal section cannot be distinguished by examining only the histogram of a mammogram. In our experience, any kind of one dimensional measurement space clustering scheme, like histogram examination, is not well suited for this purpose.

The same problem arises in the case of two dimensional grey level histograms, called co-occurrence matrices. A co-occurrence matrix represents the distribution of probability of occurrence of a pair of grey-level values separated by a given displacement vector. In other words, it indicates the frequency of occurrence of a particular grey level pair separated by the specified distance [13].
In order to establish whether there is any discriminatory information content in a two
dimensional grey level histogram, co-occurrence matrices obtained from different parts
of normal and abnormal mammograms are compared. The matrices computed for the ar­
eas show no meaningful difference between normal and MC regions especially when the
image is dense, see Figure 3.2. Some commonly used features of the co-occurrence ma­
trix, first and second moments, deviation, contrast, entropy and correlation [12], were also
considered in an attempt to develop a method for detecting MCs, but in our experience
correlation between the features and MCs was insufficient.

3.1.2.2 Region Growing Methods

Region growing is a procedure that groups pixels or sub-regions into larger regions. In
region growing, commonly, the value of each pixel is compared to a threshold specified as
a function of neighbouring pixels or region values. If its value satisfies the threshold, the
pixel and region are considered similar enough and the current region is grown to include
the pixel. If the converse is true a new region is initiated.

A wide range of region growing approaches have been proposed in the literature [6,
12, 16, 14, 17]. They can be categorised into single linkage region growing, hybrid linkage
region growing, and centroid linkage region growing methods which are described in the
following.

- Single Linkage Region Growing

In single linkage region growing, properties of neighbouring pixels are compared with
a similarity criterion. If the similarity test is satisfied, they will be linked together. Single
linkage region growing schemes are attractive for their simplicity. They have a problem
with chaining, because a single pixel satisfying the similarity criterion for two adjacent
regions is sufficient to cause the regions to merge [6].

- Hybrid Linkage Region Growing

The main difference between single linkage region growing and hybrid linkage region
growing methods lies in the use of a discontinuity measure, as a stopping criterion, to
avoid the problem of two neighbouring regions merging as a result of the similarity between a single pair of adjacent pixels. In this case, only similar adjacent pixels which fail to satisfy the discontinuity criterion, e.g. low gradient pixels, are joined together [17].

- Centroid Linkage Region Growing

In the centroid linkage region growing method, a current pixel is compared with the mean of the adjacent region. If the difference between the pixel grey level and the mean of the region is low enough, the pixel is linked to the region and the mean is updated. A similar process is used to merge two neighbouring regions. There are approaches based on this scheme which use other statistics, like the mean and variance of a local neighbourhood and their variations during the growing process, to test the similarity of each candidate pixel to the region [17]. The strength of centroid linkage is its ability to place boundaries in weak gradient areas.

3.1.2.3 Spatial Clustering

The combination of clustering in the measurement space with spatial region growing produces a new approach named spatial clustering. In essence, spatial clustering schemes combine the histogram mode seeking techniques with a region growing, or, spatial linkage techniques [16]. Typically one would start from a pixel at the peak of the histogram and grow a region to include neighbouring pixels in the spatial domain. This method can produce a better result than any of the former methods (region growing and clustering methods) for relatively large blobs.

3.1.2.4 Split and Merge

The procedures discussed as region growing methods are bottom-up processes which start from pixels as individual regions to segment the image. Other alternatives are top-down procedures which start from the whole image, as a single segment, to find segmented regions, by splitting the image into a set of distinct regions and merging adjacent partitions, provided they satisfy a similarity measure. The process of "split and merge", is repeated until every segment satisfies the similarity measure.
3.1. IMAGE SEGMENTATION: A BRIEF OVERVIEW

3.1.3 Conclusion

Before assessing the capability of the above methods to detect calcifications in mammographic images, it should be recalled that calcifications are of different variety in size, shape and coarseness, located in a fully textured background. In some cases they are difficult to detect without using a magnifier to focus on them. It is therefore important to consider the capability of the segmentation techniques in this context.

Segmentation based on discontinuity is not reliable for images with an unknown textured background, because the background can produce unpredictable noise which can mislead an edge linking step. However, it may be useful in combination with methods based on homogeneity.

Because of the small size of calcifications, the size of the associated regions is also small and consequently the measurement space statistics will not be affected by calcifications. Hence, algorithms based on measurement space clustering are not applicable to the problem at hand.

The main drawback of all the region growing techniques is over-growing into background [17] especially when it is textured. This problem may not be remedied by applying measurement space clustering, because of the small size of microcalcifications in mammograms.

Split and merge methods with fixed statistics as a similarity measure also produce an unreliable result for small regions in the large variety of textured background. Defining an adaptive statistic could be a way forward but at this point, no systematic way of adaptation has been suggested that would provide a reliable method of segmentation.

Since the existing techniques of image segmentation are not reliable enough for the purpose, we developed a two step segmentation technique which first detects the suspected blobs and then finds reliable boundaries for each blob. In the first step a Blob Detection Method (described in the next section) is applied followed by a novel region growing method, that is quite robust to textured backgrounds. As a representative of the state of the art region growing algorithms, the centroid linkage region growing technique described by Gonzalez & Woods [12] is implemented to compare its performance with our
3.2 A NEW APPROACH TO BLOB DETECTION

region growing method in Section 4.7.

3.2 A new approach to Blob Detection

The aim of blob detection is to find the location of any small region of locally high contrast in the image. The blob detector should be able to detect all the blobs in the MCs size range. We used a two-stage algorithm for this purpose. The first stage applies a morphological top-hat transform to flag all bright regions of small size. Then an adaptive thresholding technique based on the median of the local neighbourhood is employed to localise the blobs.

The set of pixels detected as suspected blobs are inserted in a list of starting points to be used for the next step of processing, namely region growing, in the original grey level image.

3.2.1 Morphological filters

Mathematical morphology has produced a class of nonlinear digital image processing operators which provides an approach to image processing based on shape and size, relevant to the problem of MCs detection [2, 10]. In mathematical morphology, information about the object size, shape, smoothness, connectivity and also orientation, can be built into an image analysis operator called a structuring element [22]. The structuring element is a tool for grey scale morphological operations. We use the top-hat transform to elaborate all the bright blobs of small size.

3.2.1.1 Top Hat Transform

Before defining the top hat transform let us introduce the two basic morphological operators: erosion and dilation. Erosion and dilation of a function, \( f(x, y) \), by a structuring element \( b(x, y) \) are defined as:

\[
(f \ominus b)(s,t) = \min \{ f(s + x, t + y) - b(x, y) | (s + x), (t + y) \in D_f; (x, y) \in D_b \} \tag{3.2}
\]

\[
(f \oplus b)(s,t) = \max \{ f(s - x, t - y) + b(x, y) | (s - x), (t - y) \in D_f; (x, y) \in D_b \} \tag{3.3}
\]

where \( D_f \) and \( D_b \) are the domains of \( f \) and \( b \), respectively.
The result of applying any of the two operators is directly related to the shape of the structuring element. Generally, erosion produces a darker image whereas dilation produces a brighter image than the original. Erosion, therefore, removes bright details based on the shape of the structuring element and dilation removes dark details.

The opening of function \( f \) by function \( b \) denoted by \( f \circ b \), is:

\[
f \circ b = (f \ominus b) \oplus b.
\] (3.4)

where \( f \oplus b \) and \( f \ominus b \) are dilation and erosion of function \( f \) by function \( b \), respectively.

The top-hat transform of a function, \( f(x, y) \), is defined as the difference between the function and its morphological opening [12].

\[
TH(f, b) = f - f \circ b
\] (3.5)

Opening involves two morphological operations, erosion followed by dilation. The initial erosion removes the small (with respect to the structuring element), light details and consequently darkens the image. The subsequent dilation increases the brightness of the image without recovering the removed details. Thus, opening removes bright details smaller than the size of the structuring element and, therefore, the top-hat transform highlights these details.

If the structuring function is flat, grey scale dilation and erosion are reduced to max and min filtering, respectively [20] and therefore the opening operation reduces to max-min operation in the local neighbourhood defined by the structuring element.

Here, we use morphological filters or more specifically the top-hat transform to enhance the details in a mammogram. The result of the top-hat transform of a mammogram in Figure 3.3(a) using a flat structuring element of size \( 9 \times 9 \) is shown in Figure 3.3(b). All the bright objects smaller than the structuring element with a variety of contrast appear as bright objects in the transformed image.

### 3.2.2 Thresholding

A local threshold is then applied to segment out the bright blobs from the transformed image. The blob pixels in the transformed image are detected as outliers with respect to
the distribution of the local background. Accordingly a median filter, which is not affected by outliers, is used to define an appropriate threshold \( T(x, y) \). Given a morphologically transformed image \( g(x, y) \), the thresholded output \( o(x, y) \), which segments out the bright blobs of interest is given by:

\[
o(x, y) = \begin{cases} 
1 & : \text{if } g(x, y) > T(x, y) \\
0 & : \text{elsewhere} 
\end{cases}
\] (3.6)

where \( T(x, y) \) is defined by:

\[
T(x, y) = k_1 + k_2 \times \text{Med} \{g(x - i, y - j) \mid -v \leq i, j \leq v\}
\] (3.7)

and \( k_1 > 1, k_2 > 1, v = (N - 1)/2 \).

\( T(x, y) \) is defined as a linear transformation of the median of image \( g(x, y) \) over a neighbourhood of \( N \times N \) where \( k_1 \) and \( k_2 \) are the parameters of the transformation. \( k_1 \) and \( k_2 \) have been empirically established to be 3 and 2.5, respectively. The median of the morphologically transformed image and the final output of the blob detector for a part of the mammogram are shown in Figure 3.3(c) and (d), respectively.

3.3 Conclusions

The capabilities of the conventional segmentation techniques were considered for the problem of MC segmentation. These methods were found unreliable for the detection of small blobs in a textured background. A novel blob detection technique which applies nonlinear filters, i.e. a Top-Hat transform and a median filter, is suggested for the detection of blobs of specific sizes.

As the blob detector segments out all blobs smaller than the structuring element, various problems may still arise such as:

- There may be many false positive blobs which are usually caused by the granularity of the mammogram and the large variations observed in the normal texture.

- Some of the detected neighbouring MCs are merged together. Consequently they appear as a single calcification which will cause the cluster detection step to fail.
3.3. CONCLUSIONS

Figure 3.3: The results of the successive steps of the blob detection algorithm on a part of image "mdb218rl" containing a cluster of MCs in a dense area.

- The shape of detected blobs can be distorted due to the nonlinear transformations used. Hence, information derived from detected blobs cannot be used directly to prune wrongly detected blobs.

These problems will be addressed in the next section by means of a novel region growing method where all the pixels detected by the blob detection method are used as starting points.
References


REFERENCES


Chapter 4

A New Region Growing Method

Here we present a new region growing method with the capability of finding the boundary of a relatively bright/dark region in a textured background. The concept of our method, like other region growing methods by pixel aggregation, is to start with a point that meets a detection criterion and grow the point in various directions to extend the region. The direction of the growing process is uniquely defined by the grey level of the current region boundary pixels. In each step, only one candidate pixel exhibiting the required property joins the region. This induces a known behaviour to our method which offers the possibility of preventing the region growing method from over-growing into a textured background [1, 3, 4] and, therefore, guaranteeing a reliable performance.

The growing procedure offers an ideal framework in which any suitable measurement can be applied to define a required characteristic of the segmented region. We use two discontinuity measurements called average contrast and peripheral contrast to control the growing process. Local maxima of these two measurements identify two nested regions, called the average contrast and the peripheral contrast regions. The method first finds the average contrast boundary of a region then a reverse test is applied to produce the peripheral contrast boundary. Since the two measurements are functions of grey level differences, their behaviour is not sensitive to intensity changes. This contrasts with the existing region growing techniques [5, 7, 10]. The method is very effective in defining the boundary of a region with fuzzy edges located in a textured background.

A number of experiments have been performed both on synthetic and real images to evaluate the new approach. The proposed scheme can be categorised as a region based
4.1 GROWING PROCESS

This section describes how the region growing expands a region from its starting point, when it includes only one pixel. The choice of a criterion to stop the growing process and of a starting point is discussed in Sections 4.2 and 4.4, respectively.

Let us assume that the process starts from an arbitrary pixel. The pixel is labelled as a region which then grows based on a similarity measure. In our approach, a boundary pixel is joined to the current region provided it has the highest grey level among the neighbours of the region. This induces a directional growing such that the pixels of high grey
4.1. GROWING PROCESS

Figure 4.1: (a) Topographical surface of a microcalcification in a homogeneous background and (b) Mapping of grey levels of the region during the growing process.

level will be absorbed first. When all the high grey level pixels in the region are absorbed, the process continues by absorbing the boundary pixels with monotonically lower and lower grey levels. When several pixels with the same grey level jointly become the candidates to join the region, the first-come first-served strategy is used to select one of them. This makes the region more compact, particularly in situations where the grey levels of the background, or, the region pixels are very homogeneous.

In order to monitor the pixels joining the region, a grey level mapping is generated. The mapping is very similar to the mapping used in the mode separating (MODESP) procedure proposed by Kittler [8] for cluster analysis. The MODESP method is a clustering procedure based on the mapping of data points from a high dimensional feature space onto a sequence in which each cluster in the space appears as a mode in the mapping. Separating surfaces between the modes in the high dimensional space are derived from
4.1. GROWING PROCESS

the points associated with distinct modes in the one-dimensional mapping function. The MODESP has never been used for the segmentation of spatially indexed data and the only similarity of our method with MODESP is the mapping used to monitor the growing process.

Consider Figure 4.1(a) which shows a small subimage with a single bright blob. To present the concept of the growing process on this data, let us assume that its starting point \( y_1 \) is the pixel with the maximum grey level of the subimage. It defines the nucleus of a region. The sequence of pixels joining the region is \( y_2, y_3, y_4 \) and so on. The graph of grey levels associated with the sequence of candidate pixels, for the region generated by the growing process, is shown in Figure 4.1(b). The mapping shows that the grey levels decrease from the highest value in the region to the background.

A similar mapping can be obtained for any measurement defined on the growing region. The mapping function defined on the sequence of pixels joining the growing region, characterises the variation of each measurement in the spatial domain. Different criteria can be used to stop the growing process and to apply a reverse check on the relevant measurements to detect the region boundary. We use the maximum possible size \( N_r \) of a region to stop the process. However, other criteria, such as minimum size of the neighbouring region or the maximum difference between current candidate and the maximum grey level inside the region, can also be applied to stop the growing process. We used the latter criterion for the segmentation of calcifications in mammographic images [6]. The size of a region is simply measured by counting the number of pixels in the mapping. This can be formalised as shown below.

The current pixel generated by the similarity measure is considered as a candidate for inclusion in the region, provided its index number \( i \) satisfies:

\[
i < N_r
\]

where \( N_r \) is the maximum expected size (number of pixels) of the region of interest. The criterion is used to avoid unnecessary growing into the background.

In the next section, we consider the use of two measurements as characteristic features of a region, to find its best boundary among all the candidate boundaries considered
4.2. DISCONTINUITY MEASURES

Figure 4.2: Schematic graph shows the CB, including candidate pixels to be joined to the region, and IB, including the outer-most pixels of the region, during the growing process. The region contains 20 pixels.

during the growing process. The applied measurements are not sensitive to the selected threshold $N_r$. Hence, the check against the threshold is introduced only to avoid unnecessary growth into a neighbouring region or homogeneous background.

4.2 Discontinuity Measures

For segmentation purposes we define a region of interest as a grey level blob, exhibiting a high contrast relative to its local background. The best boundary for the region is a set of connected pixels exhibiting predefined contrast properties. We use two different properties of the evolving region and its boundary, called *average contrast* and *peripheral contrast*, to define its nested boundaries. In the description of the proposed algorithm, four different boundaries are mentioned which are defined as follows: “Current boundary” (CB) is the set of pixels adjacent to the current region during the growing process. “Internal boundary” (IB) is defined as the boundary produced by the set of connected outer-most pixels of the current region. The current region and the two boundaries are dynamically changing during the growing process. The two mentioned boundaries, CB and IB, are shown in Figure 4.2 for a region during the growing process. “Peripheral contrast boundary” (PCB) is the boundary of the final output region produced by the region growing
4.2. DISCONTINUITY MEASURES

Method. "Average contrast boundary" (ACB) is the outermost boundary at which the growing process stops.

4.2.1 Average Contrast Measurement

The average contrast measure \( c(i) \) for a region containing \( i \) pixels is defined as the difference between the average grey level of the region and the average of its CB. This is expressed by:

\[
c(i) = \frac{1}{i} \sum_{t=1}^{i} y_t - \frac{1}{k - i} \sum_{t=i+1}^{k} y_t
\]

where \( y_1, y_2, \ldots, y_i \) is the sequence of pixels forming the current region and \( y_{i+1}, y_{i+2}, \ldots, y_k \) is the set of its CB pixels. This measurement captures the local contrast of the region based on all the grey levels inside the region and its boundary.

Let us recall that the algorithm always searches for the highest grey level in the boundary. The highest grey level pixel is then added to the growing region which systematically replaces the region boundary with pixels of lower intensity values. The region growing will produce increasing average contrast measure values, as long as the growing region continues subsuming high intensity pixels of the bright blob. Once it starts growing into the background, the rate of grey level decrease for the boundary will be less than that for its region, and consequently the average contrast will begin to decrease. Hence the maximum of this measurement during the growing process, corresponds to the point when the process starts to grow into the background. The result of the segmentation based on the maximum average contrast is the ACB of the region.

4.2.2 Peripheral Contrast Measurement

A commonly used discontinuity measure is the gradient [2]. It is well known that the gradient of a function points to the direction of the maximum rate of change of the function.

The gradient of a function \( f(x, y) \) at coordinates \((x, y)\) is defined as the vector

\[
\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}
\]
4.2. DISCONTINUITY MEASURES

The magnitude of this vector,

\[ |\nabla f| = \text{mag}(\nabla f) = \left[ \left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2 \right]^{1/2}, \tag{4.4} \]

is often approximated by a sum of absolute values of the partial derivatives as

\[ |\nabla f| \approx \left| \frac{\partial f}{\partial x} \right| + \left| \frac{\partial f}{\partial y} \right| \tag{4.5} \]

The image gradient magnitude can be estimated using various operators \( T \) in (3.1) with different functional forms. For instance, the derivatives of the image function in the \( x \) and \( y \) directions in equation (4.5) can be estimated using the Prewitt and Sobel gradient operators shown in figure 4.3. Applying the masks at any point give an estimate of the two derivatives and the summation of the absolute values of the outputs gives the magnitude of the gradient estimation at that point. These gradients are very sensitive to noise and consequently they cannot be used to stop the growing process as they may produce regions too small in size. As an alternative, we propose the use of an average gradient of the region’s boundary as a more reliable method to define the region. The average gradient, called \textit{peripheral contrast}, is estimated by computing the difference between the grey level average of the current IB and the average of the CB. The mapping of this measurement during the growing process, shows a pixel by pixel variation of an estimation of the boundary gradient of the evolving region. It is less sensitive to noise than the measurement of a pixel gradient magnitude, as it uses the difference between two neighbouring boundaries rather than that of two neighbouring pixels.

Note that for a relatively homogeneous region, the global maximum of the \textit{peripheral contrast} will be uniquely defined. However, for noisy or textured regions the \textit{peripheral contrast} will exhibit many local peaks. Each such peak can be used to segment out a distinct region which will meaningfully correspond to the information conveyed by the internal part of the region. In order to counteract the multiplicity of solutions caused by the effect of noise or texture on the \textit{peripheral contrast}, we use the last local maximum of the measurement occurring before the maximum of the \textit{average contrast} measure to determine the “peripheral contrast boundary” (PCB). We advocate the use of \textit{peripheral contrast} as the final result of segmentation.
4.3. TEST ON GAUSSIAN SHAPE IMAGE

Commonly the ACB and the actual boundary (PCB), as judged subjectively, are not very far away from each other. The difference is especially low when a bright region has a very sharp edge but it may be higher for fuzzy edge regions. We shall illustrate the difference between the two boundaries on an isotropic Gaussian blob image which has a very extensive ACB as compared to its PCB.

4.3 Test on Gaussian Shape Image

Theoretically, the highest gradient of a Gaussian shape is located one standard deviation from the mean. Equation (4.6) defines a two dimensional Gaussian shape:

\[ g(x, y) = M \exp\left\{ -\frac{1}{2} \left[ \frac{(x-u_x)^2}{\sigma^2} + \frac{(y-u_y)^2}{\sigma^2} \right] \right\} \]  \hspace{1cm} (4.6)

where \(u_x, u_y\) specify the \(x, y\) location of the centre of the Gaussian blob and \(\sigma\) specifies the spread of the grey level function. Constant \(M\) is used to normalise the output to the max-
It can be easily shown that the highest gradient magnitude of the Gaussian shape which is approximated by the peripheral contrast, is located one standard deviation from the mean. Thus the maximum peripheral contrast measure for the Gaussian shape specifies a circle with radius \( \sigma \), centred at coordinates \([ux, uy]\). 

A Gaussian shape image with a standard deviation of 25 pixels, \( \sigma = 25 \), shown in Figure 4.4(a), is used to demonstrate the relationship of the two boundaries. Let the growing process start at the highest grey level in the region, i.e. 255. The grey level mapping in Figure 4.4(e) shows that the grey levels of the sequence of pixels joining the region monotonically decrease to zero which corresponds to the background. As a result of the directional growing process, the shape of the region for the Gaussian shape is circular, even when the process continues to absorb the zero grey levels in the background. This is apparent by considering Figure 4.4(d) and noting that the grey level of all the candidate pixels beyond pixel number 21772 is zero. As one might expect, average contrast commences from a low value and smoothly increases to a maximum at point 6685 and then decreases. The maximum average contrast point in the Gaussian image defines a circular region with the radius of approximately 1.85\( \sigma \), as shown in Figure 4.4(c).

The mapping of peripheral contrast starts from a low value, increasing to a maximum at pixel number 2000 and then decreasing again to zero. The maximum peripheral contrast point corresponds to a circular region with the radius of approximately 1.01\( \sigma \) in the Gaussian image, shown in figure 4.4(b). This result agrees well with the maximum gradient region of a continuous Gaussian shape. The slight difference is caused by the effect of quantisation and the fact that our method uses the difference between the mean of two completely closed contour boundaries to calculate the peripheral contrast. Thus the effect of diagonal pixels, the distance of which is \( \sqrt{2} \), is the same as that of pixels located in the adjacent position with distance 1. The result of segmentation using this criterion is shown in figure 4.4(b).

We advocate the use of peripheral contrast as the final result of segmentation and apply the average contrast measure to find the ACB of the region of interest. Note that for sharp
4.3. TEST ON GAUSSIAN SHAPE IMAGE

Figure 4.4: The segmentation results of a Gaussian shape image with $\sigma = 25$. The region size criterion used is $N_r = 25000$. (a) Original image. (b) Segmentation result based on the peripheral contrast measure. (c) ACB segmented by the maximum contrast point. (d) The boundary produced by the region containing 25000 pixels. (e) Grey level, peripheral contrast and average contrast mappings obtained during the growing process.

edge regions, the results of segmentation using the average contrast and peripheral contrast measures are similar. In Section 4.5.1, we consider the performance of the method applying the two complementary measurements when the background is textured.
4.4 Starting Point

The starting point may be specified based on the properties of the region of interest. However, it is always possible to select randomly a set of starting points in different parts of the image and select as the final segmentation one of the results obtained. The selection process involves comparing particular features of the segmented regions with the one of interest. For microcalcification detection in mammograms, the Blob Detection Method is suggested, as described in Section 3.2. For medical image processing when the wish is to segment a known part of the image, the starting point may be specified manually by an expert.

4.5 Experimental Results

4.5.1 Textured Image

In this section, the aim is to consider the performance of the proposed procedure in situations where a noisy textured region is located in a textured background. A part of a road centre line shown in Figure 4.5(a) is used as the input. As can be seen, the dynamic grey level range of the region is quite high.

A point located at (60, 108) with the grey level of 175 is used as a starting point to segment the region. The grey level mapping is shown in Figure 4.5(e)-top. Note that, the mapping exhibits fluctuations during the growing process, two of which are particularly noticeable at pixel numbers 1100 – 1400 and 7100 – 7700. The artifact of each peak is a distinct valley in the average contrast and peripheral contrast mappings which occur in our example at pixel numbers 1300 and 7465, respectively, see Figure 4.5(e)-bottom. These two measures decrease inside the area of locally high grey level and increase after it has been covered (based on the grey level changes in the region and its boundary). The maximum average contrast measure defines the ACB of the region containing 18240 pixels, see Figure 4.5(b). The region’s texture produces more fluctuations in the peripheral contrast measure than in the average contrast measure. The peripheral contrast mapping shows that the difference between the global maximum which specifies the boundary of the region,
and local maxima which are created by texture and noise is not reliable enough to be used alone for segmentation purposes. In contrast, the global peak of the average contrast measure is not affected by the texture in the region. Hence, as mentioned before, the reliable gradient point to segment a region is the last peripheral contrast peak which is located before the global peak of the average contrast measure. This point is located at pixel number 12031, see Figure 4.5(e)-bottom. The segmented region based on this point is shown in Figure 4.5(c). This region is in full agreement with the result of human visual segmentation.

It is interesting to note that every local peripheral contrast peak will produce a segmentation which is likely to be perceptually acceptable. Figure 4.5(d) shows a region segmented out based on the first local maximum of the peripheral contrast of value 37.44 containing 511 pixels. This region can be characterised as a subregion of the region shown in Figure 4.5(c).

Grey level, contrast and peripheral contrast mappings obtained from any starting point inside the region, will eventually start appending the same pixels once all the brighter pixels are absorbed by the growing process. This happens as a result of the strategy of always appending the pixels of the largest grey level in the boundary to the current region. Thus, the segmentation results will be the same regardless of which starting point is used inside the region. Figure 4.6 shows the mapping produced during the growing process on the road centre line image, using two different starting points located at (362, 90) and (60, 108) with the grey levels of 171 and 175, respectively. The two starting points are marked on the image, Figure 4.5(a). The effect of non-homogeneity of the region is noticeable in the mappings. It can be noted that, the mappings for different starting points are different while the growing process absorbs brighter pixels inside the region. Nevertheless, after these pixels have been assigned, the mean of the grey levels inside the current region starts decreasing at a lower rate than that of the boundary grey levels. Hence, as expected, both the sequence and the corresponding measurements will converge to the same point and the mapping will exhibit the same behaviour thereafter. For the road centre line, the mapping functions become identical after pixel number 10000, Figure 4.6. The highest contrast
for the region is always located at pixel number 18240 with the grey level of 77. The highest peripheral contrast region containing 12031 pixels with the minimum grey level of 112, is specified by the peripheral contrast measure maximum of 41.72.

The use of the two contrast measures produces a unique region independent of the starting point. This will be further clarified in the next section when we consider the region growing procedure on another real image.

4.5.2 Experiments on Medical Images

This section shows the performance of our method on medical images where each region can be categorised as a bright blob separated from its neighbour by a low grey level boundary. First we show that our method is not sensitive to threshold \( N_r \) of rule (4.1). We then give an interpretation of the grey level mapping when \( N_r \) is too high in comparison to the size of the region of interest. The segmentation result obtained for Magnetic Resonance images (MR) of a head are presented and discussed. For each region we specify an arbitrary internal starting point.

Figure 4.7 shows an MR image of a head. The aim is to outline the corpus callosum, brain stem, pituitary gland and cerebellum. As can be seen, edges of the cerebellum and corpus callosum are very fuzzy at the boundaries of these regions. Consequently, neither boundary finding methods nor region based methods can reliably determine the boundaries.

We first start with the segmentation of the brain stem. A very high threshold is used \((N_r = 20000)\) in comparison to the size of the region of interest, to provide an opportunity to consider the behaviour of the discontinuity measurements in relation to the neighbouring regions. The grey level and contrast mappings obtained during the growing process are shown in Figure 4.8. The highest average contrast at pixel index 4604, determines the location of the ACB and the last peripheral contrast measure maximum before the maximum average contrast point specifies the final boundary for the region.

The grey level mapping shows local valleys which are induced by the grey levels at the boundary of two neighbouring regions. Each visible valley in the grey level mapping,
is the result of the switch between the absorption of decreasing grey levels of the pixels in the boundary of the region being grown and the absorption of pixels of increasing grey levels, leading to the nearest local peak of the neighbouring region. As mentioned before, the peripheral contrast of the grey level mapping on the left side of the valley is related to the rate of grey level decrease in the boundary of the region and the size of its CB. The size of the region affects the rate, because the number of pixels in the boundary is a function of region size. Hence, the bigger the size, the lower the peripheral contrast. This measurement on the right side of the valley is related to the rate of grey level increase along the pathway forged by the growing process towards a local hill of a neighbouring region. The latter is very sharp because the growing process takes a pixel width path to the top of the hill and then continues to cover its surface. Thus the difference between a valley minimum and the following peak in the grey level mapping, shows the difference between the maximum grey level of the hill and the maximum grey level at which the two neighbouring regions meet. If the difference is quite high and the number of pixels in the new region is high enough, it is a strong clue for the existence of a new significant region. Otherwise the new hill is a local peak or noise in the region being grown.

The effect of these variations is even more clear in the average contrast mapping. The growth into a neighbouring region, causes a more rapid increase in the mean of the grey levels of the pixels within the region's boundary, compared to that of the region itself. Consequently a local peak in the grey level mapping causes a local valley in the average contrast mapping. The local peak in the contrast measure corresponds to the ACB of its corresponding region. Its related peripheral contrast peak will then specify the best boundary to the region. The contrast mapping in Figure 4.8 shows the sequence of such peaks and valleys.

Figure 4.7(c) shows the boundary corresponding to the last peripheral contrast maximum, before the second distinct average contrast peak which segments a region containing 5677 pixels. The two peripheral contrast maxima located before the second and third average contrast peaks are located at point 11904 and 14448 (see Figure 4.8-bottom). The corresponding boundaries are shown in Figure 4.7(d) and (e), respectively. As mentioned
before, each boundary has a meaningful information regarding different possible regions 
produced by the process which can be of interest in target detection.

The five distinct parts of the MR image as segmented by using the average contrast and 
peripheral contrast peaks are shown in Figure 4.9. For each segmented region in the image 
a starting point is selected. We tested the algorithm using every starting point in the five 
regions but the segmentation results were the same (there was zero difference between 
the results produced by any starting point inside a region). The independence of the seg­ 
mmentation results from the choice of a starting point is an important characteristic of the 
approach. Figure 4.9(b) shows the result of segmentation based only on the average con­ 
trast mapping. As can be seen, the ACB of each region is well characterised. Figure 4.9(c) 
and (d) show the distinct regions produced by applying each of the two measurements 
separately.

In another experiment we examined the performance of the method on the MR image 
when Gaussian noise was added to the original image. The segmentation results obtained 
on the noisy image were compared with the results produced by applying the method to 
the original image. The segmentation error rate is defined by the percentage of pixels in­ 
correctly labelled by the region growing method and for a given noise level, it is averaged 
over different noise sequences. The error for various levels of standard deviation, σ, of the 
Gaussian noise for the five regions in the MR image is plotted in Figure 4.10. The error 
rate for the scalp is very low, less than 10%, even when σ = 20. This is because of the rela­ 
tively sharp edge between the scalp and other tissues. The error rate as a function of noise 
is much higher for the cerebellum, the corpus callosum and the pituitary gland. High sen­ 
sitivity of the method to the noise for those regions is caused by the relatively low contrast 
between the tissues and their background (fuzzy edges) and by being located in a close 
vicinity to neighbouring regions. The latter is particularly important as occurrence of high 
intensity grey level noise at the boundary between two neighbouring regions may cause 
the two regions to amalgamate. In such situations, application specific measurements, eg. 
shape measures, can be applied to prevent the growing process from absorbing the neigh­ 
bouring regions.
The above discussion is also applicable to segmenting out a dark region from a brighter background if the whole process is reversed. In such a case, the minimum peripheral contrast measure defines the final boundary for the dark region. This is demonstrated by applying the method to segment out the cavities in another MR image, shown in Figure 4.11(a). The segmentation results of our method using three arbitrary starting points, one in each cavity, are shown in Figure 4.11(b). The results are again in full agreement with the results of human visual segmentation.

4.6 Adaptation of the Method for Microcalcification Boundary Extraction

The method can be adopted for microcalcifications (MCs) segmentation purposes with minor modification. The mappings of the measurements during the growing process show the pixel by pixel variation of the peripheral contrast and average contrast in the spatial domain over the evolving region. Any peak of the mappings specifies the boundary for a region which could be of interest. Since we are more concerned with the regions of small size, a criterion based on size is used to specify the most relevant peak. We illustrate this procedure for two conditions: when a single MC exists and when two MCs are adjacent. The procedure for other situations when more than two adjacent MCs exist can be considered with the same logic as two MCs.

Consider Figure 4.12(a) which shows a small subimage with a single MC. To present the concept of the method on this data, let us assume that its starting point $y_1$ is the pixel with the maximum grey level of the subimage. The sequence of pixels joining the region is $y_2, y_3, y_4$ and so on, see Figure 4.12(e). The graph of the grey levels associated with the sequence of candidate pixels for the region generated by the growing process is shown in Figure 4.12(f)-top. The mapping shows that the grey levels decrease from the highest value in the region to the background. The peripheral contrast and average contrast mappings shown in Figure 4.12(f)-bottom start from low values, increasing to a maximum at pixel number 39 and 88, respectively. These pixels point to regions with the highest average contrast and peripheral contrast measures achieved during the growing process as shown in
Figure 4.12(c) and (d) respectively. The two boundaries are shown over the original image to demonstrate their differences for a blob with fuzzy edge boundaries.

In a special situation, when two MCs are projected very close to each other, the PCB may specify a single boundary to cover them. Consequently they appear as a single calcification which will cause the microcalcification detection step to fail. The potential problem can be prevented by applying a criterion based on the minimum acceptable size of a neighbouring region as an MC.

Let pixel $y_k$ be the current local min of the grown region. The second criterion is defined to check how many subsequently appended pixels with index $l > k$ satisfy:

$$l - k < N_n \quad \text{where} \quad y_l > y_k \quad \text{for} \quad l = k + 1, k + 2, \ldots, k + N_n, \ldots, N_r$$

(4.7)

Index number $l$ is used to count the number of pixels in the sequence of grey levels joining the region, with values greater than the minimum grey level $y_k = y_{\text{min}}$, and $N_n$ is the minimum acceptable size for a neighbouring region to be considered as a distinct region. If the size of a neighbouring region is greater than $N_n$, which means a neighbouring MC exists, the process for PCB finding terminates and a region corresponding to the highest peripheral contrast measure is segmented.

In summary, two coarse criteria are suggested for the MC detection purpose: the first criterion, (4.1), uses the maximum possible size $N_r$ of the region to stop the growing process and the second criterion, (4.7), uses the minimum expected size $N_n$ of the neighbouring regions aspiring to be MCs to stop the growing process based on the peripheral contrast measurement. This ensures that the PCB of the region being grown will not cover an adjacent region. Therefore, criterion (4.7) is a prerequisite for a local peak in the peripheral contrast mapping to define the PCB of the region.

The effect of the two criteria is illustrated by considering the growing process for two MCs projected next to each other. Figure 4.13(a) shows a subimage containing two neighbouring MCs. The criteria used are $N_r = 500$ and $N_n = 5$. We consider the different mappings, shown in Figure 4.13(e), produced by the growing process when the starting point is coincident with the pixel of maximum grey level inside the MC located on the right hand side of the subimage. The mapping shows that grey level first decreases from the maxi-
4.6. ADAPTATION OF THE METHOD FOR MC BOUNDARY EXTRACTION

Minimum to a minimum value inside the MCs region and then goes up through the highest gradient direction to the local maximum grey level in the neighbouring region. The peripheral contrast is approximated by the mean difference grey level of two neighbouring boundaries. Since the size of the neighbouring region exceeds \( N_n = 5 \) pixels, the growing process based on the peripheral contrast measure is terminated and the maximum peripheral contrast peak located at index number 11 is selected to define its corresponding boundary. This boundary is shown in black on Figures 4.13(b) and (c). Returning to the grey level mapping, the boundary is located in the intersection of two regions where some, but not all of the pixels with the same grey level are joined to the region. The contrast boundary for the region is shown in Figure 4.13(d). The region growing algorithm is then restarted from the left region and as a result the PCB, shown by dark in Figure 4.13(b) is detected. The boundary specified by the highest contrast measure value is exactly the same as the one obtained by commencing from the right region, shown in Figure 4.13(d).

In order to prevent segmenting more than one set of boundaries for each blob in the image, a simple step is included to shift the starting point to the nearest local peak before applying the region growing process. The process is as follows; if the first new pixel has a grey level higher than the starting point, the growing process continues until it reaches the first turning point in the grey level mapping (local peak in spatial domain). The starting point is then moved to the turning point. If the local peak is not covered by any of the PCBs found so far, the growing process is initiated, otherwise the starting point is discarded and the next blob pixel is considered. The effect of this process is that only the first selected pixel among a set of starting points belonging to a local peak will trigger the region growing process. Note that the starting points may not be joined to the grown region.

As explained in Section 4.2, boundaries specified by the two measurements are different when the region has fuzzy edges and, conversely, similar when it has a sharp edge. For a spike (noise) the two boundaries are similar. They cover only the spike, which is usually one pixel. This specification is used to remove the spike noises by removing all the regions in which their contrast boundaries cover only one pixel.
4.7 A Comparison of the Proposed Method with the Centroid Linkage Region Growing Method

In order to compare the proposed method with one of the commonly used region growing methods on the MCs detection problem, a centroid linkage region growing algorithm, based on similarity and discontinuity, is implemented. As a similarity criterion, the difference between the mean of the region and the grey tone of the candidate pixel is used. The discontinuity measure adapted is the Roberts' gradient at the candidate pixel.

In each step, all of the 8 neighbours of the new pixel are labelled as candidate pixels. To join a candidate pixel to a region it should satisfy two criteria. First, the absolute difference of the candidate pixel from the mean of the region should be less than a threshold, $T_m$. Second, the maximum edge magnitude obtained from the Roberts' operator involving the candidate pixel should be less than a threshold, $T_e$. We considered the result of this algorithm for various values of $T_m$ and $T_e$. The best result was obtained when $T_e = 5$ and $T_m = 10$.

We used image 4.14(a) which contains calcifications of various sizes and contrast for the comparison. The segmentation result of our algorithm with $N_r = 500$ and $N_n = 5$ is shown in Figure 4.16 and the segmentation result of the centroid linkage algorithm is shown in Figure 4.15(a). The maximum grey level in each region identified in the annotated image is used as a starting point for region growing. The annotation of the image 4.14(a) is shown in Figure 4.14(b).

The performance of the two methods can be visually evaluated by comparing the segmentation result with the annotation in Figure 4.14(b). The result of the centroid linkage method shows over-growing for blobs located in a textured background and under-growing for some well defined blobs. The problem increases if the starting point is located in the border of a region. If a starting point is near the border of a region the result of the region growing algorithm is unpredictable. Also, we considered the result of segmentation using only the similarity measure of the centroid linkage algorithm by selecting a very large value for the discontinuity measure criterion ($T_e = 100$). The result is shown in Figure 4.15(b). Over-growing is more extensive than for the method using a discontinuity
measure, which shows that the use of discontinuity results in a better performance.

Comparing the results shown in Figures 4.15(a) and 4.16 with the annotation, Figure 4.14(b), clearly demonstrates that our algorithm outperforms the centroid linkage region growing method. The agreement of the proposed method with the annotation is excellent and more importantly, the proposed method is not sensitive to a change in the location of the starting point.

4.8 Conclusions

A new method of region growing by pixel aggregation, using novel similarity and discontinuity measures has been presented. The unique feature of the proposed approach is that in each step at most one candidate pixel will exhibit the required properties to join the region. This makes the direction of the growing process more predictable. Two new discontinuity measures named average contrast and peripheral contrast which use grey level difference information to produce the final segmentation result are proposed and their properties analysed. The use of the two discontinuity measures guarantees the robustness of our region growing approach to intensity changes. This contrasts with the sensitivity to grey level shifts commonly exhibited by conventional region growing techniques [5, 7, 10].

Since the growing process is directional, i.e. pixels join the grown region according to a ranking list, the method does not necessarily include all the pixels with the same grey level in the region. This contrasts with thresholding methods where all the pixels exceeding a certain threshold are included in the segmented region [9]. Based on our experimental results, our method appears to be more reliable and consistent than other region growing and thresholding methods when the aim is the segmentation of bright objects from a dark background or vice versa [5, 7, 9].

The result of the method does not appear to be affected by the presence of a reasonable amount of noise. Hence, it can be used for segmenting raw images without any need to apply a smoothing filter or perform preprocessing procedures to improve the signal to noise ratio. This property of the proposed method is in sharp contrast to standard segmentation techniques which are commonly disturbed by noise.
The most significant features of the method are:

- Insensitivity to the starting point location.

- Insensitivity to a reasonable amount of noise, and to region and/or background being textured.

- No statistical information is needed concerning the region.

We should emphasise that the proposed method is applicable for the segmentation of bright/dark regions in a dark/bright background without using any a priori knowledge about the region.

References


Figure 4.5: The segmentation results of a part of road centre line. (a) Input image. (b), (c) Boundary produced by applying average contrast and peripheral contrast measures, respectively. (d) Segmentation results based on the first local peripheral contrast maximum at pixel number 511. (e) Grey level, average contrast and peripheral contrast mappings during the growing process.
Figure 4.6: Mapping of grey level, average contrast and peripheral contrast measures during the growing process for two starting points at (362, 90) and (60, 108).
Figure 4.7: (a) Original MRI image. (b) Segmentation result of brain stem. (c), (d) and (e) Segmentation results based on different locally highest *peripheral contrast* regions at pixel numbers 5677, 11904 and 14448, respectively.
Considered region
Peripheral contrast region

---

Figure 4.8: The mappings for brain stem during the growing process starting at pixel (304,165), $N_r = 20000$. 

Figure 4.9: (a) Segmentation results of an MR image of head. (b) The ACB of each region specified by the *average contrast* measurement. (c) And (d) shows the segmented regions based on the two measurements.
Figure 4.10: Segmentation error rate of the regions in the MRI image for different levels of Gaussian noise (image range is 0 – 255).

Figure 4.11: The segmentation results of holes in the brain.
Figure 4.12: (a) An MC in a dense mammogram. (b) Two boundaries detected by the region growing method. (c), (d) The boundaries outlined by the peripheral contrast and average contrast measures, respectively. (e) Topographical surface of MC in (a). (f) Grey level, average contrast and peripheral contrast mappings during the growing process.
Figure 4.13: Segmentation results for two neighbouring MCs using criteria $N_r = 500$ and $N_n = 5$. (a) Original image. (b) Boundaries detected by the region growing method. (c) Two PCBs outlined for the two regions. (d) ACB which is similar for the two regions. (e) Grey level, average contrast and peripheral contrast mappings during the growing process for the MC identified by black boundary in (b). In order to have a good resolution in the peripheral contrast mapping, only the mappings functions from point 0 to 212 are shown.
Figure 4.14: Part of a mammogram containing microcalcifications and its annotation.
(a) Discontinuity and similarity measures are used.

(b) Only similarity measure is used.

Figure 4.15: Segmentation result of image by the centroid linkage region growing method.
Figure 4.16: Segmentation result of image by the proposed region growing method.
Chapter 5

Pattern Recognition

The segmentation technique described in the two previous chapters finds all the bright blobs in the image and outlines two boundaries for each blob. The boundaries are extracted based on the two measures of contrast derived by the proposed region growing method. In the next processing step the segmented blobs are to be labelled as microcalcifications or normal background regions.

In this chapter the application of pattern recognition techniques for the task of classifying the segmented regions into one of the two categories is considered. The attention is given to all three main stages of a pattern recognition system design (measurement extraction, feature selection and classification). The first stage, measurement extraction, is concerned with computing a number of region descriptors to represent the extracted region properties. A set of 39 region descriptors extracted from the two boundaries and their associated regions are computed to constitute the measurement space. Most of the region descriptors have already been used by other researchers for different purposes including microcalcification detection (Chen [3], Woods [18]). However some of the features reflect the unique characteristics of our region growing method and are therefore new. A list of the region descriptors can be found in Appendix A. This chapter, therefore, focuses on the design of feature selection and classification steps of the decision making system.
5.1 Feature Selection

The aim of feature selection is to reduce the dimensionality of pattern representation. Lower-dimensional pattern descriptors are commonly referred to as features. In general the complexity of a classifier is directly dependent on the number of dimensions of the pattern space. Hence, a dimensionality reduction will result in the reduction of the complexity.

Another important justification for dimensionality reduction comes from the classification performance. Having a finite number of training patterns, which is the case in real situations, gives rise to the "peaking phenomenon" problem or the curse of dimensionality [5]. This problem concerns the relationship between the probability of correct recognition and the number of features used. Initially the probability improves as new features are added, but at some point the inclusion of further features may result in an actual degradation in performance. Dimensionality reduction may therefore provide a decrease in error rates [1].

5.1.1 Problem Formulation

Formally the goal of feature selection is to choose a small subset of \( d \) features \( x = \{x_j | j = 1, 2, ..., d\} \), out of the available \( D \) measurements \( Y = \{y_k | k = 1, 2, ..., D\} \), so as to achieve the highest possible performance of the recognition system. The problem involves designing a reliable search method to choose an optimal feature subset by maximising a suitable criterion function. Different figures like a discriminant function, eg. inter-class and intra-class distances, or the classification error rate have been suggested as criterion functions for feature selection [5].

It is well known that the only way to guarantee the selection of an optimal subset of features is to use a simple but computationally expensive search [4], which examines all the \( \binom{D}{d} \) subsets of size \( d \). This number is excessive even for moderate values of \( D \) and \( d \). For example for our problem, we may want to choose a subset of 13 features out of the 39 measurements which would require evaluation of more than \( 8 \times 10^9 \) different feature sets. Obviously exhaustive search would be prohibitive for the real world application. An
5.1. FEATURE SELECTION

alternative method called Branch and Bound has been proposed by Narendra [10] which guarantees the optimal solution provided the criterion satisfies the set inclusion monotonicity condition. The method avoids testing all the \( \binom{D}{d} \) possible subsets by rejecting suboptimal subsets without direct evaluation.

Some other computationally feasible procedures are proposed at the cost of achieving only a suboptimal solution to the problem [5, 11]. Those methods include the use of genetic algorithms [16], simulated annealing [15] and various other search techniques [8]. The sequential forward search (SFS) method was originally proposed by Marill [9]. Its counterpart called sequential backward method was presented by Whitney [17]. A mixture of the two methods, called plus l take away r [8], provides a much more effective solution to the problem. More recent contributions to feature selection methods include the sequential floating search methods of Pudil et. al. [11]. Pudil demonstrated that the floating search methods are more reliable than other techniques such as genetic algorithms and simulated annealing.

A complete review of feature selection algorithms is beyond the scope of this thesis. Here we only present a brief description of the sequential search methods used in our study. A review of the classical feature selection techniques can be found in [5] and more recent contributions in [11].

5.1.2 Sequential Forward/Backward Search Methods

These algorithms are "bottom-up" or "top-down" sequential procedures depending on whether the selection process starts from the best measurement and then includes sequentially the best unused measurement at each step or from the full set of features and removing sequentially the worst measurement in each step. The former method is called Sequential Forward Search (SFS) [17] whereas the latter is called Sequential Backward Search (SBS) [9].

To formalise the SFS algorithm, suppose \( k \) features have already been selected from the set of available measurements, \( Y \), to form feature set \( X_k \). The \( k + 1 \)st feature is then chosen from the set of available measurements, \( Y - X_k \), so that
5.1. FEATURE SELECTION

\[ J(X_{k+1}) = \max_j J(X_k \cup \{x_j\}), \quad x_j \in Y - X_k \]

*Initialization:* \( X_0 = \emptyset \)

where \( J \) is the criterion to be maximised by the feature selection algorithm.

The SBS algorithm starts from the complete set of available measurements \( Y \), and eliminates one measurement at a time. Suppose \( k \) features have been removed from the set of measurements \( X_0 = Y \) to form feature set \( X_k \). Then the next feature to be removed to form set \( X_{k+1} \) is defined by

\[ J(X_{k+1}) = \max_j J(X_k - \{x_j\}), \quad x_j \in X_k \]

*Initialization:* \( X_0 = Y \)

The two methods are very fast and easy to implement, but both suffer from the so-called “nesting effect”. It means that in the case of the top-down search, SBS, the discarded features cannot be reselected while in the case of the bottom up search, SFS, the features once selected cannot be later discarded. The result is that the methods may fail to produce optimal results.

5.1.3 Plus l Take Away r Algorithm

The “nesting effect” can be overcome by alternating the process of depletion and augmentation of the feature set. This is achieved by sequentially adding \( l \) measurements to the feature set and removing \( r \) features from the feature set. This algorithm for \( l > r \) can be described as follows; Let \( X_k \) be the current feature set.

- **Step 1:** Apply SFS \( l \) times to generate feature set \( X_{k+l} \)
- **Step 2:** Apply SBS \( r \) times to obtain feature set \( X_{k+l-r} \)
- **Step 3:** Stop if \( k + l - r = d \) otherwise set \( k := k + l - r \) and return to Step 1

The procedure for \( l < r \) is the same as above with **Step 1** and **Step 2** interchanged. The algorithm reduces to the SFS algorithm when \( r = 0 \) and SBS algorithm when \( l = 0 \).
5.1. FEATURE SELECTION

This procedure is more flexible than the SBS and SFS algorithms but there is no systematic way to choose \( l \) and \( r \) to achieve the optimal result. This problem is overcome by using the floating search feature selection methods described in the next section.

5.1.4 Sequential Forward Floating Selection and Sequential Backward Floating Selection Methods

Floating search methods have been proposed by Pudil et. al. [11] in 1994. In these methods, the criterion used for feature selection is applied to switch between the forward and backward selection in the Plus \( l \) Take Away \( r \) algorithm. The use of float values for \( l \) and \( r \) provides a near optimal solution. According to the dominant direction of search (forward or backward) the method is called sequential forward floating selection (SFFS) or sequential backward floating selection (SBFS).

The Sequential Forward Floating Selection Algorithm

Assume \( k \) features have already been selected from the complete set of measurements \( Y \) to form set \( X_k \) with the corresponding criterion function \( J(X_k) \). In addition, the values of \( J(X_i) \) for all preceding subsets of size \( i = 1, 2, ..., k - 1 \), are known and stored.

- **Step 1 (Inclusion).** Using the SFS method, select feature \( x_{k+1} \) from the set of available measurements, \( Y - X_k \), to form set \( X_{k+1} \). Therefore
  \[
  X_{k+1} = X_k \cup \{ x_{k+1} \}
  \]
  which means the new set contains \( k + 1 \) features.

- **Step 2 (Conditional Exclusion).** Find the least significant feature \( x_r \), \( 1 \leq r \leq k + 1 \), in the set \( X_{k+1} \). If
  \[
  J(X_{k+1} - \{ x_r \}) > J(X_k)
  \]  
  then exclude \( x_r \) from \( X_{k+1} \) to form a new feature set \( X'_k = X_{k+1} - \{ x_r \} \), else set \( k := k + 1 \) and return to Step 1.

If \( k = 2 \), then set \( X_k = X'_k \) and \( J(X_k) = J(X'_k) \) and return to Step 1 else go to Step 3.
5.1. FEATURE SELECTION

The algorithm to find the least significant feature, \( x_r \), is similar to the SBS method. \( x_r \) is the least significant feature in set \( X_{k+1} \) when

\[
J(X_{k+1} - \{x_r\}) \geq J(X_{k+1} - \{x_j\}), \quad 1 \leq j \leq k + 1.
\]  (5.2)

- **Step 3 (Continuation of conditional exclusion).** Find the least significant feature \( x_s \), \( 1 \leq s \leq k - 1 \), in the set \( X'_k \). If \( J(X'_k - \{x_s\}) > J(X_{k-1}) \) then exclude \( x_s \) from \( X'_k \) to form a newly reduced set \( X'_{k-1} \) else set \( X_k = X'_k \) and \( J(X_k) = J(X'_k) \) and go to **Step 1**. Therefore,

\[
X'_{k-1} = X'_k - \{x_s\}
\]

and \( k = k - 1 \). Now if \( k = 2 \), then set \( X_k = X'_k \) and \( J(X_k) = J(X'_k) \) and return to **Step 1** else repeat **Step 3**.

The algorithm is initialised by setting \( k = 0 \) and \( X_0 = \emptyset \), and the SFS is used until a feature set of cardinality 2 is obtained. Then the algorithm is continued with **Step 1**.

**The Sequential Backward Floating Selection Algorithm**

The SBFS algorithm is very similar to the SFFS algorithm with the difference that it starts from the full set of measurements and uses the SBS algorithm to exclude one feature in each step unless the inclusion of the most significant measurement in the set of excluded measurements produces a higher performance in comparison to the performance which is previously achieved using the same number of features. Otherwise a new measurement is included and the procedure repeated.

To formalise the algorithm, assume \( k \) features have already been removed from the complete set of measurements \( \mathbf{X}_0 = \mathbf{Y} \) to form feature set \( \mathbf{X}_k \) with the corresponding criterion function \( J(\mathbf{X}_k) \). Furthermore the values of all supersets \( \mathbf{X}_i, i = 1, 2, ..., k - 1 \), are known and stored.

- **Step 1 (Exclusion).** Use the SBS method to remove the least significant feature \( x_{k+1} \) from the current set \( \mathbf{X}_k \) to form a reduced feature set \( \mathbf{X}_{k+1} \). 
5.1. FEATURE SELECTION

- **Step 2** (*Conditional Inclusion*). Find among the excluded features the most significant feature \( x_r, 1 \leq r \leq k + 1 \), with respect to the set \( \bar{X}_{k+1} \). If

\[
J(\bar{X}_{k+1} \cup \{x_r\}) > J(\bar{X}_k)
\]

then include \( x_r \) to the set \( \bar{X}_{k+1} \) to form a new feature set \( \bar{X}'_{k+1} = \bar{X}_{k+1} \cup \{x_r\} \), else set \( k := k + 1 \) and return to Step 1.

If \( k = 2 \), then set \( \bar{X}_k = \bar{X}'_k \) and \( J(\bar{X}_k) = J(\bar{X}'_k) \) and return to Step 1 else go to Step 3.

The algorithm to find the most significant feature, \( x_r \), with respect to the set \( \bar{X}_{k+1} \) is similar to the SFS method. \( x_r \) is the most significant feature with respect to the set \( \bar{X}_{k+1} \) when

\[
J(\bar{X}_{k+1} \cup \{x_r\}) \geq J(\bar{X}_{k+1} \cup \{x_j\}), \quad 1 \leq j \leq k + 1.
\]

- **Step 3** (*Continuation of Conditional Exclusion*). Find among the excluded features the most significant feature \( x_s \) with respect to the set \( \bar{X}'_k \). If \( J(\bar{X}'_k \cup \{x_s\}) > J(\bar{X}'_{k-1}) \) then include \( x_s \) in the set \( \bar{X}'_k \) to form a new enlarged set \( \bar{X}'_{k-1} = \bar{X}'_k \cup \{x_s\} \) and set \( k := k + 1 \) else set \( \bar{X}'_k = \bar{X}_k \) and \( J(\bar{X}'_k) = J(\bar{X}_k) \) and return to Step 1.

Now if \( k = 2 \), then set \( k := 2 \), then set \( \bar{X}_k = \bar{X}'_k \) and \( J(\bar{X}_k) = J(\bar{X}'_k) \) and return to Step 1 else repeat Step 3.

The algorithm is initialised by setting \( k = 0 \) and \( \bar{X}_0 = \bar{Y} \), and the SBS method is used until a feature set of cardinality \( D - 2 \) is obtained (it means until the two least significant features are excluded to form \( \bar{X}_2 \)). Then the algorithm continues with Step 1.

The floating search methods correct any wrong decisions made in the previous steps and in this way they can yield a near optimal solution. Pudil et al. [11] have shown that the SFFS and SBFS yield comparable results to the Branch and Bound search but computationally the former methods are much faster.
5.2 Classification

Classification techniques can be categorised into two well-known groups, namely parametric and non-parametric, according to the assumption that can be made about the probability distributions of the patterns in each of the classes. Parametric classifiers exploit the assumption that the class conditional probability densities have a known parametric form. The knowledge of the form can be used to derive a parametric decision rule. Once the parameters of the decision rule are estimated from the training data the classifier can be applied to classify unknown patterns. A well known method within this category is the Gaussian classifier. This is widely used in statistical pattern recognition because in many applications a feature vector can be well-modelled by a Gaussian distribution.

Non-parametric methods estimate the unknown class densities or a posteriori probabilities at a point using the training set available. They do not require any a priori mathematical model for the underlying patterns [5]. A group of non-parametric methods derive the local statistics based on distances between the point and the patterns in the training set. The most commonly used classifier within this group is the K-nearest neighbour classifier (K-NN). Another group of methods that can compute the relevant functions are Neural Networks. This group can be considered as adaptive classifiers that learn through examples in the training set. These include feed-forward neural networks such as Multi-Layer Perceptron (MLP) and kernel-based classifiers such as Radial Basis Functions (RBFs).

A brief description of the four different types of classifiers is presented in the next section applied to the problem of microcalcification detection.

5.2.1 K-Nearest Neighbour Classifier

The K-NN decision rule is a very simple and powerful method of pattern classification. It computes the distance from an unknown test pattern to every training pattern and identifies the K nearest training samples. The test sample is then classified to the class represented by the majority of the K-NNs. The assumption is that the a posteriori class conditional probabilities of the test sample and its nearest neighbours are equal.

Statistical decision theory dictates that a large training set is needed in order to achieve
5.2. CLASSIFICATION

a good estimation accuracy. In a large sample case, independent of a particular metric, \( K \)-NN can be expected to approximate the Bayes error performance. But, in practice only a finite number of samples is available and, therefore, the classification performance is dependent on the choice of metric. This has been demonstrated by Short and Fukunaga [14], Fukunaga and Hostetler [6] and Brown and Koplowitz [2].

Short and Fukunaga [14] showed that under mild assumptions, the locally optimum metric for \( K \)-NN voting in the two class case is the Euclidean distance measure projected onto the local gradient of the a posteriori class probability function. Devijver [5] presented a modified version of the Short and Fukunaga’s algorithm [14] in which the gradient direction is defined as the difference between the means of two classes in the local neighbourhood. The procedure involves the following:

i ) Find the set \( \chi \) of \( m \) neighbours to a test pattern \( x \) with Euclidean metric.

ii ) Calculate the gradient direction \( V = M_1 - M_2 \) where \( M_i \) is the mean value of class \( \omega_i \) based on the samples in \( \chi \).

iii ) Project the training samples in \( \chi \) onto \( V \) and use the Euclidean metric in the one dimensional subspace to find the \( K \)-nearest neighbours to \( x \).

Thus a test sample, \( x \), is assigned to the class, \( \omega_i \), which has the most samples, \( k_i \), among the \( K \) nearest samples. The selection of suitable values for the two parameters \( K \) and \( m \) is discussed in Section 5.2.1.1.

The mild assumptions imposed are that the conditional densities and a posteriori probabilities are sufficiently smooth so that they may be approximated locally as linear functions.

5.2.1.1 Parameter Estimation for the \( K \)-NN

Values of \( K \) and \( m \) are important parameters of the \( K \)-NN classifier. As \( K \) and \( m \) increase from 1, the decision rule relies on a larger and larger neighbourhood of the test feature vector which causes the decision making to be influenced by an increasing number of training patterns. Hence, the performance of the classifier based on different values of \( K \) and \( m \)
Figure 5.1: K-NN classifier performance versus $K$ and $m$ for three different conditions, is estimated using LOO error estimation method. The performances are shown when (a) non-normalised patterns, (b) normalised on mixture patterns and (c) normalised on abnormal patterns are used.
5.2. CLASSIFICATION

will be related to the distribution of the training patterns in the two classes in the corresponding neighbourhoods. A reliable way to identify the best values of the two parameters is to estimate the classification performance for a variety of $K$ and $m$ using the leave-one-out (LOO) method and select the combination for which the classifier delivers the best performance.

In the LOO method, one sample is excluded from the training set and the classifier is designed using the remaining $N - 1$ samples. The excluded sample is then used for testing the classifier. This operation is repeated $N$ times to test all the training samples. The number of misclassified samples is counted to obtain an estimate of the error. Since each test image measurement is excluded from the design sample set, the independence between the design and test sets is maintained. The number of correctly classified samples is then counted to obtain an estimate of the classification performance.

In addition to the above two parameters which must be specified for the $K$-NN algorithm to obtain the best results it is also necessary to define the acceptance threshold. Changing the threshold biases the decision towards one of the two classes. A biased decision is desirable for applications where the cost of misclassification for one of the classes is more than that associated with another class. The aim in MC detection is to detect a high true positive rate while the false positive rate is kept as low as possible. We use a biased threshold when $k$ is even during the feature selection to provide a higher sensitivity in favour of the abnormal class.

To specify $K$ and $m$, we run the program for different $K$ from 1 to 30 (about 10% of the smallest class in our training set described in Section 5.3.1) and $m$ from $K$ to 90. When $m = K$ the results are equal to the $K$-NN with Euclidean distance. A wide range is used for $m$ to ensure that various numbers of samples are used to estimate the gradient for the locally optimum metric.

Figure 5.1 shows the performance of the $K$-NN classifier for different $m$ and $K$. The results are presented for three conditions: two differently normalised and a non-normalised patterns. The two normalisations are performed, first by a linear transformation of features using the difference between the maximum and minimum values of each feature in
the mixture pattern and the other using the difference between the maximum and minimum values of each feature in the class of abnormal patterns in the training set. As can be seen, the global behaviour of the classifier performance is more smooth when \( m \) is higher than \( K \) for all the three conditions. The performance for normalised features using patterns of abnormal class is higher than that of the two other cases. The main reason for the higher performance is that the background can vary considerably and therefore the variation of features corresponding to the normal background is unpredictable. Consequently, the distribution of normalised features is affected by its outliers produced by the extreme values of the normal background, which adversely affects the performance of the classifier.

These considerations identified the best combination of \( K \) and \( m \) (\( K = 5 \) and \( m = 17 \)) which are used in the following experiments whenever the \( K \)-NN classifier is applied.

### 5.2.2 Gaussian classifier

The Gaussian classifier exploits the assumption that the probability density of each class has a Gaussian shape. For a two class problem the Gaussian classifier computes the Mahalanobis distances \( \Delta_i(x) \) between pattern \( x \) and the mean of each class and compares their difference to a threshold. The Mahalanobis distance is defined by:

\[
\Delta_i(x) = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)
\]  

where \( \mu_i \) and \( \Sigma_i \) are the mean and covariance matrix of the design set for class \( \omega_i \), respectively.

### 5.2.3 Multi-layer Perceptron

The multi-layer perceptron is a biologically motivated classifier. It consists of a large number of simple processing units in which information is processed to produce a single output. Every unit performs a weighted sum of its inputs and produces a single output value using an activation function which involves amplifying/thresholding the sum. The topology of the network refers to the way the processing units are interconnected. The knowledge of the network is coded in the values of the interconnection weights which are up-
5.2. CLASSIFICATION

dated during the learning procedure. Various learning procedures, activation functions and interconnections of the units produce different network designs [7].

The generic structure of a fully connected MLP network with a single hidden layer is shown in Figure 5.2. The response of $k$th hidden unit, $h_k$, and the $j$th output are given by

$$h_k = g\left(\sum_i v_{ki}x_i\right);$$

$$o_j = f\left(\sum_k w_{jk}h_k\right),$$

where indices $i$ and $k$ sum over the input and hidden units respectively. The structure of a single unit (or neuron) in the MLP network is shown in figure 5.3. If the activation function $g$ is sigmoidal and $f$ is linear, then the network can uniformly approximate any continuous function provided a sufficiently large number of hidden units and enough training patterns are available. This property including the simplicity of the MLP neural network explains why it has been applied to a wide range of classification and function approximation problems.

Learning in the network of Figure 5.2 entails adapting the weights, including $v_{ki}$ and
5.2. CLASSIFICATION

Figure 5.3: Structure of a single neuron in the MLP neural network.

\[ w_{ki}, \text{ in order to minimise a cost function such as the Mean Square Error (MSE) function:} \]

\[ E = \frac{1}{2} \sum_j \sum_p (t_j^p - o_j^p)^2 \]

where \( t_j^p \) is the target value for the \( p \)th pattern, and \( o_j^p \) is the actual output for the \( p \)th pattern.

The well known back-propagation algorithm for training MLPs performs the steepest descent on the gradient of \( E \) in the weight space [1]. Using this algorithm, each weight of links can be updated using

\[ \Delta w_{jk} = \eta \delta_j o_k \]

where \( \eta \) is the learning rate, and \( \delta_j = (t_j - o_j)f'(h_j) \) for the output layer, and \( \delta_k = \sum_j w_{jk} \delta_j \) for all other layers with the \( k \)th layer preceding the \( j \)th layer. Details of the "back-propagation of error" learning rule, and of other more complicated learning methods can be found in [7].

A popular fully connected feed-forward neural network with three layers is used in this experiment. A sigmoidal function is used for hidden and output units which produce a real value output between 0 and 1 based on the weighted sum of inputs. The initial weights are randomly chosen between 0 and 1 for every input. There is no precise rule available to specify the number of hidden units to achieve a good performance. Thus, a suitable topology of the network is normally determined by trial and error.
5.2. CLASSIFICATION

5.2.4 Radial basis functions network

The RBF classifier is a kind of neural network, with a structure similar to that shown in Figure 5.2, built to estimate the conditional probability density functions, \( P(x|\omega_i) \), for each class, \( \omega_i \). Each density is estimated as a finite mixture of uncorrelated Gaussians \( G(x; \mu_j, \sigma_j) \), centred at \( \mu_j \) and with a diagonal covariance matrix of components \( \sigma_j \)

\[
P(x|\omega) = \sum_{j=1}^{M} W_j G(x; \mu_j, \sigma_j)
\]

where \( W_j \) is the weight associated with the \( j \)-th component of the mixture. The network is then trained using the log-likelihood as an optimality criterion. Such a network is shown in figure 5.4. The number \( M \) of hidden neurons of the pdf approximating network is selected using a method referred to as MPL (Maximum Penalised Likelihood)

\[
MPL = \text{log-likelihood} + \text{Penalty}
\]

\[
\text{Penalty} = -\frac{dp \log N}{2N}
\]

where \( p \) is the total number of parameters to be estimated, \( d \) is the data dimensionality and \( N \) is the number of training samples. For details see [12, 13].

Once the densities, \( P(x|\omega_i) \), have been estimated, the classifier is built through the usual Bayes rule. For the two class problem, \( \omega = 1, 2 \), the classifier is built as shown by the schematic diagram in Figure 5.5.
5.3 Experimental Method

The feature selection process is applied to select sets of relevant features from the set of available measurements, see Table A.1. These features are used as a basis for decision making in the classification stage.

5.3.1 Feature selection

The performance of two simple classifiers, K-nearest neighbour (K-NN) and Gaussian classifiers, is used as the criterion function to find a suitable combination of features. This

\[ \begin{align*}
\text{Class } \omega_1 & \rightarrow \text{ RBF network with } m_1 \text{ neurons} \rightarrow P(X|\omega_1) \\
\text{Train set } X_1 & \\
\text{Class } \omega_2 & \rightarrow \text{ RBF network with } m_2 \text{ neurons} \rightarrow P(X|\omega_2) \\
\text{Train set } X_2
\end{align*} \]

\[ P(\omega_1) P(X|\omega_1) \leq P(\omega_2) P(X|\omega_2) \]

Figure 5.5: The use of Bayes decision rule for decision making using the RBF classifier.

Having an estimate of the pdfs not only defines a decision boundary between the two classes, but it also gives an insight about the data structure. It allows us to identify the presence of outliers: points that should neither be classified as normal, nor as abnormal, because there is little evidence that they come from either class.

In order to detect possible outliers, two thresholds have been defined:

\[ T_{\omega_1} = \min_{x \in X_{\omega_1}} P(x|\omega_1) \]

\[ T_{\omega_2} = \min_{x \in X_{\omega_2}} P(x|\omega_2) \]

where \( X_{\omega_1} \) and \( X_{\omega_2} \) are the training sets for the two classes. Points, such as

\[ P(x|\omega_1) < T_{\omega_1} \]

\[ P(x|\omega_2) < T_{\omega_2} \]

are considered as outliers.
The results of feature selection

Performance

Number of Features

Figure 5.6: The results of SFFS and SBFS feature selection methods using the LOO error estimation method on the training set for the Gaussian classifier and the $K$-NN classifier using the optimal distance.

criterion (performance measure) is computed using the LOO error estimation method on the training set. The training set was formed by comprising 960 regions selected from 5 normal images and 320 single microcalcifications from the three abnormal images containing microcalcifications spreaded over the images. The 39 independent measurements extracted from each region are used to provide the initial measurements.

The two floating search methods, SFFS and SBFS [?], are applied to maximise the performance of each classifier by selecting relevant sets of features from the original 39 region descriptors. Figure ?? shows the performance of the two classifiers for various numbers of features selected by the floating search methods. Both classifiers exhibit inferior performance when the number of features is low. Because of the curse of dimensionality, the performance declines for both classifiers when too many measurements are used. The best performance for both classifiers is achieved for a moderate number of features.

Since the performance of the $K$-NN classifier does not significantly rise when the num-
ber of features is greater than 11, this set with the performance of 97.3% is used as the best set for the K-NN classifier. The performance reduces to 96.6% for the feature set corresponding to the first local peak in the error criterion. The set contains 7 features. The result of feature selection is different for the Gaussian classifier. The maximum performance of 96.3% is achieved with a set containing 17 features while the first significant local peak produces a set of 13 features yielding a 96% correct classification rate, see Figure 5.6.

The four sets of features containing the 7, 13, 11 and 17 features are called thereafter as FS1, FS2, FS3 and FS4 respectively. List of the features are as following:

- **SF1**: 3, 5, 6, 10, 12, 14, 39
- **SF2**: 1, 3, 5, 8, 12, 14, 16, 19, 20, 32, 39
- **SF3**: 3, 4, 5, 9, 14, 17, 18, 20, 22, 24, 28, 29, 39
- **SF4**: 1, 2, 3, 5, 7, 11, 13, 14, 15, 17, 18, 19, 21, 24, 28, 33, 39

5.3.2 Classification

The sets of features obtained as described in Section 5.3.1 were used to build two different MLP and RBF classifiers. The classifiers yielding the best performance on the training set were then tested on an independent database. The best MLP architecture comprised an input layer with 13 units, a hidden layer with 9 units and an output layer with 2 units. The best RBF network had 13 units for the normal class (no microcalcifications) and 7 units for the abnormal class. The classifier uses 7 features. The K-NN and the Gaussian classifiers use 11 and 17 features respectively.

5.4 Performance Measures

As discussed in the Introduction, two figures of merit, image identification and cluster of microcalcification, are used in this experiments. However the two figures are not independent, they represent two different aspects of the system. The first figure, image identification, is very important when the application of the detection system for a screening
5.5. EXPERIMENTAL RESULTS

program is considered. The second figure, cluster of microcalcification detection, is im-
portant for prompting all the possible abnormalities in an image with the aim of aiding
the interpretation performed by radiologists.

Our aim is to identify a quenching point which will guarantee a 100% true positive rate. To facilitate presentation of the results on such a quenching point, two false posi-
tives called Error-1 and Error-2 representing the two figures, cluster of microcalcification
detection and image identification, are defined. Error-1 shows the number of falsely de-
tected clusters of microcalcifications per image when all the clusters are labelled correctly
and Error-2 reflects the percentage of normal images misclassified as abnormal when all
the abnormal images are labelled correctly.

5.4.1 Clustering

A hierarchical nearest mean clustering routine is applied to check the classifier results for
the existence of a cluster of MCs. $x, y$ coordinates of the highest grey level in a classified
region are used to constitute the input pattern for the clustering method. A threshold of 1
cm$^2$ is used as a discontinuity measure to distinguish a new cluster. If a cluster has more
than two objects all the objects are labelled as a cluster of MCs. The centroid of each clus-
ter is computed and compared with known clusters. If the computed cluster centroid is
within the area of a known cluster, which is labelled in the database, then a true positive
(TP) detection is identified, otherwise the cluster is a false positive (FP).

5.5 Experimental Results

The performance of the four classifiers for the two different figures of merit were sepa-
rately examined on an independent test set. The eight images used for training set were
excluded from the dataset and the remaining images 19 microcalcifications and 200 nor-
mal images are used to provide an independent test set. The ROC curves for each figure
of merit is plotted in Figure 5.7.

A comparison based on "cluster of microcalcifications detection", shown in fig 5.7-(a),
indicates that the MLP classifier performs the best. It can detect all the clusters of micro-
5.5. EXPERIMENTAL RESULTS

Figure 5.7: ROC curve comparison of performance in cluster of MCs detection, (a), and in image identification, (b), by the 4 classifiers. The False Positives (FPs) also include the false detections on the set of normal images.

calcifications correctly while the FP cluster per image (on the normal and abnormal images) is less than 0.65. This figure is less than 0.75 FP clusters per image achieved with the Gaussian classifier and about 1.2 FP clusters per image for the RBF and K-NN classifiers.

Another ROC curve showing the percentage of correctly classified images versus falsely labelled normal images is shown in fig 5.7-(b). The curve demonstrates the comparative performance of the classifiers for various levels of confidence in abnormal image identification. The best performance is achieved by the RBF classifier, with a less than 15% false positive image rate when all the abnormal images are labelled correctly. This figure for the MLP, K-NN and Gaussian classifiers is 19%, 24% and 32% respectively.

Since the performance of the RBF classifier for the other figure of merit, “cluster of microcalcifications detection”, is worse than that of the MLP, it appears that the RBF detects lots of clusters in a smaller number of images while the MLP detects fewer clusters but spread over a higher number of normal images. It is not surprising that as far as the
5.5. EXPERIMENTAL RESULTS

Figure 5.8: (a) shows a bright spot in the background which has $P(x|\omega_1) << T_{\omega_1}$ and $P(x|\omega_2) << T_{\omega_2}$, while (b) shows an anomaly (perhaps a cyst) that is not a microcalcification.

discrimination between the two classes is concerned, the MPL performs better, since the optimality criterion used for the training was the best separating surface, whereas for the RBF network, the criterion of optimality was the best representation for the data structure. The learning of one class is performed ignoring the other class.

In contrast the RBF has a great advantage in comparison to the MLP: It can flag unexpected cases, such as outliers. Interestingly, instances of outliers have been confirmed in the two examples shown in Figure 5.8. The two examples reported are a bright pixel in the background (certainly an outlier that could lead to a false alarm) and a bright spot that has been identified as a microcalcification, but that is really another type of anomaly, perhaps a cyst: an anomaly different from the type of abnormality for which the network was trained.

Figure 5.9(a) shows a part of a mammogram with two clusters of microcalcifications with the segmentation results shown in figure 5.9(b). The two black circles marked on image 5.9(a) show the location of two clusters of microcalcifications, based on the MIAS annotation. All the correctly detected blobs are outlined in black and those detected outside
5.6. CONCLUSIONS

A true cluster of microcalcifications are outlined in white. As can be seen, all of the microcalcifications in the clusters are detected by the blob detector and correctly outlined by the region growing method. The results of applying the MLP classifier with three different a priori probabilities are shown in Figure 5.9(c,d,e). The first and the third a priori probabilities are those that guarantee 100% image identification and 100% cluster of microcalcifications detection respectively and the second one is selected to be a half way between the two. Using any of the a priori probabilities, the algorithm detects the two clusters of microcalcifications when one falsely located cluster is detected inside a single cyst. Using the first a priori, see Figure 5.9(c), seven suspected regions classified as microcalcifications are detected outside the two correctly detected clusters. They include three single microcalcifications spread over the image, a single bright blob at the bottom of the image and a false positive cluster with three regions in a cyst. Using a higher a priori probability for microcalcifications, the method detects a higher number of correct microcalcifications in the image while the number of wrongly detected blobs is also increased. Figures 5.9(c, d and e) show that as the number of truly detected microcalcifications increases from 22 through 25 to 30, the number of FP microcalcifications also increases from 7 through 8 to 13. Other classifiers behave in a similar fashion to the MLP classifier.

5.6 Conclusions

A systematic method has been proposed for the detection of microcalcifications in mammograms. The method relies on the information conveyed by the two boundaries defined by the proposed region growing method. It involves a four step procedure. The first step is a blob detection method which uses two nonlinear (or morphological) filters, median filter and top-hat transform, to detect regions of small size. All pixels detected as blobs are considered as starting points (seed pixels) for the next step of processing — region growing. The second step involves a specially designed directional region growing method which applies two measures, average contrast and peripheral contrast to find the best fitting boundaries for each blob region. The two boundaries, PCB and ACB, and their associate regions represent the characteristics of the region and its local background. Since the local peak
of each blob is used as a starting point for region growing, the method always reaches the same termination point. Spike noise is removed by discarding any region which is only one pixel in size. The classification of each blob as a microcalcification or background is based on features computed over the two regions associated with each blob.

Four different sets of features were used to design four classifiers. A comparison of the four classifiers using an independent set of 209 mammograms for the two figures of merit was performed. The final processing step involved the detection of clusters of microcalcifications. The hierarchical nearest mean clustering method was used for this purpose. The method reported all the clusters of at least 3 microcalcifications in an area of less than 1cm².

An extensive test over abnormal and normal images showed promising results of 100% true positive versus less than 15% false positive image identification using the RBF classifier. The algorithm detects less than 0.65 FP clusters per image when all the clusters are detected correctly. This performance is achieved using the MLP classifier. This result confirms the reliability of the method in microcalcifications clusters detection. The performance of the method is very encouraging but it is anticipated that it could further be improved by combining classifiers. This idea will be presented in the next chapter.

References


REFERENCES


(a) Original image.
(b) Peripheral contrast boundary.

(c) Final result using the MLP classifier.
Figure 5.9: (a) Part of a mammogram containing two clusters of microcalcifications. (b) PCB of suspected regions superimposed on the original mammogram. (c, d and e) The final results produced by applying the MLP classifier with three different a priori probabilities for labelling the suspected regions, shown in figure (b), as microcalcifications.
Chapter 6

Classifier combination

In the previous chapter, four different individual classifiers were used for labelling suspected regions as microcalcifications or abnormal background. Many recent studies in the pattern recognition literature suggest that object labelling performance can be improved by means of combining the opinions of individual experts. This is analogous to using the opinion of several specialists, instead of only one, to finalise the decision in medicine.

This chapter presents a theoretical framework for the combination of soft decisions generated by experts employing mixed (some shared and some distinct) object representations. By taking the confidence of the individual experts into account, weighted benevolent fusion strategies are derived. This provides a basis for combining classifiers and illustrates that a substantial gain in performance can be achieved by fusing the opinions of multiple experts. These strategies are experimentally tested and their effectiveness is considered.

6.1 Introduction

Classifier combination as a method of improving classification error probability has been of great interest to the pattern recognition community in recent years. Ample experimental evidence gathered in several application domains demonstrates that classifier combination offers an effective way to improve the performance of the recognition system.

The idea is to design several decision rules and to combine their outputs in order to reach a consensus decision about the object identity. Although the majority of the fusion
strategies advocated in the literature are largely heuristic, recently a few attempts to underpin a subset of these strategies by a common theoretical framework have been reported [1, 3].

From the point of view of their analysis, the approaches to multiple expert fusion can be divided into four categories. The first two groups include strategies applicable for fusing expert opinions based on identical measurements [3, 2] and distinct measurements [1] respectively. The third category comprises multistage combination rules [5] whereby objects are classified by a simple classifier using a small set of cheap features in combination with a reject option. For the more difficult objects more complex procedures based on different features are used. Finally, the last family of approaches encompasses data dependent fusion schemes [4] where the decision about the class membership of each unknown pattern is made by the locally most reliable expert.

For the first category, it has been shown in [3] for discriminant function classifiers that classifier combination reduces the classification error rate by means of obtaining a better estimate of the class boundaries. For the second category Kittler et al. [1] showed that many existing combination schemes can be developed from a common Bayesian framework. This is extended to take into account the confidence of individual experts in the computed aposteriori probabilities.

In the first two approaches each expert can be deemed to have the same influence on the final decision. However, it can be argued that a nonuniform weighting of classifiers outputs should outperform a corresponding combination scheme which assigns equal weights to the experts’ opinions. The support for this argument derives from the fact that equal weights represent just a single point in the potential weight space. By exploring other weight combinations it is guaranteed to achieve at least as good performance as standard combination rules but hopefully much better.

This chapter considers a mixed situation where the individual experts deploy representations which are partly shared and partly distinct and return soft decisions. In the first section, a common theoretical framework for multiple expert fusion is presented and used as a basis for deriving two fusion strategies. An analysis of the sensitivity of the strategy
to estimation errors is then performed to enhance the understanding of its properties. This framework is then extended to take into account the confidence in the individual expert opinions. We show that this leads to combination strategies which incorporate weighting factors. The weighting factors can be supplied by the experts themselves or where this is impracticable, they can be determined by means of training.

The chapter is organised as follows. In the next section, the theoretical foundations are developed for multiple expert fusion when partly shared and partly distinct representations are used. An analysis of the sensitivity of the strategies to estimation errors is performed to enhance the understanding of their properties. In Section 6.2.1 the necessary formalism is extended to weighted fusion strategies. In Section 6.3 the developed fusion strategies for the two conditions (distinct and mixed representations) are applied to the problem of clusters of microcalcification detection in mammographic image analysis. The section presents experimental results achieved with the different combination strategies, and compares them with the results achieved by the individual classifiers and the conventional classifier combination techniques. Finally, the last section summarises the results and offers concluding remarks.

6.2 Theoretical Framework

Consider an image labelling problem where object Z is to be assigned to one of m possible semantic categories \( \{\omega_1, ..., \omega_m\} \). Let us assume that we have R experts each representing the given object by a measurement vector which has a number of components shared with all the offered classifiers and the rest are unique. Denote the measurement vector used by the i-th expert by \( x_* \). In the measurement space, each class \( \omega_k \) is modelled by the probability density function \( p(x_*|\omega_k) \) and its a priori probability of occurrence is denoted \( P(\omega_k) \). We shall consider the models to be mutually exclusive which means that only one class can be associated with each object.

According to the Bayesian theory, given measurements \( x_i, i = 1, ..., R \), the object, Z, should be assigned to class \( \omega_j \), i.e. its label \( \theta \) should assume value \( \theta = \omega_j \), provided the
aposteriori probability of that interpretation is maximum, i.e.

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } P(\theta = \omega_j | x_1, \ldots, x_R) = \max_k P(\theta = \omega_k | x_1, \ldots, x_R)
\]  

(6.1)

Let us rewrite the aposteriori probability \( P(\theta = \omega_k | x_1, \ldots, x_R) \) using the Bayes theorem. We have

\[
P(\theta = \omega_k | x_1, \ldots, x_R) = \frac{p(x_1, \ldots, x_R | \theta = \omega_k) P(\omega_k)}{p(x_1, \ldots, x_R)}
\]  

(6.2)

where \( p(x_1, \ldots, x_R | \theta = \omega_k) \) is the class conditional joint probability density and \( p(x_1, \ldots, x_R) \) is the unconditional measurement joint probability density. Since the latter is class independent, in the following we can concentrate only on the numerator terms of (6.2).

6.2.1 Combining Classifiers Employing Mixed Pattern Representations

When the representations used by the individual classifiers are partly shared by the components of each pattern, vector \( x_i \) can be divided into two groups, shared and distinct patterns. This forms vectors \( y \) and \( \xi_i \), i.e. \( x_i = [y^T, \xi_i^T]^T \), where the vector of measurements \( y \) is shared by all the \( R \) classifiers whereas \( \xi_i \) is specific to the \( i \)-th classifier. We shall assume that given a class identity, the classifier specific part of the pattern representation \( \xi_i \) is conditionally independent from \( \xi_j \) if \( j \neq i \).

Now, let us express \( p(x_1, \ldots, x_R | \theta = \omega_k) \) as

\[
p(x_1, \ldots, x_R | \theta = \omega_k) = p(\xi_1, \ldots, \xi_R | y, \theta = \omega_k) p(y | \theta = \omega_k)
\]  

(6.3)

Using the assumption that the classifier specific representations \( \xi_i \) \( i = 1, \ldots, R \) are conditionally statistically independent, we can write

\[
p(x_1, \ldots, x_R | \theta = \omega_k) = [\prod_{i=1}^R p(\xi_i | y, \theta = \omega_k)] p(y | \theta = \omega_k)
\]  

(6.4)

which can be expressed as

\[
p(x_1, \ldots, x_R | \theta = \omega_k) = [\prod_{i=1}^R \frac{P(\theta = \omega_k | y, \xi_i) p(y | \xi_i) P(\omega_k | y) p(y)}{P(\omega_k) p(y)}] p(y | \theta = \omega_k)
\]  

(6.5)
and finally
\[ p(x_1, \ldots, x_R|\theta = \omega_k) = \left[ \prod_{i=1}^{R} \frac{P(\theta = \omega_k|x_i)p(x_i)}{P(\omega_k|y)p(y)} \right] \frac{P(\omega_k|y)p(y)}{P(\omega_k)} \] (6.6)

Let us pause to look at the meaning of the terms defining \( p(x_1, \ldots, x_R|\theta = \omega_k) \). First of all \( P(\theta = \omega_k|x_i) \) is the \( k \)-th class aposteriori probability computed by each of the \( R \) classifiers whereas \( P(\omega_k|y) \) is the \( k \)-th class probability based on the shared features. \( p(x_i) \) and \( p(y) \) are the mixture and shared measurement densities of the representations used for decision making by each of the experts. Since the measurement densities are independent of the class labels they can be cancelled out by the normalising term in the expression for the aposteriori probability in (6.2) and we obtain the decision rule

\[ \text{assign } \theta \rightarrow \omega_j \text{ if } \left[ \prod_{i=1}^{R} \frac{P(\theta = \omega_j|x_i)}{P(\theta = \omega_j|y)} \right] P(\theta = \omega_j|y) = \max_{k=1}^{m} \left[ \prod_{i=1}^{R} \frac{P(\theta = \omega_k|x_i)}{P(\theta = \omega_k|y)} \right] P(\theta = \omega_k|y) \] (6.7)

for equal a-priori class probabilities. Decision rule 6.7 combines the individual classifiers outputs in terms of a product. Each factor in the product for class \( \omega_k \) is normalised by the aposteriori probability of the class given the shared representation.

Now let us consider the ratio \( \frac{P(\theta = \omega_k|x_i)}{P(\theta = \omega_k|y)} \) and suppose it is close to one. We can then write \( P(\theta = \omega_k|x_i) = P(\omega_k|y)(1 + \Delta_{\omega_k}) \). Substituting into (6.7) and linearising the product by expanding it and neglecting all terms of second order and higher, the decision rule becomes

\[ \text{assign } \theta \rightarrow \omega_j \text{ if } (1 - R)P(\theta = \omega_j|y) + \sum_{i=1}^{R} P(\theta = \omega_j|x_i) = \max_{k=1}^{m} [(1 - R)P(\theta = \omega_k|y) + \sum_{i=1}^{R} P(\theta = \omega_k|x_i)] \] (6.8)

Note that the classifier combination rules (6.7) and (6.8) are expressed in terms of the aposteriori class probabilities returned by the individual classifiers using mixed representations and the aposteriori class probability based on the shared representation. Each classifier provides an independent estimate of the latter. It is therefore sensible to average these values to obtain a more reliable estimate \( \hat{P}(\theta = \omega_k|y) \), i.e. \( \hat{P}(\theta = \omega_k|y) = \)
6.2. THEORETICAL FRAMEWORK

\[ \frac{1}{R} \sum_{i=1}^{R} P_i(\theta = \omega_k|y) \] where \( P_i(\theta = \omega_k|y) \) is the aposteriori probability computed by the \( i-th \) classifier.

Further, it is worth noting that when the shared features are non informative, the aposteriori probabilities \( \hat{P}(\theta = \omega_k|y), \forall k \) will be comparable and will be close to apriori probability. Therefore, for equal a-priori probabilities, the term \((1 - R)\hat{P}(\theta = \omega_k|y)\) can be omitted from both sides of the decision rule (6.8) giving a combination rule

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \\
\sum_{i=1}^{R} P(\theta = \omega_j|x_i) = \max_{k=1}^{R} \sum_{i=1}^{R} P(\theta = \omega_k|x_i)
\] (6.9)

Even if the shared features are informative, it may be beneficial to ignore this term if the estimation errors on \( \hat{P}(\theta = \omega_k|y), \forall k \) are non-negligible, as the effect of these errors on the decision rule will be amplified by the factor \((1 - R)\). Based on the experimental results reported in Section 6.3, the decision rule (6.9) is an important alternative for the combination of multiple experts employing mixed mode representations.

6.2.2 Error Sensitivity

Let us consider the case when the output of individual experts is an estimate of the aposteriori class probabilities which deviates from the true probability. Let us denote the errors of the estimates of the aposteriori probability for class \( \omega_k \) for the mixed and shared representations by \( e_k(x_i) \) and \( e_k(y) \), respectively. In a real situation, (6.8) becomes:

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \\
(1 - R)(P(\theta = \omega_j|y) + e_j(y)) + \sum_{i=1}^{R} (P(\theta = \omega_j|x_i) + e_j(x_i)) = \\
= \max_{k=1}^{R} [(1 - R)(P(\theta = \omega_k|y) + e_k(y)) + \sum_{i=1}^{R} (P(\theta = \omega_k|x_i) + e_k(x_i))] 
\] (6.10)

which can be rewritten as

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \\
(1 - R)P(\theta = \omega_j|y)[1 + \frac{e_j(y)}{P(\theta = \omega_j|y)}] + \sum_{i=1}^{R} (P(\theta = \omega_j|x_i)[1 + \frac{\sum_{i=1}^{R} e_j(x_i)}{\sum_{i=1}^{R} P(\theta = \omega_j|x_i)}]) = 
\]
6.2. THEORETICAL FRAMEWORK

\[ \max_{k=1}^{m} \{(1 - R)P(\theta = \omega_k | y)[1 + \frac{\varepsilon_k(y)}{P(\theta = \omega_k | y)}] + \right. \]
\[ \left. \left[ \sum_{i=1}^{R} P(\theta = \omega_k | x_i)[1 + \frac{\sum_{i=1}^{R} \varepsilon_k(x_i)}{\sum_{i=1}^{R} P(\theta = \omega_k | x_i)}] \right] \right\} \tag{6.11} \]

A comparison of (6.8) and (6.11) shows that two different error factors,

\[ [1 + \frac{\varepsilon_k(y)}{P(\theta = \omega_k | y)}] \tag{6.12} \]

and

\[ [1 + \frac{\sum_{i=1}^{R} \varepsilon_k(x_i)}{\sum_{i=1}^{R} P(\theta = \omega_k | x_i)}] \tag{6.13} \]

affect the outcome. Comparing the error factors (6.12) and (6.13), shows that the sensitivity to error of the former is much more dramatic than the sensitivity of the latter as it is amplified by the factor of \( (1 - R) \). Note that since the a posteriori class probability is less than unity, the estimation errors on \( \hat{P}(\theta = \omega_k | y) \), \( \forall k \) will be also amplified by \( \frac{1}{P(\theta = \omega_k | y)} \). In contrast, for the complex representation the errors are not amplified. On the contrary, their compounded effect, which is also computed as a sum, is scaled by the sum of the a posteriori probabilities. For the most probable class this sum is likely to be greater than one which will result in the dampening of the errors.

These considerations demonstrate that the decision rule (6.9) is likely to be more resilient to noise as compared with (6.8). This finding is also observed experimentally in Section 6.3.2.

6.2.3 Weighting Factors in Classifier Combination

Let us return to equation (6.6) and introduce the following notation:

\[ P(\theta = \omega_k | x_i)p(x_i) = P(\theta = \omega_k | y)p_i(1 + \delta_{ki}) \tag{6.14} \]

and

\[ P(\theta = \omega_k | y)p(y) = P(\theta = \omega_k)p_{y}(1 + \delta_{ky}) \tag{6.15} \]

where \( p_i \) is a nominal reference value of the mixture density \( p(x_i) \) and similarly \( p_y \) is a reference value for \( p(y) \). A suitable choice of e.g. \( p_i \) is for instance \( p_i = \max_{x_i} p(x_i) \). Sub-
6.2. THEORETICAL FRAMEWORK

Substituting (6.14) and (6.15) for the appropriate terms in (6.6) we find

\[
\prod_{i=1}^{R} \frac{P(\theta = \omega_k | x_i) p(x_i) P(\omega_k | y) p(y)}{P(\omega_k) P(y)} = \frac{1}{(p(y))^R} \prod_{i=1}^{R} p_i \prod_{j=1}^{R} (1 + \delta_{ki}) p_y (1 + \delta_{ky}) \tag{6.16}
\]

Ignoring \((p(y))^{-R}\) on the grounds that this term is class independent and the confidence in \(P(\omega_k | y)\) has already been taken into account in \(p(y)\), we can expand the product and neglect any terms of second and higher order to approximate the right hand side of (6.16) as

\[
\prod_{i=1}^{R} p_i \prod_{j=1}^{R} (1 + \delta_{ki}) p_y (1 + \delta_{ky}) = p_y \prod_{i=1}^{R} p_i + p_y \prod_{j=1}^{R} p_i [\sum_{i=1}^{R} \delta_{ki} + \delta_{ky}] \tag{6.17}
\]

Substituting (6.17) and (6.14) into (6.6) and eliminating \(p(x) \prod_{i=1}^{R} p_i\) we obtain a sum decision rule

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } -R + \sum_{i=1}^{R} \frac{P(\omega_j | x_i) p(x_i)}{P(\omega_j | y) p(y)} + \frac{P(\omega_j | y) p(y)}{P(\omega_j) p_y} = \max_{k=1}^{R} [-R + \sum_{i=1}^{R} \frac{P(\omega_k | x_i) p(x_i)}{P(\omega_k | y) p_i} + \frac{P(\omega_k | y) p(y)}{P(\omega_k) p_y}] \tag{6.18}
\]

This approximation will be valid provided that \(\delta_{ki}\) satisfies \(|\delta_{ki}| << 1\). It can be easily established that this condition will be satisfied if \(P(\omega_k | x_i) p(x_i)/p_i P(\omega_k | y) - 1\) is small in absolute value sense. Note that this condition will hold when the amount of information about class identity of the object gained by observing \(x_i\) in relation to that gained from observing \(y\) is small and the observation is representative for the distribution of \(x_i\) which means that \(p(x_i)\) will be close to the reference value \(p_i\). Similar assumptions apply to the term \(\delta_{ky}\). However, whatever approximation error is introduced when the conditions do not hold, it has been shown in [1] that the adoption of the approximation has the benefits of reduced sensitivity to estimation errors as compared to the product rule, which will justify even the introduction of relatively gross errors at this step.

Before proceeding any further, it may be pertinent to ask, why we did not cancel out the unconditional probability density functions \(p(x_i)\) and \(p(y)\) from the decision rule. The main reason is that this term conveys very useful information about the confidence of the expert in the observation made. It is clear that an object representation for which the value
of the probability density is very small for all the classes will be an outlier and should not be classified by the respective expert. By retaining this information, the sum information fusion rule will automatically control the influence of such outliers on the final decision. In other words, the expert fusion rule in (6.18) is a weighted average rule where the weights reflect the confidence in the soft decision values computed by the individual experts. Thus our decision rule (6.18) can be expressed as

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \\
\sum_{i=1}^{R} w_j(x_i)P(\omega_j|x_i) + w_j(y)P(\omega_j|y) = \max \left[ \sum_{i=1}^{R} w_k(x_i)P(\omega_k|x_i) + w_k(y)P(\omega_k|y) \right] 
\]  

(6.19)

The main practical difficulty with the weighted average expert opinion combiner as specified in (6.19) is that not all experts will have the inner capability to output such information. For instance, it would not be provided by a multilayer perceptron and many other classification methods. We shall therefore limit our objectives somewhat and identify the weights \( w_i \) which will reflect the relative confidence in the experts expectation. This can be done easily by selecting weight values by means of minimising the empirical classification error count produced by the decision rule

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \\
\sum_{i=1}^{R} w_i P(\omega_j|x_i) + w_y P(\omega_j|y) = \max \left[ \sum_{i=1}^{R} w_i P(\omega_k|x_i) + w_y P(\omega_k|y) \right] 
\]  

(6.20)

in which the data dependence of the weights has been suppressed. In other words we find \( w_y \) and \( w_i, \ i = 1,..,R, \ [w_y + \sum_{i=1}^{R} w_i] = 1 \) such that \( e = \frac{1}{N} \sum_{k=1}^{N} \eta(Z_k) \) where \( Z_k, \ k = 1, N \) is the k-th training sample and \( \eta(Z_k) \) takes values

\[
\eta(Z_k) = \left\{ \begin{array}{ll}
0 & \beta_k = \eta_k \\
1 & \text{otherwise}
\end{array} \right.
\]

(6.21)

is minimised. In (6.21), \( \beta_k \) is the true class label of object \( Z_k \) and \( \eta_k \) is the class label assigned to it by the decision rule (6.20). The optimisation can easily be achieved by an exhaustive search through the weight space.
6.3 Experimental Results

The aim of the experiments is to demonstrate the benefits of multiple expert fusion in mammographic image analysis. In particular the problem is to label, either as microcalcifications or non-microcalcifications, all suspected regions which are detected in mammographic images by the segmentation method described in Chapters 3 and 4.

As in our case experimental studies showed that the shared features on their own were not very informative and they were included in the mixed mode representations anyway, in the following we focus only on the first term in the decision rule (6.20) and simplify it to:

\[
\text{assign } \theta \rightarrow \omega_j \text{ if } \sum_{i=1}^{R} w_i P(\omega_j|x_i) = \max_{k=1}^{R} \sum_{i=1}^{R} w_i P(\omega_k|x_i) \quad (6.22)
\]

when a priori class probabilities are equal. The weighted averaging combiner is schematically represented in Figure 6.1.

6.3 Experimental Results

The aim of the experiments is to demonstrate the benefits of multiple expert fusion in mammographic image analysis. In particular the problem is to label, either as microcalcifications or non-microcalcifications, all suspected regions which are detected in mammographic images by the segmentation method described in Chapters 3 and 4.
The same database as described in Section 1.3.1 containing 227 digitised mammo­
grams was used in this experiment. The 8 images used for training were excluded from
the database and the remaining images were divided into two sets, Set A and Set B. Set A
contained 10 abnormal (5 malignant and 5 benign) and 100 normal images and Set B is
made up of 9 abnormal (5 malignant and 4 benign) and 100 normal images.

To consider the independence of representations used, the scaled correlation matrices
of the features are examined. These matrices for each class separately are shown in Fig­
ure 6.2. The grey-level in the figure is related to the absolute value of the matrix element
which is to have unit entries on the diagonal. So brightness of the off-diagonal element of
the matrix is related to the degree of correlation between the corresponding representa­
tions. As the off-diagonal elements of the matrices are small, the assumption of indepen­
dence could be deemed to hold.

![Figure 6.2: Scaled correlation matrix of features used by the four classifiers. The first four features are shared by all four classifiers.](image-url)
### 6.3. EXPERIMENTAL RESULTS

<table>
<thead>
<tr>
<th>Classifier</th>
<th>No. of Features</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS1/RBF</td>
<td>7</td>
<td>0.44 12 %</td>
<td>0.98 15 %</td>
</tr>
<tr>
<td>FS2/MLP</td>
<td>13</td>
<td>0.40 21 %</td>
<td>0.62 19 %</td>
</tr>
<tr>
<td>FS3/K-NN</td>
<td>11</td>
<td>0.62 21 %</td>
<td>1.15 22 %</td>
</tr>
<tr>
<td>FS4/Gaussian</td>
<td>17</td>
<td>0.49 24 %</td>
<td>0.09 32 %</td>
</tr>
</tbody>
</table>

Table 6.1: The number of features used by each classifier and the errors produced on the two independent test sets (Set A and Set B).

#### 6.3.1 Individual Experts

The four different classification experts were implemented: RBF, MLP, K-NN and Gaussian classifiers using the feature sets FS1 to FS4 respectively. The Receiver Operating Characteristic (ROC) curve was then used to identify the a priori probabilities which will guarantee a 100% true positive detection.

The two different false positive figures of merit, Error-1 and Error-2, were adopted as a basis for fusion strategy assessment. The performance of the individual classifiers using the two different figures of merit is presented in Table 6.1. The MLP classifier achieves a minimum error for both test sets while the K-NN classifier yields the worst performance in terms of Error-1. The RBF classifier gives the best performance on both test sets in terms of Error-2 while the Gaussian classifier produces the worst result on both sets.

#### 6.3.2 Combiners using Classifiers Employing Mixed Pattern Representations

In the first experiment, we consider the influence of the four features shared by all the four classifiers on the sum rule. This experiment involves comparing the results achieved by the individual classifiers using the two different sets of features (shared and the mixed representations) with that of the sum rules presented by (6.8) and (6.9).

The classification rates achieved by the four individual experts and the equally weighted mean combination of the classifiers are presented in Table 6.2. The results show that an improvement is gained by averaging the aposteriori class probabilities based on only the shared representation. Table 6.3 presents the performance rates achieved using
### 6.3. EXPERIMENTAL RESULTS

<table>
<thead>
<tr>
<th>Classifier</th>
<th>No. of Features</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Error-1</td>
<td>Error-2 (%)</td>
</tr>
<tr>
<td>RBF</td>
<td>4</td>
<td>1.0</td>
<td>40</td>
</tr>
<tr>
<td>MLP</td>
<td>4</td>
<td>.62</td>
<td>36</td>
</tr>
<tr>
<td>K-NN</td>
<td>4</td>
<td>.70</td>
<td>31</td>
</tr>
<tr>
<td>Gaussian</td>
<td>4</td>
<td>.90</td>
<td>32</td>
</tr>
<tr>
<td>Mean Comb.</td>
<td>4</td>
<td>.55</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 6.2: The performance of the classifiers and their equally weighted average combination when only the shared features are used for classification purpose.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error-1</td>
<td>Error-2 (%)</td>
</tr>
<tr>
<td>Mean Comb. (6.8)</td>
<td>0.37</td>
<td>20</td>
</tr>
<tr>
<td>Mean Comb. (6.9)</td>
<td>0.30</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 6.3: Errors produced using equations (6.8) and (6.9) on the two independent data sets.

Multiple expert fusion as presented by the combination rules (6.8) and (6.9). When comparing results shown in Table 6.3 with Table 6.1, it is apparent that both combiners improve the classification rates in comparison to the best classifier, but rule (6.9) outperforms rule (6.8) which suggests that the class aposteriori probabilities based on the shared features are subject to substantial estimation errors.

### 6.3.3 Equally Weighted Combiners

In this experiment we consider the conventional combination strategies when mixture representations are used. Five commonly used multiple expert fusion schemes (mean, max, median, min and vote combiners) described in [1] were investigated. In these strategies all experts are deemed to carry the same weight. The errors produced by the five different combination strategies on the two independent test sets A and B are shown in Table 6.4. Comparing the results for different combiners, we find that the mean combiner outperforms the other combiners in three out of the four cases (Set A Error-1 and -2 and Set B Error-2). The median outperforms the mean combiner only when Error-1 is considered
6.3. EXPERIMENTAL RESULTS

<table>
<thead>
<tr>
<th>Combiner</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error-1 (%)</td>
<td>Error-1 (%)</td>
</tr>
<tr>
<td>Mean</td>
<td>0.30</td>
<td>0.78</td>
</tr>
<tr>
<td>Max</td>
<td>0.54</td>
<td>0.82</td>
</tr>
<tr>
<td>Median</td>
<td>0.31</td>
<td>0.62</td>
</tr>
<tr>
<td>Min</td>
<td>0.64</td>
<td>1.50</td>
</tr>
<tr>
<td>Vote</td>
<td>0.50</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 6.4: Errors produced on the two independent test sets by applying conventional classifier combination techniques (equally weighted combinations of the four classifiers).

for test Set B while it is the second best classifier combination scheme in the other three cases. The worst performance for both figures of merit is achieved by the \( \min \) combiner. This is the effect of the \( \min \) combiner relying on the lowest confidence classifier [1].

Comparing the results of multiple expert fusion in Table 6.4 with the results produced by the individual experts shown in Table 6.1, we note that all the strategies excluding the \( \min \) combiner perform better than the worst individual expert. This statement is correct for both errors considered here. On Set A, the \( \text{mean} \) combiner performs better or as well as the best individual expert with the exception of Error-2 on Set B. Among the rest, the \( \text{median} \) outperforms the best individual expert only in terms of Error-1. These comparisons illustrate that, although on the whole fusion offers a higher performance in comparison with the worst expert, it may fail to outperform the best individual expert.

6.3.4 Weighted Average Combiner

This section examines the benefit of incorporating weighting factors in multiple expert fusion as described by decision rule (6.22). In each experiment, one independent data set is used to determine the best combination of weights for the weighted average combiner and the performance of the resulting combiner is tested on the other data set. The role of Set A and B is then interchanged. This experiment is performed for each figure of merit separately. The best set of weights is obtained using an exhaustive search method by changing the weights incrementally between zero and one, with a step size of 0.2.

The best combination of weights to minimise Error-1 on Set A is \((0.14, 0.0, 0.43, 0.43)\)
6.3. EXPERIMENTAL RESULTS

<table>
<thead>
<tr>
<th>Weights for the weighted average combiner</th>
<th>Error-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF</td>
<td>K-NN</td>
</tr>
<tr>
<td>0.14</td>
<td>0.0</td>
</tr>
<tr>
<td>0.00</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 6.5: The best results for Error-1 produced by applying an exhaustive search to find the best set of weights for the weighted average combiner. The first row shows the best combination of weights for Set A and the result of applying the weights to Set B. The second row shows the best weights for Set B and the results of those on Set A.

<table>
<thead>
<tr>
<th>Weights for the weighted average combiner</th>
<th>Error-2 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF</td>
<td>K-NN</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 6.6: The best results for Error-2 produced by applying an exhaustive search to find the best set of weights. The first row shows results on set B when Set A is used for training. In the second row, Set B is used for training and Set A for testing.

for the RBF, K-NN, MLP, Gaussian classifiers respectively. Thus the best performance for Set A is achieved when the aposteriori probabilities produced by RBF, K-NN, MLP and Gaussian classifiers are weighted by 0.14, 0.0, 0.43, and 0.43, respectively. The same combination of weights on Set B produces a lower Error-1, 0.54, than that obtained by the mean combiner, 0.78. The same experiment on Set B demonstrates that a combination of weights producing the highest performance on Set B (0, 0, .5, .5) gives a better performance on Set A. As shown in Table 6.5, the mean combination of only two classifiers yielding the best performance on Set B gives a better performance than the mean combination of the four classifiers on Set A. The results in both cases are not only better than the best classifier used but also better than the results produced by any of the equally weighted combiners. From the weights associated with individual classifiers in Table 6.5 and the performance of each single classifier for Error-1 in Table 6.1, we find that the classifier with the best performance, MLP, is always included in the set of input classifiers and the worst classifier, K-NN, is always excluded from the set of classifiers used for the weighted average combiner.
Similar experiments were performed to find the best combination of weights when Error-2 is used as a performance measure. The results of this experiment are shown in Table 6.6. The best performance for both sets, Set A and Set B, is achieved by combining the RBF, K-NN, MLP and Gaussian classifiers with the weights of 0.5, 0.3, 0.2 and 0.0, respectively. The results for both data sets are much better than those obtained by the best individual classifier. By considering the weights in combination with the performance of the individual classifiers presented in Table 6.1, we observe that the best individual expert (RBF), which produces minimum Error-2, has the highest weight among the set of available experts while the worst individual expert (Gaussian) takes the lowest weight, 0.

From these observations it is apparent that fusing the best set of designs will result in a better performance than a simple averaging of a larger number of classifiers regardless of their performance.

The same rule, (6.22), may be used to combine the beliefs of two or more radiologists based on their confidence. The relative weights could be established by considering the performance of the weighted average combination of their confidence on a set of biopsy proven mammograms, for various combinations of weights.

In order to compare the performance of the weighted average combiner with that of the best individual classifier, we consider the results produced by the weighted combiner using the weights 0.5, 0.3 and 0.2 for the RBF, K-NN, MLP classifiers, respectively. The results of our method on a representative from the abnormal mammograms is shown in Figure 6.3. The location of the clusters of microcalcifications specified by the radiologists shown in Figure 6.3-b is marked by a dark circle in respect to the boundary of the breast. Figure 6.3-c and -d show the regions detected as microcalcifications using the MLP classifier and the weighted average classifier combiner. These results are produced using the a priori probabilities which guarantee that all the abnormal images are labelled correctly. This means at least one correct cluster of microcalcifications is detected on every abnormal image. The MLP classifier, detects three microcalcifications in the area annotated by radiologists and five individual regions in different parts of the mammogram, see Figure 6.3-c. The weighted average combination of classifiers detects four single microcalcifications.
correctly in the annotated area and the same number of regions outside the cluster. In order to facilitate a visual comparison, the cluster of microcalcifications detected is shown in Figure 6.4 under a zoom factor of about 10. Having a higher number of microcalcifications detected in the cluster (the annotated part) demonstrates that a greater accuracy can be expected if the weighted average combiner method is used as compared to the best classifier chosen for the same purpose.

6.4 Conclusions

The problem of combining multiple classifiers which employ mixed mode representations consisting of some shared and some distinct features was addressed. Two combination strategies were first developed and experimentally compared to demonstrate their effectiveness for mixed mode representation. The experimental results demonstrate that the sum decision rule is a justifiable strategy even when some of the features are shared.

By taking the confidence of the individual experts into account, the theoretical framework is then expanded and a weighted benevolent fusion strategy for the combination of soft decisions is presented. This strategy was successfully applied to the problem of automatic detection of microcalcifications in digital mammograms. The experimental results demonstrate that substantial gains in performance in image interpretation can be achieved by fusing the opinions of multiple experts which leads to a superior performance as compared with an equal weighting of experts opinions. Although the performance of the equally weighted combiner is better than the worst expert, only weighted fusion strategy is always likely to outperform the best expert.

The experimental results also warn about a possible degradation in performance caused by the inclusion of a poorly performing expert. The use of weighting factors in multiple expert fusion provides a practical way of preventing such an event.

References

REFERENCES


(a) Abnormal mammogram, mdb218rl, with a cluster of microcalcifications.

(b) Annotation of the mammogram is marked by a black circle in respect to the breast boundary.
Figure 6.3: Results produced on a full size glandular abnormal mammogram. (The original mammogram is shown contrast-enhanced for display purposes.)

(c) Results produced using the MLP classifier.

(d) Results produced by the weighted average combiner.
(a) A part of image mdb218rl including the cluster of microcalcifications.
(b) Detected cluster of MCs when the MLP classifier is used.
Figure 6.4: Annotation is marked by a black circle and boundaries of the detected microcalcifications are shown in black or white when it is inside or outside of the cluster respectively. The weighted average combiner detects more true positive microcalcifications and it is therefore less likely to under-detect a cluster of microcalcifications.
Chapter 7

Conclusion

7.1 Summary of the Research

The schematic diagram in Figure 7.1 shows the different steps of our calcification detection technique. In step one, during the automatic blob detection, the digitised grey level image will be segmented into binary regions representing the location of the blobs suspected of being microcalcifications. The set of detected pixels, called seeds of suspected blobs, are inserted in a list of starting points to be used for the next step of processing, namely region growing in the original grey level image. The region growing method, which is applied to the original image, starts from the seed points and grows through the highest grey levels in the boundary of the region, to extract two different boundaries based on two criteria used (peripheral contrast and average contrast). The suspected regions and their associated boundaries (outlined by the region growing method) provide useful information for pattern recognition purposes.

A set of 39 measurements is computed from the suspected regions and their associate boundaries for each blob. Four different sets of features, called region features, of the candidate regions are selected to constitute the input patterns which are then classified either as microcalcifications or normal background. Four different classifiers are employed to classify each suspected blob, and their soft outcome are combined using a novel weighted combination rule, as shown in figure 6.1, to label each pattern. The last section of the algorithm uses clustering to detect clusters of microcalcifications.

Extensive tests on 219 abnormal and normal mammograms in the MIAS database were
performed using the single classifiers and a combination of the four classifiers. These experiments demonstrate that the RBF classifier gives the best performance for image identification, with a promising results of 100% TP versus less than 15% FP. Our experimental results here also demonstrate that the best classifier for cluster of microcalcifications de-
tection is the MLP classifier with an outstanding result of 0.65 FP clusters per image when all the clusters are detected correctly. These results confirm the reliability of the method in microcalcification detection.

Although the performance of the method is very encouraging, a better performance is achieved when the weighted combination of classifiers is used for the classification task. The result utilising the same database, where half of the images were used for training (to determine the relevant weighting factors) and the other half for testing and vice versa, is 7% – 12% FP images when all the abnormal images were labelled correctly. The same experiment also demonstrates that the weighted combination of classifiers produces a better performance of 0.46 – 0.54 FP clusters per image, when all the clusters are detected correctly. The experiments here demonstrate that the performance achieved using the combination of classifiers is superior not only to the best classifier used but also to the conventional classifier combination techniques.

7.2 An Overview of Contributions and Findings

The main contributions of this research to the field of image processing and pattern recognition are three novel techniques; i) a novel blob detection, ii) a unique region growing technique, iii) a weighted combination of classifiers.

- The novel blob detection technique build of nonlinear filters (Top Hat transform and median filter) is capable of detecting blobs of microcalcification size.

- The boundaries extracted by the region growing method do not change in the presence of moderate amounts of noise.

- Based on the experimental studies on real images, it can be concluded that the region growing method appears to be very useful for medical application purposes. This conjecture is supported by the reported results of our method on MRI images.

- Experimental tests demonstrate the independence of the segmentation results from the starting point location on a noise free image.
From the experimental testing, our region growing method appears to be more reliable and consistent than the thresholding techniques.

As the performance of our method is not affected by the presence of a reasonable amount of noise, it can be used for segmenting raw images without any need to perform preprocessing techniques to improve the signal to noise ratio. This property of the proposed method is in sharp contrast to standard segmentation techniques whose performance is adversely affected by noise.

The application of the floating feature selection [1] method appears to be very useful for selecting relevant sets of features.

Based on comparative studies, the RBF and the MLP neural networks perform better than the K-nearest neighbour and the Gaussian classifiers.

The experimental results using the weighted combination of classifiers demonstrate that substantial gains in performance can be achieved by fusing the opinions of multiple experts.

A theoretical framework for the combination of soft decisions generated by experts employing distinct object representations is developed. By taking the confidence of the individual experts into account, we derived the weighted benevolent fusion strategies which were applied successfully to the problem at hand.

The results demonstrate that using weighting factors in combining experts’ opinions leads to a superior performance.

Although the performance of the equally weighted combiner is better than the worst expert, only weighted fusion is most likely to outperform the best expert.

The experimental results here also warn about a possible degradation in performance caused by the inclusion of a poorly performing expert. The use of weighting factors in multiple expert fusion provides a practical way of preventing such an event.
7.3. Future Work

In the majority of approaches considered in Chapter 2, a set of candidates are generated by image segmentation methods and verified by considering the properties of the candidates. Our approach is consistent with this group of microcalcification detection schemes. The main problem with this kind of approaches is that a low sensitivity in the earlier stages of processing cannot be improved by further processing. Consequently, during the early stages of processing, the number of true detected targets should be 100% whilst the false positive rate should be kept low.

A major goal of the next phase of this research is to improve the sensitivity of the first stage of processing. Iterative segmentation methods will be considered in pursuit of this goal. Such techniques as contextual segmentation or relaxation methods, make fuzzy or probabilistic classification decisions at every step in each iteration. They update the decisions at successive iterations, based on the decisions made at the preceding iterations at neighbouring pixels. In other words, the initial non-contextual classification is successively refined using the local context at each iteration.

In this work only linear weighting factors were investigated. The idea of weighting may be extended to nonlinear weighting functions for each classifier combination. However, the weighting factors may also be used for other classifier combination techniques mentioned in Chapter 6.

In order to ensure that the excellent experimental results achieved here can be replicated in practice, it would be desirable to test the system in a clinical setting for a period of time. Such a test will determine the robustness of the system for national screening applications.
A complementary step to the research studies carried out here will be to explore further the use of pattern recognition approaches by taking into account cluster features and their variations to label the detected clusters into malignant or benign categories.

Lastly, this study could be extended to adapt the present system for the detection of masses, spiculated regions and other kinds of abnormalities. Such an adaptation would have to be made to the first two steps of our algorithm to achieve this goal.

References

Appendix A

Feature space

This appendix presents the variety of region descriptors applied to produce the 39 measurements. Most of the measurements which are region oriented have already been used in many image processing applications [1, 2] including microcalcification detection [3]. Some new difference measurements are also produced by applying conventional region descriptors on the two regions and their associated boundaries produced for each blob.

The full set of measurements is shown in Table A.1. In order to simplify the description of the measurements, the following notations described in Chapter 4 are used in the table;

- ACR: Average contrast region.
- ACB: Average contrast boundary.
- PCR: Peripheral contrast region.
- PCB: Peripheral contrast boundary.

Four novel measurements defined by the unique characteristics of our segmentation method are marked by a star *.

References


<table>
<thead>
<tr>
<th>Type of feature</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Size of PCR</td>
<td>Number of pixels in PCR</td>
</tr>
<tr>
<td>2- Local contrast</td>
<td>Mean difference between PCR and ACR excluding pixels in PCR</td>
</tr>
<tr>
<td>3- Intensity of PCR</td>
<td>Mean grey level of pixels in PCR</td>
</tr>
<tr>
<td>4- PCR contrast</td>
<td>Mean difference between PCR and ACB</td>
</tr>
<tr>
<td>5- PCB contrast 1</td>
<td>Mean difference between PCR and PCB</td>
</tr>
<tr>
<td>6- PCB contrast 2</td>
<td>Mean difference between PCR and pixels covering ACB</td>
</tr>
<tr>
<td>7- Maximum peripheral contrast *</td>
<td>Peripheral contrast value pointing to PCR</td>
</tr>
<tr>
<td>8- Non-homogeneity of region 1 *</td>
<td>Minimum peripheral contrast value before PCR is segmented</td>
</tr>
<tr>
<td>9- Maximum average contrast *</td>
<td>Maximum average contrast value during the growing process</td>
</tr>
<tr>
<td>10- Non-homogeneity of region 2 *</td>
<td>Minimum average contrast value before ACR is segmented</td>
</tr>
<tr>
<td>11- Local variations 1</td>
<td>Difference between maximum grey level in PCR and minimum grey level in ACB</td>
</tr>
<tr>
<td>12- Local variations 2</td>
<td>Difference between maximum grey level in PCR and minimum grey level of pixels covering ACB</td>
</tr>
<tr>
<td>13- Local variations 3</td>
<td>Difference between maximum grey level in PCR and minimum grey level of pixels covering PCB</td>
</tr>
<tr>
<td>14- Local variations 4</td>
<td>Difference between maximum grey level in ACR and minimum grey level in ACB</td>
</tr>
<tr>
<td>15- Boundary variations 1</td>
<td>Difference between maximum and minimum grey levels in ACB</td>
</tr>
<tr>
<td>16- Boundary variations 2</td>
<td>Difference between maximum and minimum grey levels in PCB</td>
</tr>
<tr>
<td>17- Edge strength in PCB</td>
<td>Average Prewitt edge gradient [1] of pixels in PCB</td>
</tr>
<tr>
<td>18- Maximum edge gradient in PCB</td>
<td>Maximum Prewitt edge gradient [1] in PCB</td>
</tr>
<tr>
<td>19- Gradient range in PCB</td>
<td>Difference between maximum and minimum Prewitt’s edge gradient [1] in PCB</td>
</tr>
<tr>
<td>20- Fluctuations in ACR</td>
<td>Grey level variance of pixels in ACR</td>
</tr>
<tr>
<td>21- Fluctuations in ACB</td>
<td>Grey level variance of pixels in ACB</td>
</tr>
<tr>
<td>22- Fluctuations in PCB</td>
<td>Grey level variance of pixels in PCB</td>
</tr>
<tr>
<td>23- Compactness of ACR</td>
<td>Complexity of ACB with respect to a circle</td>
</tr>
<tr>
<td>24- Compactness of PCR</td>
<td>Complexity of PCB with respect to a circle</td>
</tr>
<tr>
<td>Type of feature</td>
<td>Interpretation</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>25- Central moment $u_{11}$ of ACR</td>
<td>Second order central moment of ACR</td>
</tr>
<tr>
<td>26- Perimeter of ACR</td>
<td>Number of pixels, length, in ACB</td>
</tr>
<tr>
<td>27- Perimeter of PCR</td>
<td>Number of pixels, length, in PCB</td>
</tr>
<tr>
<td>28- Intensity of local background</td>
<td>Mean of one pixel wide covering of the external boundary (ACB)</td>
</tr>
<tr>
<td>29- Contrast of local background</td>
<td>Difference between mean of the two boundaries covering PCB and ACB</td>
</tr>
<tr>
<td>30- Eccentricity [2] of ACR</td>
<td>Ratio of the large to the small eigenvalues obtained using pixels in ACR</td>
</tr>
<tr>
<td>31- Eccentricity [2] of PCR</td>
<td>Ratio of the large to the small eigenvalues obtained using pixels in PCR</td>
</tr>
<tr>
<td>32- Fluctuations in PCR</td>
<td>Grey level variance of pixels in PCR</td>
</tr>
<tr>
<td>33- Area variations</td>
<td>Difference between the area of PCR and ACR</td>
</tr>
<tr>
<td>34- Holes in PCR</td>
<td>Number of pixels outlined by PCB but are not included in PCR</td>
</tr>
<tr>
<td>35- Holes in ACR</td>
<td>Number of pixels outlined by ACB but are not included in ACR</td>
</tr>
<tr>
<td>36- Central moment $u_{11}$ of PCR</td>
<td>Second order central moment [2] of PCR</td>
</tr>
<tr>
<td>37- Diameter of PCR</td>
<td>Distance of two furthest points in PCR</td>
</tr>
<tr>
<td>38- Diameter of ACR</td>
<td>Distance of two furthest points in ACR</td>
</tr>
<tr>
<td>39- Fluctuations in local background</td>
<td>Variance over one pixel wide covering of the external boundary (ACB)</td>
</tr>
</tbody>
</table>

Table A.1: List of features used for microcalcification detection.