Visualisation and Topological Aspects of Higher Dimensional Data

Final Report for CS39440 Major Project

Author: Samuel Jackson (slj11@aber.ac.uk)
Supervisor: Prof. My Supervisor (rrz@aber.ac.uk)

March 24, 2015
Version: 1.0 (Draft)

This report was submitted as partial fulfilment of a MEng degree in Software Engineering (G601)

Department of Computer Science
Aberystwyth University
Aberystwyth
Ceredigion
SY23 3DB
Wales, UK

Declaration of originality

In signing below, I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for plagiarism and other unfair practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the sections on unfair practice in the Students' Examinations Handbook and the relevant sections of the current Student Handbook of the Department of Computer Science.
- I understand and agree to abide by the University's regulations governing these issues.

Signature .....

Date .....

Consent to share this work

In signing below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Signature .....

Date .....

Acknowledgements

I am grateful to...

I’d like to thank...

Abstract

Include an abstract for your project. This should be no more than 300 words.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Background & Objectives

1.1 Mammography

Breast cancer is the leading cause of death among women and is the most common form of cancer found in women [6]. Early screening of breast cancer using mammography has been shown to reduce the mortality rate of women [4, 7].

Mammography is the analysis of female breast tissue through the use of X-ray radiology with the goal of producing high resolution images of the structure within the female breast. The composition of the parenchymal patterns and tissue density revealed by in a mammographic evaluation can be used in the early detection of breast cancer.

Qualitatively speaking the composition of breast tissue can be split into four distinct categories. These are Nodular densities (corresponding to Terminal Ductal Lobular Units (TDLUs), linear densities (corresponding to ducts, vessels, and fibrous strands), homogeneous, structureless densities (corresponding to fibrous supporting tissue), and radiolucent areas (corresponding to adipose tissue). Typical markers used in the detection of cancer can are the presence of clusters of microcalcifications, masses, architectural distortions, breast density and parenchymal patterns [2, 5].

1.1.1 Risk Assessment

Mammograms provide a non-invasive means to assess the risk of a patient developing cancer given a set of mammographic images. There are multiple classification systems used in the classification of mammographic risk. The composition of breast tissue can be categorised using the Breast Imaging Reporting And Data System (BI-RADS) [3]. BI-RADS classifies mammograms based on the density of tissue (and therefore risk) in the mammogram.

1.2 Features

1.2.1 Shape Features

1.2.2 Texture Features

1.3 Dimensionality Reduction

1.3.1 Linear

1.3.2 Non Linear

1.4 Visualisation

1.5 Analysis

1.6 Research Method

Chapter 2

Experiment Methods

2.1 Overview

2.2 Techniques

2.2.1 Features

2.2.2 Dimensionality Reduction

2.2.3 Visualisation

2.3 Datasets

2.3.1 Synthetic Data

2.3.2 Real Data

2.4 Implementation

2.4.1 Languages

2.4.2 Libraries

Chapter 3

Results and Conclusions

3.1 Comparison of Real and Synthetic Datasets

3.2 Investigation of Mapping

Chapter 4

Critical Evaluation

4.1 Conclusions

4.2 Evaluation of the Project

4.3 Future Work

Appendices

Appendix A

Third-Party Code and Libraries

Appendix B

Code samples

Annotated Bibliography

[1] I. T. Gram, E. Funkhouser, and L. Tabár, "The tabar classification of mammographic parenchymal patterns," *European journal of radiology*, vol. 24, no. 2, pp. 131–136, 1997.

[2] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiology Biomarkers & Prevention*, vol. 15, no. 6, pp. 1159–1169, 2006.

[3] A. C. of Radiology. BI-RADS Committee and A. C. of Radiology, *Breast imaging reporting and data system*. American College of Radiology, 1998.

[4] I. U. P. on Breast Cancer Screening *et al.*, "The benefits and harms of breast cancer screening: an independent review," *The Lancet*, vol. 380, no. 9855, pp. 1778–1786, 2012.

[5] M. P. Sampat, M. K. Markey, A. C. Bovik, *et al.*, "Computer-aided detection and diagnosis in mammography," *Handbook of image and video processing*, vol. 2, no. 1, pp. 1195–1217, 2005.

[6] R. Siegel, J. Ma, Z. Zou, and A. Jemal, "Cancer statistics, 2014," *CA: a cancer journal for clinicians*, vol. 64, no. 1, pp. 9–29, 2014.

[7] R. A. Smith, D. Manassaram-Baptiste, D. Brooks, V. Cokkinides, M. Doroshenk, D. Saslow, R. C. Wender, and O. W. Brawley, "Cancer screening in the united states, 2014: a review of current american cancer society guidelines and current issues in cancer screening," *CA: a cancer journal for clinicians*, vol. 64, no. 1, pp. 30–51, 2014.

[8] J. N. Wolfe, "Breast patterns as an index of risk for developing breast cancer," *American Journal of Roentgenology*, vol. 126, no. 6, pp. 1130–1137, 1976.