Efficient Dendronic Creation, Visualisation and Analysis for the Detection of Stellates in Digitised Mammograms

Thesis submitted for the degree of

DOCTOR of PHILOSOPHY (Ph.D.)

by

ROBERT MITCHELL

2006

Department of Electrical Engineering

University of Technology

Sydney, Australia

CERTIFICATE OF ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Production Note:
Signature removed prior to publication.

(Signed)

This research is dedicated to the memory of Mrs. Margaret Macfarlane, a wise lady and wonderful grandmother.

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Hung Nguyen, Emeritus Prof. Barry Thornton and Dr. Arthur Hung for their invaluable advice and encouragement over the past years.

To Chris and the rest of my family, thank you for putting up with late nights studying, proof-reading and the roller-coaster that is life. To the Watkins family, my appreciation cannot be expressed for the kindness you showed by welcoming us into your house whilst I was completing this research.

ABSTRACT

Breast cancer can be controlled and treated successfully when detected at early stages. Since survival rates are highest and recurrence as well as treatment costs are lowest if the cancer is detected and treated at an early stage, it is critical to diagnose breast cancer in its earliest stage.

However, the mammographic appearance of normal breast tissue varies widely and the signs of breast cancer are subtle. The evaluation of screening mammograms is a very difficult task where the radiologist must balance the requirement of high sensitivity for abnormalities (leading to high cancer detection) and high specificity for normality (keeping unnecessary biopsies to a minimum).

Computer Aided Detection of abnormalities in breast screening can potentially improve individual radiologists' performance, however, despite a great deal of research literature on image processing in mammography, the detection of cancerous mass lesions is still very difficult for many reasons. Masses are often of varying size, shape and density, at the same time exhibiting poor image contrast. In addition, many mass lesions and normal parenchymal tissues surrounding them look similar on mammograms.

A novel method utilising dendronic analysis of mammograms has been carried out in this research. A dendrone is a hierarchical thresholding structure that can be automatically generated from a complex image. The dendrone structure captures the connectedness of objects and sub-objects during successive brightness thresholding. Based upon connectedness and changes in intensity contours, dendronic representations of objects in images capture the course to fine unfolding of finer and finer detail, which is invariant to lighting, scale and placement of the object within the image.

Complex images can be autonomously analysed using the output of the dendrone to determine if they contain the signatures of particular target objects of interest, in this case, the signatures of stealth-like stellate mass lesions within mammograms. Dendronic analysis of mammograms has been overlooked due to the computational expense in the creation of the image dendrogram. Thus, in this research, a modified creation algorithm has been developed to dramatically decrease creation times. Further, a three-dimensional data visualisation implementation has been developed to better visualise and understand the dendronic representation of the image.

In the analysis of 50 mammogram images containing 15 cases of cancerous stellate mass lesions, the method proposed in this research detected 86.7% of the stellates. However, it is also evident that the number of false positives produced required further classification of the output.

These results indicate that dendronic analysis of digitised mammogram images has the potential to provide a robust tool to aid in the screening of breast cancer. The analysis is completely automated with very little *a-priori* information required and shows that dendronic analysis is an excellent tool in the detection of stellates. Moreover, the technique can now be performed within a clinically acceptable timeframe.

TABLE OF CONTENTS

CERTIFICATE OF ORIGINALITY ---------- II

ACKNOWLEDGMENTS ---------- IV

ABSTRACT ---------- V

TABLE OF CONTENTS ---------- VII

LIST OF FIGURES ---------- X

LIST OF TABLES ---------- XV

CHAPTER 1 INTRODUCTION ---------- 1

1.1 Background ---------- 1
- 1.1.1 Breast Screening ---------- 4
- 1.1.2 Mammography ---------- 7
- 1.1.3 Digitised Mammograms ---------- 11

1.2 Other Research in Breast Cancer Detection ---------- 13
- 1.2.1 Pre-processing ---------- 14
- 1.2.2 Pixel Level Features ---------- 16
- 1.2.3 Features for Mass Detection ---------- 16
- 1.2.4 Segmentation of Regions ---------- 18
- 1.2.5 Classification of Regions ---------- 18

1.3 Research Intent ---------- 20

CHAPTER 2 DENDROGRAMS ---------- 23

2.1 What is a Dendrogram? ---------- 24

2.2 Spatial Image Dendrograms ---------- 26
- 2.2.1 Repetitive Thresholding ---------- 27
- 2.2.2 Spatial Image Dendrograms and the Image Histogram. ---------- 35

2.3 Properties of Image Dendrograms ---------- 41
- 2.3.1 Discretisation Effects ---------- 41
- 2.3.2 Filtering Properties ---------- 41
- 2.3.3 Invariance Properties ---------- 42

2.3.4 Robustness -- 42
2.3.5 Significance Extraction --- 43

CHAPTER 3 NEW IMAGE DENDROGRAM CREATION -------------- 51

3.1 Introduction -- 52

3.2 Initialisation --- 54
3.2.1 Histogram Creation --- 54
3.2.2 Creation of the Label Matrix --------------------------------------- 56

3.3 Segmentation --- 57
3.3.1 Segmentation of Histogram Bins ------------------------------------ 58
3.3.1.1 Segmentation Case One --------------------------------------- 59
3.3.1.2 Segmentation Case Two --------------------------------------- 61
3.3.1.3 Segmentation Case Three ------------------------------------- 64

3.4 Copying Cluster Information -- 69

3.5 Accommodating MIS --- 71

3.6 Results --- 73

3.7 Discussion -- 80

CHAPTER 4 THREE DIMENSIONAL DENDROGRAM VISUALISATION -- 83

4.1 Introduction -- 83

4.2 Dendrogram Creation --- 84
4.2.1 Programming Language --- 84

4.3 Visualisation of the Dendrogram ------------------------------------- 85

4.4 Application Environment --- 93

4.5 Discussion -- 97

CHAPTER 5 BREAST BORDER DETECTION ------------------------ 99

5.1 Introduction -- 99

5.2 Dendrone Slenderness Ratio --- 102

5.3 Average Border Gradient --- 104

5.4 Breast Border Detection Process ------------------------------------ 106

5.5 Results --- 110

5.6 Discussion ---------- **114**

CHAPTER 6 STELLATE DETECTION ---------- **118**

6.1 Introduction ---------- **118**

6.2 Stellate Suspect Detection ---------- **119**

6.3 Stripping ---------- **120**
6.3.1.1 Compactness ---------- 123

6.4 Image Database and Detection Criteria ---------- **126**

6.5 Stellate Suspect Detection Results ---------- **128**

6.6 Suspect Object Classification ---------- **139**
6.6.1 Shape and Orientation Parameters ---------- 139
6.6.2 Texture Parameters ---------- 146
6.6.3 Artificial Neural Network ---------- 151
6.6.3.1 Neural Network Detection Results ---------- 156

6.7 Discussion ---------- **159**

CONCLUSION ---------- **162**

APPENDIX ---------- **177**

REFERENCES ---------- **204**

LIST OF FIGURES

Figure 1-1 *Typical Screening Process.*-- 5

Figure 1-2 *Typical Region of Interest with Stellate.*---------------------------- 8

Figure 1-3 *Typical Mammogram Images – MLO views are in the top row and CC views are on the bottom row.* --------------------------------------11

Figure 1-4 *General framework for malignant mass detection methods.*------13

Figure 2-1 *Samples of Two Parameters (Left), and their Dendrogram (Right)* --25

Figure 2-2 *Image (Left) and Visualisation of Intensity as Topological Elevation.* ---28

Figure 2-3 *An artificial 2-D Intensity Terrain - Initially Flooded with Water.* --29

Figure 2-4 *Water Level Decreases to Reveal First Island (Cluster of Connected Pixels).* ---30

Figure 2-5 *The Original Island Remains and a New One Appears.* ----------31

Figure 2-6 *Two Clusters Merge to Form One Cluster.*------------------------31

Figure 2-7 *The Main Branch Grows and Another Island Appears.*-----------32

Figure 2-8 *The Next Four Levels of Segmentation.*----------------------------33

Figure 2-9 *Hierarchically Segmented. Imaginary One Dimensional Dataset (a) and Corresponding Dendrogram (b).* ---------------------------------34

Figure 2-10 *An artificial image consisting of a black background, two bright discs, one with a brighter rectangle and triangle within it (a). Its corresponding dendrogram (b).* ---44

Figure 2-11 *The same image with high amplitude noise added (a). The main branches of the corresponding dendrogram are still persistent* (b). ----45

Figure 2-12 *The filtered dendrogram. In (a), the MIS is set to 100. In (b), the MIS is set to 1500, and the structure is nearly identical to that of the original image dendrogram without noise.* -------------------------------46

Figure 2-13 *Image reconstructed from the filtered dendrogram (MIS = 1500).* ---47

Figure 2-14 *The four objects detected by the dendronic structure.* ----------49

Figure 3-1 *A matrix of pixel objects are created then pointers to each pixel object are stored in the histogram bins accordingly.*---------------------55

Figure 3-2 *Neighbourhood locations using 4-connectivity.* -------------------58

Figure 3-3 *Bin_1 is segmented - There are no qualified labels in Q, therefore a new cluster, C_1 is created, and $^PP_{1,1}$ from Bin_1 is added to the cluster.* ---60

Figure 3-4 An empty *Cluster,* C_2 is created on the next level. C_2 becomes the parent of C_1, and C_1 a child of C_2.---60

Figure 3-5 *The first pointer in H_2 is segmented. $Q_{1,2}$ is found to have one neighbour cluster pointer, $^PC_1$. The uppermost parent of $^PC_1$ is found to be $^PC_2$.* --61

Figure 3-6 *For location (3,4), there is no qualified label, and a new cluster, C_3 is created and pixel pointer $^PP_{3,4}$ is added to it.* -----------------------62

Figure 3-7 Clusters C_4 and C_5 are created in level 3, as parents of clusters C_2 and C_3, respectively.---63

Figure 3-8 *H_3 is completed, and most of H_4.* ----------------------------------65

Figure 3-9 *Cluster C_8 is merged into cluster C_7 and pixel pointer $^PP_{3,3}$ is added to it.* --65

Figure 3-10 *Location (1,3) is found to have two qualified labels, whose uppermost parents are $^PC_8$ and $^PC_9$.* --------------------------------------66

Figure 3-11 *Cluster C_9 is merged into clusterC$_8$ and cluster C_7 becomes a child cluster of C_8.*---67

Figure 3-12 *The completed dendrogram. Cluster C_9 corresponds to the entire image, as it contains all the pixels held within its child clusters through heirarchical linkage.* ---68

Figure 3-13 *The dendrone branch with Cluster C_7 as the base, the pixel locations it contains and the cluster of pixels it correlates to within the image (red).*---68

Figure 3-14 *Cluster C_2 created on the next level, level 2. It becomes the parent of Cluster C_1, and then segmentation of H_2 can commence.* -----70

Figure 3-15 *Benchmark Mammogram - ACGLCC.* ----------------------------73

Figure 3-16 *Dendrogram Creation Times for Different Values of Stride.* ---74

Figure 3-17 *264x462 Pixel Image for comparison.* ----------------------------75

Figure 3-18 *Creation time comparison. Adjusted results are also presented to account for processor speed.* --77

Figure 4-1 *Two-dimensional Dendrogram and its Image.*----------------------85

Figure 4-2 *Three-dimensional representation of a Dendrone branch.*-------87

Figure 4-3 *Artificial mammogram with two objects present within the breast (a). Image coordinates for two-dimensions (b) and Warped with intensity (c). Dendrone created for first ten levels (d) and showing its three-dimensional axes (e).* ---88

Figure 4-4 *VTK Rendering Pipeline for Dendrogram Visualisation.* --------90

Figure 4-5 *Document object in memory with two attached view objects, the Image and Dendrogram views.* --94

Figure 4-6 *Application Interface. Dendrogram View on the left and Image View on the right.* ---94

Figure 4-7 *Dendrogram is zoomed and a branch has been selected (red). The cluster of pixels that the branch represents is highlighted in red in the Image View also.* ---95

Figure 4-8 *Application process for tumour detection.* ------------------------96

Figure 5-1 *Mammogram (Right) and Corresponding Dendrogram (Left).* 100

Figure 5-2 *Two examples of dendrones. A 'skinny' dendrone (a), and a more 'bushy' dendrone (b).*-- 102

Figure 5-3 *Template for border gradient calculation.* ---------------------- 104

Figure 5-4 *Exaggerated breast border profile of intensity. Three contours of the breast are shown in red. The gradient of the border at each of the contours increases as the intensity increases.*--------------------------- 106

Figure 5-5 *Breast and marker detection comparison: Using Slenderness Ration alone - (a) and (c). Using Slenderness ratio as well as Average Border Gradient - (b) and (d).*--- 108

Figure 5-6 *ABQLML border Detection.*--------------------------------------- 110

Figure 5-7 *ABULCC border Detection.*--------------------------------------- 111

Figure 5-8 *ABRRML border Detection.* -------------------------------------- 111

Figure 5-9 *ABWLCC border Detection.* -------------------------------------- 112

Figure 5-10 *ACELCC border Detection.* -------------------------------------- 112

Figure 5-11 *ADBRML border detection.*-------------------------------------- 113

Figure 5-12 *Two examples of detected breast regions. A typical example (a) and an example of a mismatch in shape (b).*----------------------------- 115

Figure 6-1 *A region of interest and its corresponding dendrogram.*------- 120

Figure 6-2 *Stripping sequence. Left to right, top to bottom. First image corresponds to the full dendrogram of the region. Last two images are the remaining mass, along with the red dendrone that encodes it in the otherwise yellow dendrogram.* --- 122

Figure 6-3 *Original mammogram (a) and its normal tissue 'Stripped' revealing suspect objects (b).* --- 123

Figure 6-4 *A Mass with two concentric circles – The Green embracing circle, and the Blue circle of equivalent area to the mass.* --------------------- 124

Figure 6-5 *Removal of elongated suspects – Removed suspects are in green, remaining suspects are in red.* --- 125

Figure 6-6 *Initial Detection FROC Curves for Stride = 1.*------------------ 130

Figure 6-7 *Initial Detection FROC Curves for Stride = 2.*------------------ 130

Figure 6-8 *Best Case FROC Curves for Initial Detection.* ------------------ 131

Figure 6-9 *Initial Suspect Detection FROC by Case - A Comparison.*---- 133

Figure 6-10 *Abwlml image with the results of 'Stripping'. Radiologist's outlines for the breast border (pink) and large stellate (yellow) can be seen.* -- 134

Figure 6-11 *Aczlcc image with the results of 'Stripping'. Radiologist's outlines for the breast border (pink) and large stellate (yellow) can be seen.* -- 135

Figure 6-12 *Abqlcc with radiologist's outline (a). Stellate has been detected, however, it has been ruled out due to being incompact (b). Had the child cluster of that branch been selected, the stellate would not have been ruled out (c).*-- 137

Figure 6-13 *Abtrml with radiologist's outline (a). Stellate has been detected, however, it has been joined with other objects and ruled out due to being incompact (b). Had one of the child clusters of that branch been selected, the stellate would not have been ruled out (c).* --------------- 137

Figure 6-14 *Hierarchical Repartment of Compactness of malignant tumours, plotted against the tumour dendrone branch length (log scale).* ------ 141

Figure 6-15 *Illustration of blood vessel and stellate distances from the nipple.* ---------- 143

Figure 6-16 *Illustration of blood vessel and stellate orientations relative to the nipple.* ---------- 144

Figure 6-17 *Histograms of Pixel Intensity for Normal Tissue (a). Malignant Mass (b).* ---------- 146

Figure 6-18 Examples skewed distributions ---------- 149

Figure 6-19 *Feed Forward Artificial Neural Network.* ---------- 152

Figure 6-20 *Classification FROC, both by Image and by Case.* ---------- 156

Figure 6-21 *FROC Area of Interest.* ---------- 157

Figure 6-22 *Two examples of detected breast regions. A typical example (a) and an example of a mismatch in shape (b).* ---------- 168

Figure 6-23 *Original image (a), Dendrogram with pectoral branch in red (b) and the corresponding region of the pectoral muscle (c).* ---------- 170

LIST OF TABLES

Table 3-1 *Benchmark Mammogram Details.* 74

Table 3-2 *Creation time comparison for 264x462 pixel image.* 76

Table 3-3 Percentage of Time Spent in Each Creation Segment (Chen, 2000). 78

Table 3-4 Percentage of Time Spent in Each Creation Segment – New Image Dendrogram Construction Method. 78

Table 6-1 *Mammogram Image Database.* 126

Table 6-2 *Dendrogram Creation Parameters for Initial Detection Optimisation.* 128

Table 6-3 *Stripping Parameters for Initial Detection Optimisation.* 129

Table 6-4 *Initial Detection Parameters*.. 132