









Assessment of Obstetric Ultrasound Images using Machine Learning

Bahbibi Rahmatullah

St Catherine's College



Supervisors: Prof J. Alison Noble and Dr Aris T. Papageorghiou Trinity Term 2012

> Institute of Biomedical Engineering Department of Engineering Science University of Oxford





















Abstract

Ultrasound-based fetal biometry is used to derive important clinical information for identifying IUGR (intra-uterine growth restriction) and managing risk in pregnancy. Accurate and reproducible biometric measurement relies heavily on a good standard image plane. However, qualitative visual assessment, which includes the visual identification of certain anatomical landmarks in the image is prone to inter- and intra-reviewer variability and is also time-consuming to perform. Automated anatomical structure detection is the first step towards the development of a fast and reproducible quality assessment of fetal biometry images. This thesis deals specifically with abdominal scans in the development and evaluation of methods to automatically detect the stomach and the umbilical vein within them.

First, an original method for detecting the stomach and the umbilical vein in fetal 05 4506832 pustaka upstedumy abdominal scans was developed using a machine learning framework. A classifier solution was designed with AdaBoost learning algorithm with Haar features extracted from the intensity image. The performance of the new method was compared on different clinically relevant gestational age groups.

Speckle and the low contrast nature of ultrasound images motivated the idea of introducing features extracted from local phase images. Local phase is contrast invariant and has proven to be useful in other ultrasound image analysis application compared with intensity. Nevertheless, it has never been implemented in a machine learning environment before. In our second experiment, local phase features were proven to have higher discriminative power than intensity features which enabled them to be selected as the first weak classifiers with large classifier weight.





















Abstract

Third, a novel approach to improving the speed of the detection was developed using a global feature symmetry map based on local phase to select the candidate locations for the stomach and the umbilical vein. It was coupled with a local intensity-based classifier to form a "hybrid" detector. A nine-fold increase in the average computational speed was recorded along with higher accuracy in the detection of both the anatomical structures.

Quantitative and qualitative evaluations of all the algorithms were presented using 2384 fetal abdominal images retrieved from the image database study of the Oxford Ultrasound Quality Control Unit of the INTERGROWTH-21st project.

Finally, the "hybrid" detection method was evaluated in two potential application scenarios. The first application was clinical scoring in which both the computer algorithm and four experts were asked to record presence or absence of the stomach and the umbilical vein in 400 ultrasound images. The computer-experts agreement was found to be comparable with the inter-expert agreement. The second application concerned selecting the standard image plane from 3D abdominal ultrasound volume. The algorithm was successful in selecting 93.36% of the images plane defined by the expert in 30 ultrasound volumes.

















Acknowledgements

I would like to thank my supervisors Prof. Alison Noble and Dr. Aris Papageorghiou for their help, support and guidance throughout this research degree. I am truly grateful for the opportunity to do this research project with the funding support of the Sultan Idris Education University (UPSI) Malaysia and Malaysian Ministry of Higher Education (MOHE). I would also like to thank Dr Ippokratis Sarris, Dr Christos Ioannau and Dr Caroline Knight for their collaboration and contribution in this work. I am grateful to all of my colleagues within the Biomedical Image Analysis (BioMedIA) Laboratory for thought-provoking discussions, support and kindness. A special thanks to the examiners who have kindly agreed to examine this thesis.

I would also like to thank my family and friends for all their support and encouragement throughout this research. Most importantly, I wish to thank my mother Fatimah Abdullah for her non-stop encouragements and prayers throughout my academic work.

Lastly I cannot express enough thanks to my husband Arfian and my three children, Abdullah, Bilal and Hasanah for their love, patience and support. This thesis is dedicated to them.



















Table of Contents

A	DSTr	ют	······································			
AcknowledgementsIII						
T	able	of Conter	ntsIV			
L	ist of	Figures	VII			
L	ist of	Tables	XII			
C	hapt	er 1 Intro	duction1			
	1.1	Motivation	1			
	1.2	Contribution	ons4			
	1.3	Thesis Out	line6			
C	hapt	er 2 Liter	ature Review7			
	2.1		on7			
	02.206	Ultrasound	I-based Fetal Biometry Perpustakaan Tuanku Bainun 9 pt			
		2.2.1	Weeks 6 - 13 of gestation9			
		2.2.2	Weeks 13 - 25 of gestation			
		2.2.3	Weeks 26 - 42 of gestation			
		2.2.4	Other Anatomical Measurements13			
		2.2.5	3D Ultrasound			
	2.3	INTERGR	OWTH-21 st 14			
	2.4	Challenges	s in Fetal Ultrasound Imaging15			
		2.4.1	Quality of Images15			
		2.4.2	Inter- and Intra-Operator Variability16			
	2.5	Qualitative	Measures in Fetal Biometry Images18			
	2.6	Automated	I Image Analysis in Obstetric Ultrasound21			
		2.6.1	Automated Fetal Biometric Measurement21			
		2.6.2	Other Measurements			
		2.6.3	Summary26			
	2.7	Machine L	earning in Medical Imaging26			















2.8	Conclusi	ons31
Chapt	ter 3 Ana	atomical Object Detection in 2D Fetal Abdominal Ultrasound
		32
3.1		tion
3.2		und on the Object Detection Framework
	3.2.1	Haar Features
	3.2.2	Integral Image
	3.2.3	Adaptive Boosting (AdaBoost)
3.3		ental Setup
	3.3.1	Image Module
	3.3.2	Feature Module
	3.3.3	Learning Module44
	3.3.4	Detector Module46
	3.3.5	Datasets46
	3.3.6	Validation Methodology48
3.4	Results a	and Discussion48
3.5	Conclusi	ons
05-450 Chapt	ter 4 Loc	Pustaka upst edu my Perpustakaan Iuanku Bainun Pustaka Bainun Pust
Detect	tion	57
4.1	Introduc	tion57
4.2		und on Local Phase and Monogenic Signal
4.3		ents
	4.3.1	Features67
	4.3.2	Scale Selection67
	4.3.3	Classifier Training73
	4.3.4	Validation Measures75
4.4	Results	76
4.5	Discussi	on80
4.6	Conclus	ons85
Chapt	ter 5 Fea	ture Symmetry for Efficient Object Detection86
5.1		tion87
	5 1 1	Feature Symmetry 87













Table of Contents

	5.1.2	Scale Selection				
5.2	Experiments					
5.3	Results and	d Discussions94				
5.4	Conclusion	ns106				
Chapt	er 6 Two	Pilot Studies to Illustrate Potential Clinical Utility107				
6.1	Pilot Study 1: Comparison with Inter-Experts Agreement					
	6.1.1	Experiments				
	6.1.2	Results				
	6.1.3	Discussion				
	6.1.4	Conclusion				
6.2	Pilot Study	y 2: Selection of Optimal Plane from Ultrasound Volume115				
	6.2.1	Experiments				
	6.2.2	Results119				
	6.2.3	Discussion				
	6.2.4	Conclusions				
Chapt	er 7 Sum	mary and Future Work124				
7.1	Summary.	124				
7.2	Summary 12 06832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Future Work PustakaTBainun 12					
	7.2.1	Other Anatomical Objects Detection				
	7.2.2	Multi-class Object Detection				
	7.2.3	Testing on Data from Other Ultrasound Machines126				
	7.2.4	3D Object Detection and Planar Slicing127				
Appen	dix A	Observer Agreement Statistics in Clinical Imaging128				
A.1		Percentage of Agreement 128				
A.2		Cohen's Kappa (κ)129				
A.3		Prevalence-Adjusted Bias-Adjusted Kappa (PABAK)130				
A.4		Benchmark Scale of Agreement Statistics				
Biblio	graphy	134				



















List of Figures

during image quality training session
Figure 2.1: An ultrasound showing a fetus measured to have a crown-rump length (CRL) of 73.5mm9
Figure 2.2: Fetal biometric measurements from fetal head ultrasound images showing the three standard measurements (BPD, OFD and HC).
Figure 2.3: Fetal biometric measurements from fetal abdomen ultrasound images showing the three standard measurements (APAD, TAD and AC).
Figure 2.4: Fetal biometric measurement from fetal thigh ultrasound image showing the FL measurement.
Figure 2.5: Different quality of fetal abdomen ultrasound images at various gestational ages (17, 21, 28, 33 and 38 weeks)
Figure 2.6: Examples of fetal ultrasound scan that satisfy Salomon's scoring and diagram showing visible landmarks in standard fetal (a) head and (b) abdominal planes20
Figure 3.1: The position of the ultrasound probe for taking the standard abdominal image plane is represented by dotted lines
Figure 3.2: Examples of good section for fetal abdominal circumference in 17 weeks and 30 weeks fetuses.
Figure 3.3: Two examples of the wrong section for fetal abdominal circumference in 16 weeks (stomach is invisible) and 38 weeks fetuses (umbilical vein is elongated)
Figure 3.4: Prototypes of Haar features and unary feature used in our algorithm36
Figure 3.5: Example of integral image application
Figure 3.6: Example of (a) stomach images (b) umbilical vein images and (c) background images used for training
Figure 3.7: Fetal abdominal area extraction from original image using ellipse fitting43











Figure 3.8: The first ten selected Haar features by AdaBoost are shown superimposed on some example images from the training set for (a) the stomach and (b) the umbilical vein45
Figure 3.9: ROC curve to analyse the effect different number of weak classifiers (WC) (50, 100, 150, and 200) in the classification of (a) the stomach and (b) the umbilical vein46
Figure 3.10: ROC curves for the detection of (a) the stomach and (b) the umbilical vein in different gestational age (GA) groups. GA is defined in weeks from conception51
Figure 3.11: True positive results for stomach detection in different fetal scans at (a) 18 weeks (b) 28 weeks and (c) 38 weeks.
Figure 3.12: True positive results for umbilical vein detection in different fetal scans at (a) 16 weeks (b) 26 weeks and (c) 39 week
Figure 3.13: True negative results for stomach detection in different fetal scans at (a) 17 weeks (b) 30 weeks and (c) 38 weeks.
Figure 3.14: True negative results for umbilical vein detection in different fetal scans at (a) 17 weeks (b) 28 weeks and (c) 38 weeks
Figure 3.15: False positive results for stomach detection in different fetal scans at (a) 19 weeks (b) 29 weeks and (c) 39 weeks
Figure 3.16: False positive results for umbilical vein detection in different fetal scans at (a) 16 weeks (b) 26 weeks and (c) 38 weeks.
Figure 3.17: False negative results for stomach detection in different fetal scans at (a) 19 weeks (b) 29 weeks and (c) 38 weeks
Figure 3.18: False negative results for umbilical vein detection in different fetal scans at (a) 16 weeks (b) 30 weeks and (c) 40 weeks
Figure 4.1: Illustration of the importance of phase where Fourier magnitude spectrum and Fourier phase spectrum were taken from separate images. Inverse Fourier transform was then performed to produce a new image
Figure 4.2: Example of a log-Gabor filter. The transfer function of the filter viewed on both (a) linear and (b) logarithmic frequency scales
Figure 4.3: (a) Original abdominal ultrasound images and its corresponding local phase images for filter scale of (b) 30 (c) 50 and (d) 100 pixels











Figure 4.4: Example of local phase images produced with the filter scale of (a) 50 (b) 150 and (c) 250 pixels which are integrated to produce (d) the multi scale images, (e) - (g) are the original intensity images for each row of local phase images
Figure 4.5: ROC analysis for the classification result of (a) the stomach and (b) the umbilical vein in the validation set using local phase features from multi scale (MSLP) and single scale (SSLP) filters
Figure 4.6: The first five features together with its weight (α) designated by AdaBoost algorithm for the stomach detection superimposed on the sample image
Figure 4.7: The first five features together with its weight (α) designated by AdaBoost algorithm for the umbilical vein detection superimposed on the sample image75
Figure 4.8: ROC curves for the detection of (a) the stomach and (b) the umbilical vein using three different types of feature sets
Figure 4.9: Comparison of ROC curves between "Intensity+MSLP" and "Intensity only" methods in different gestational age groups
Figure 4.10: True positive results for stomach detection by the "Intensity+MSLP" method (blue box with α_T) where it corrected the false detection result achieved by using "Intensity" features (red box with α_{INT}).
Figure 4.11: True positive results for umbilical vein detection by the "Intensity+MSLP" method (blue box with α_T) where it corrected the false detection result achieved by using "Intensity" features (red box with α_{INT}).
Figure 4.12: Example of images where objects were failed to be detected correctly. The blue box and α_T represent the detection by the "Intensity+MSLP" method, and the "Intensity" method detection is represented by the red box and α_{INT} .
Figure 4.13: The false positive results for umbilical vein detection in images at 38 weeks. The umbilical veins were too elongated, hence not acceptable under the scoring criteria
Figure 5.1: Examples of feature symmetry images produced using different scales combinations and threshold with three feature significance values91
Figure 5.2. Two examples of the global detector application93
Figure 5.3: ROC curves for the detection of (a) the stomach and (b) the umbilical vein, using "Local", "Global" and "Hybrid" methods96









Figure 5.4: Comparison of ROC curves between "Hybrid" and "Local" methods in different gestational age groups
Figure 5.5: Examples of stomach detection where the false negative detection by the "Local" method were corrected by the "Hybrid" method
Figure 5.6: Examples of umbilical vein detection where the false negative detection by the "Local" method were corrected by the "Hybrid" method.
Figure 5.7: Examples of the misdetection of the stomach in the $38^{+0} - 42^{+6}$ weeks images
Figure 5.8: Examples of the umbilical vein misdetections in the $38^{-0} - 42^{+6}$ weeks images
Figure 5.9: The three negative stomach images where false positive detection (high α_T score) by the "Local" method were corrected by true negative result (low α_T score) using the "Hybrid" method
Figure 5.10: The negative image that was missed by both "Local" and "Hybrid" methods. The scores achieved by both methods were higher than the threshold value and resulted in a false positive detection
Figure 5.11: Examples of negative umbilical vein images where false positive detection (high α _T score) by the "Local" method were corrected by true negative result (low α _T score) using the "Hybrid" method
Figure 6.1: Illustration of different slices acquired from a 3D volume at different positions on the fetus
Figure 6.2: Flowchart showing the implementation of the training and testing phase
Figure 6.3: Graphs showing the normalized classifier scores achieved by the detector for each image plane in two sample volumes.
Figure 6.4: Precision and recall values (in percentages) achieved for the selection of standard planes from 30 fetal abdominal volumes.
Figure 6.5: Graph plot and image planes for Volume 7
Figure 6.6: Graph plot and image planes for Volume 18

















List of Figures





Figure 6.7: Graph	plot and image	planes from	the volume	with 100%	precision	and recal
values	***********************	***************************************				122

Figure 6.8: Graph plot and image planes from the volume with lowest recall percentage....123



















List of Tables

Table 2.1: Objective scoring system for still images (Salomon et al., 2006)20
Table 2.2: Summary of research in fetal ultrasound image analysis in chronological order24
Table 3.1: AdaBoost Algorithm (modified from (Viola and Jones, 2004))40
Table 3.2: Number of features extracted from a 100x100 window
Table 3.3: Details of the number of positive (+) and negative (-) images in the training, validation and testing datasets
Table 3.4: Distribution of images in the testing datasets for different gestational age groups.
Table 3.5: Overall performance evaluation for the stomach and the umbilical vein detection50
Table 4.1: Number of unary features extracted from a 100x100 window67
Table 4.2: Filter scales for the stomach and the umbilical vein
Table 4.3: The scale of filter, the weight and the accuracy of the first ten features selected by AdaBoost for the parameter determination of single-scale local phase implementation69
Table 4.4: Details on the selected local phase (LP) features in the stomach and the umbilical vein trained classifier
Table 4.5: Performance evaluation for the detection of the stomach and the umbilical vein between "Intensity+MSLP" and "Intensity only" methods in different gestational age groups. 79
Table 5.1: Different combinations of band-pass filter scale used to produce the feature symmetry measure
Table 5.2: The filter scales combination and the weight of the first ten features selected by AdaBoost from the pool of unary features extracted from feature symmetry images90
Table 5.3: The overall performance of the three different methods in the detection of the stomach and the umbilical vein in fetal abdominal images
Table 5.4: Performance evaluation for the detection of the stomach and the umbilical vein between "Local" and "Hybrid" methods in different gestational age groups95











100 ultrasound images (Dataset 1) between the automated method and the experts
Table 6.2: Confusion matrices for the classification of the stomach and the umbilical vein in 100 ultrasound images (Dataset 1) between the experts
Table 6.3: Confusion matrices for the classification of the stomach and the umbilical vein in 300 ultrasound images (Dataset 2) between the automated method and the experts
Table 6.4: Confusion matrices for the classification of the stomach and the umbilical vein in 300 ultrasound images (Dataset 2) between the experts
Table 6.5: Percentage of agreement and adjusted kappa value between the automated method (AM) and the experts (E1, E2, E3, E4) for Dataset 1
Table 6.6: Percentage of agreement and adjusted kappa value between the automated method (AM) and the experts (E1, E2, E3, E4) for Dataset 2
Table A.1: Landis and Koch – Kappa's Benchmark Scale
Table A.2: Fleiss – Kappa's Benchmark Scale
Table A.3: Altman – Kappa's Benchmark Scale













Chapter 1 Introduction

1.1 Motivation

Small for gestational age (SGA) refers to the situation when a fetus is smaller than expected for the number of weeks of pregnancy. Newborn babies with SGA are often associated with having intrauterine growth restriction (IUGR), which is a more specific condition where the fetus fails to reach its growth potential. 18 million babies are born every year with low birth weight because of IUGR and/or prematurity, resulting in significant short-and long term morbidity and mortality (Lawn et al., 2005). Growth restricted fetuses have poorer neonatal outcomes and it is recognised that developmental delay associated with IUGR leads to significant health care and developmental problems during childhood and most likely in adult life (Barker, 2006). Recognition of the serious risks associated with IUGR has elevated its diagnostic importance among perinatologists. Thus, obtaining accurate assessment of fetal growth and gestational age from fetal biometry for identifying risks to the fetus/neonate is very important.

Historically, X-ray was used to measure fetal dimensions (e.g. fetal head, pelvic dimension) (Shenton, 1922) before the development of ultrasound. The development of two-dimensional (2D) ultrasound made it possible to measure the dimensions of bones and soft tissue structures of the fetus faster and more reliably than with x-rays. 2D ultrasound is currently considered to be the first choice for a safe, non-invasive, accurate and cost-effective investigation in the fetus. It has progressively become an indispensable obstetric tool and plays an important role in pregnancy management. Comprehensive ultrasound examination during pregnancy includes standard fetal biometric measurements, which are primarily used to estimate the gestational age of the fetus, to track fetal growth patterns, to estimate fetal











weight and to detect abnormalities. A detailed description of ultrasound-based fetal biometry used for age estimation and growth assessments is given in Section 2.2.

Fetal biometry is determined from standardized ultrasound planes taken from the fetal head, abdomen and thigh. The acquisition of optimal image planes from which these measurements are taken is crucial to allow for accurate and reproducible biometric measurements, and also to minimize inter- and intra-observer variability. Criteria and description of standard fetal biometric planes are presented in Section 2.5.

The importance of quality control for the scanning procedures and measurements has been emphasized (Dudley, 2006, Ville, 2008) and a quality control policy based on image scoring has been proposed (Salomon et al., 2006). To highlight challenges in scanning and acquiring the standard image plane, scans made by several different sonographers for finding the abdominal measurements after they had been briefed on the scanning protocol for a growth study known as INTERGROWTH-21st are shown in Figure 1.1. Even though all the scans shown in the figure are magnified satisfactorily, the appearance of the stomach and the umbilical vein are inadequate in some of the scans. According to Salomon's grading, scans in the first column are acceptable with the stomach and the umbilical vein clearly identified and in the correct position. However, elongated umbilical vein appearance in the second column's images indicates that the plane is too angled.











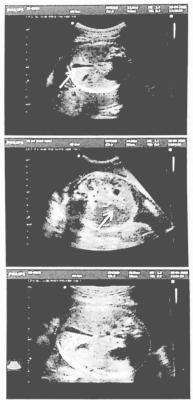
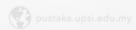


Figure 1.1: Scans acquired by different sonographers for finding abdominal measurement during image quality training session. The stomach and the umbilical vein (white arrows) are clearly visible in the images in the first column. However, elongated umbilical vein in images in the second column indicates that the plane is too angled.

There are several limitations of visual image quality assessment. The process of image review, which includes the visual identification of certain anatomical landmarks in the image, requires significant human resources. Extensive qualitative analysis is also time-consuming and costly. Furthermore, there is an issue of inter- and intra-reviewer variability and also bias imposed by a human reviewer. An automated image scoring system, which 1) can perform the evaluation quickly, 2) is robust to appearance variations of the visual object of interest, and 3) efficient and economical for any scale of implementation would be a valuable support to the quality control process.







This thesis deals specifically with the development of automated methods for the detection of two important landmarks (the stomach and the umbilical vein) in fetal abdominal ultrasound scans using machine learning. The plane containing these two landmarks is described in the early proposal for using the abdominal circumference measurement for fetal weight estimation (Campbell and Wilkin, 1975). The plane containing these two landmarks was adopted in constructing the widely used chart for abdominal circumference size (Chitty et al., 1994) and also proposed in the image quality scoring system (Salomon and Ville, 2005).

1.2 Contributions

The main contributions of this thesis are summarized below:

- 1. The development of an original method to detect the stomach and the umbilical vein in fetal abdominal scan using a machine learning technique (Chapter 3). Parts of this
 - chapter have been published at peer-reviewed conferences:



- Quality Control of Fetal Ultrasound Images: Detection of Abdomen Anatomical Landmarks using Adaboost. IEEE International Symposium on Biomedical Imaging (ISBI), 2011.
- Image Analysis Using Machine Learning: Anatomical Landmarks Detection in Fetal Ultrasound Image. IEEE Signature Conference on Computers, Software, and Applications (COMPSAC), 2012.
- 2. The investigation of introducing features extracted from the local phase image into the machine learning framework for the detection of the two anatomical landmarks











¹ Local phase concept used in this thesis is different than the term 'phase' used in the radio frequency signal used at acquisition time for ultrasound images. Detailed explanation can be found in Chapter 4.

(stomach and umbilical vein) (Chapter 4). Part of this chapter has been published at a peer-reviewed conference:

- i. Multi-Scale Local Phase Features for Anatomical Object Detection in Fetal Ultrasound Images. Medical Image Understanding and Analysis Conference (MIUA), 2012.
- 3. The development of a faster and more accurate detector using a hybrid approach for the detections of the stomach and the umbilical vein (Chapter 5). Part of this chapter has been published at a peer-reviewed conference:
 - Integration of Local and Global Features for Anatomical Object Detection in i. Ultrasound. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2012.
- 4. The evaluation of the proposed detection method in two clinical application scenarios (Chapter 6). Parts of this chapter has been published at a peer-reviewed conference:
 - i. Automated Selection of Standardized Planes from Ultrasound Volume. MICCAI Workshop on Machine Learning in Medical Imaging (MLMI), 2011.

and as abstracts in the following clinical meetings:

- ii. A Pilot Study of Automated Image Scoring for Quality Control Purposes in the Context of Multicentre Studies: Abdominal Circumference. World Congress on Ultrasound in Obstetrics and Gynecology, 2011.
- Automated Fetal Biometry Image Landmark Detection for Confirming Correct iii. Image Planes: Abdominal Circumference. World Congress on Ultrasound in Obstetrics and Gynecology, 2012.











iv. Automated Standard Plane Selection from Fetal Abdominal Ultrasound Volumes using a Machine Learning Algorithm. World Congress on Ultrasound in Obstetrics and Gynecology, 2012.

1.3 Thesis Outline

Chapter 2 describes the background knowledge on fetal growth restriction and the current clinical practice which uses ultrasound for its assessment along with its challenges and the quality control process. The chapter also provides the review of related image analysis work in fetal ultrasound domain and the application of machine learning for detection purposes in medical imaging.

Chapter 3 describes the initial method used for the detection of important anatomical landmarks in fetal abdominal ultrasound images using machine learning framework.

Chapter 4 deals with utilizing features from multi-scale local phase images in the same detection framework. The efficiency of the new feature sets are compared to the performance intensity-based features used in Chapter 3.

Chapter 5 introduces a new hybrid approach for the enhancement of the performance and the speed of the detection. A multi-scale feature symmetry measure derived using local phase is combined with the local intensity-based detector (developed in Chapter 3) is utilized for fast object detection and its detection performance is analysed.

Chapter 6 evaluates the application of the proposed algorithm in two potentials scenarios: comparison with experts' agreements in recording the presence and absence of the anatomical structures in fetal abdominal scan and utilizing the algorithm for the selection of standard plane from 3D volumes.

Chapter 7 concludes the thesis and discusses directions for future work.









